Roadmap Thematic Digital Competence Center - v1.0

Domain Natural and Engineering Sciences

Lead authors

Anthony Brown, Universiteit Leiden
David Groep, Nikhef
Marta Teperek, 4TU.ResearchData

Contributing authors

Irene Dedoussi, TUD
Vera Sarkol, CWI
Wim Som de Cerff, KNMI

NWO-secretaries

Erik van Aert, Domain Sciences
Paul Blank, Domain Applied and Engineering Sciences

All authors have been listed in alphabetical order by surname (per category).

Prologue

Digital competences cover a broad range of topics, in particular in the natural and engineering sciences ('NES'). Beyond data competences they also comprise the competences needed to effectively work software needed to understand, process, enrich, analyse and visualise the digital artefacts and extract knowledge therefrom. Digital competences are also essential for the use of the facilities for computing and processing, the frameworks to bring together software, data, and compute facilities.

Methodologies, models and approaches used are different for each of the disciplines within the NES domain. Yet the domain has much in common: similar software libraries and programming techniques are used in very different fields, the challenges related to voluminous data, or to lack of standardisation and interoperability are shared between many disciplines.

Several initiatives are already in place to aid research groups in becoming more effective and more prolific users of digital tools. At the institutional level, the local Digital Competence Centres (LDCCs) support research groups with data management plans, research software engineers, and local expertise. The LDCC network enables practice exchange between the local DCCs. Efforts are underway to stimulate the hands-on exchange of expertise between institutions to support research groups that span more than one institution, or where digital facilities from national peers could be of help. Examples are the SURF initiative for aligning institutional HPC capabilities, the Netherlands eScience Center, SURF, and DANS.

So what does a Thematic DCC add to this landscape? First and foremost, it is there to -*complement*, not to replace or control, these existing groups. It collaborates with the LDCCs where it can provide a national expertise map and aid them in finding appropriate experts from the national, European, or global arena. It connects research communities to existing infrastructures, links them with the Roadmap Facility experts and 'digital side' of the European Research Infrastructures and ESFRI clusters, and discipline-specific tools and software that is out there. The TDCC can thus facilitate the re-use of expertise gained in participation of research groups in global efforts.

The second important aspect of a TDCC is to exchange best practice across disciplines and institutes. The TDCC will support the creation of a national expertise map, thus fostering the re-use of interoperable tools and software, with the associated training, to make adoption of existing tools more attractive than developing a new one from scratch. This will increase expertise among the researchers themselves so that they can identify and steer initiatives in their own field.

The 'NES TDCC' should be considered as part of a collaborating suite of thematic DCCs. The three TDCCs will collaborate with one another and share results.

The broad consultation in preparation to this roadmap not only identified 'generic' needs such as training and networking, or 'adoption of FAIR practices'. Specific NES domain challenges include: long term data preservation; the need for integration of tools and workflows that bring the FAIR practice to life; and software sustainability. While the level of these challenges is obviously different for each discipline, how to *identify where the best practice is*, is where the TDCC will help.

The combination of generic challenges (the expertise map and community building, training) and the more NES-specific challenges (reusable data in the computational environment, software sustainability) together inspired the 'bottleneck projects' that receive immediate attention when the TDCC starts. Beyond the bottleneck projects the TDCC is set up to evolve in line with the needs of the domain.

A lean organisational structure is put in place to guide the evolution of the TDCC on its own road to sustainability. The guiding principles of the TDCC organisation are: (i) activities and knowledge is open to everyone, (ii) the work programme is coherent and the projects within it contribute towards the shared work programme, (iii) who contributes to the TDCC activities helps steer its direction, (iv) digital competences are best developed in collaboration.

Description of domain

The domain of natural and engineering sciences is large and diverse: not only with regards to the multitude of disciplines covered within the domain (from astronomy to civil engineering), but also with regards to its fragmentation and diverse needs for data and software within the domain.

Institutions, organisations, infrastructures, public/private parties

Research Universities and Institutes

In addition to the Dutch universities, which are the main places where research within the domain of natural and engineering sciences is conducted, the NWO-I, the Institutes Organisation of NWO, manages nine national research institutes: AMOLF, ARCNL, ASTRON, CWI, DIFFER, Nikhef, NIOZ, NSCR and SRON. With the exception of NSCR (Netherlands Institute for the Study of Crime and Law Enforcement), all of them are part of the domain of natural and engineering sciences. Some institutes have elaborate (international) standards for interoperability of digital objects, such as ASTRON, Nikhef, NIOZ, CWI and SRON.

Coordination and networking

There are multiple organisations that coordinate and connect the digital research landscape within the NES domain in the Netherlands. These include Large Scale Research Infrastructures (see also below) which play an important coordinating role, as well as national institutes, which stimulate research within specific disciplines and themes, such as Nikhef, SRON, ASTRON, AMOLF, ARCNL, CWI or KNMI. In addition, 4TU.ResearchData plays a role in coordination and networking specifically within the domain of natural and engineering sciences, in particular through its community building activities.

Repositories for data and software sharing

The NPOS report 'Eindrapport Verkenning en optimalisering nationaal datalandschap' (2020) offers an overview of repositories for data across all the domains within the Netherlands based on the analysis of the www.re3data.org database in March 2020. There are 22 repositories that cater for data from Natural Sciences and 8 for data from Engineering Sciences. Interestingly, only 3 out of 30 repositories within the entire domain are certified, and out of these, only 4TU.ResearchData provides support for sharing and long-term preservation of both research data and software and for disciplines from the entire NES domain.

It should be noted that the NES domain has strong international connections and links to international infrastructure, thus for disciplinary research international repositories are also in widespread use (see also the section about the International landscape).

Training

RDNL has been recognised as the key training provider by the aspiring Data Stewards[1]. In addition, SURF and the eScience Centre have been providing training on computational skills. However, there is no dedicated organisation to provide training for data and software management specifically for researchers from the natural and engineering sciences domain.

Infrastructure

Out of 122 large infrastructures identified in the national report on Dutch large scientific infrastructures (2021), 35% of them are within the domain of natural and engineering sciences. Within the domain, the large infrastructures are further split into four groups: Astronomy & Particle Physics, Materials, Technology and GeoSciences. As the report concludes, these infrastructures frequently consist of "large-scale and expensive measuring equipment and produce large quantities of data. Over two-thirds of the infrastructures in this domain are fully international in nature, such as Magnum-PSI, KM3NeT, Extremely Large Telescope (E-ELT),

[1] Mijke Jetten, Marjan Grootveld, Annemie Mordant, Mascha Jansen, Margreet Bloemers, Margriet Miedema, & Celia W.G. van Gelder. (2021). Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship (1.1). Zenodo. https://doi.org/10.5281/zenodo.4623713

Laser Interferometer Space Antenna (LISA), ITER and European Spallation Source (ESS). Examples of national initiatives are: Climate and Meteorological Network (C-MetNet), National Characterisation Centre for Sustainable Materials (NC2SM) and Aerodynamics and Propulsion Laboratory (APP-lab)". Design and construction of these infrastructures regularly involves participation in expensive, high-profile international initiatives.

Need for structural funding

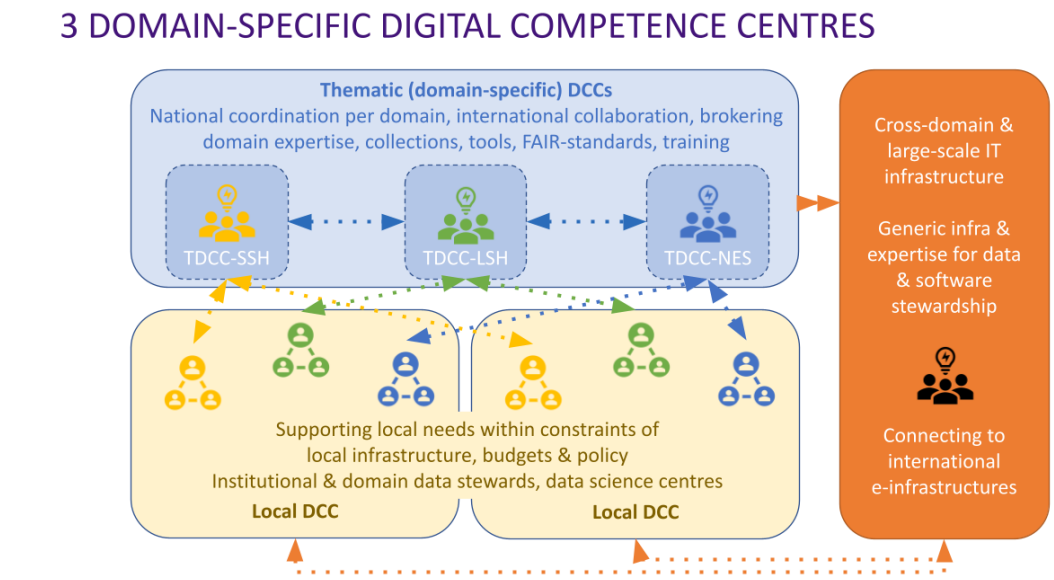
As noted in the report 'Topwetenschap vereist topinfrastructuur' from June 2017, for the Netherlands to have infrastructure to support competitive, high quality research in the 21st century, investment not only needs to increase, but also to change from project-funding to strategic, structural support. This implies that NWO should implement funding schemes that do not make the artificial distinction between "science" and "infrastructure". Rather, strategic funding should be provided for both aspects for any given scientific endeavour.

Local DCCs

To facilitate building of digital competencies across knowledge institutions, NWO has provided impulse financing for the creation of local Digital Competence Centers at universities, university medical centres, universities of applied sciences, as well at NWO and KNAW institutes (see the report: Integrale aanpak voor digitalisering in de wetenschap - Uitvoeringsplan investeringen digitale onderzoeksinfrastructuur). DCCs have been created at knowledge institutions as the centres of expertise on data and software. The impulse financing allowed institutions to appoint data stewards and Research Software Engineers and to increase collaboration and coordination for building digital competencies.

Most of the local DCCs (perhaps with the exception of DCCs at university medical centres) are not domain specific. Rather, they provide support for researchers from all research disciplines and domains practised at knowledge institutions. In contrast to local DCCs, the role of the thematic DCC would be to connect and support domain-specific research from across local DCCs. As such, thematic DCCs will offer support complementary to all the services and support already offered by the local DCCs (see Figure 1). There will be an active collaboration and coordination between the local and thematic DCCs to establish effective ways of cooperation.

As such, the TDCCs will alleviate the work pressure on local DCCs by addressing the issues which cross institutional borders and are better addressed in disciplinary (cross-institutional) settings (e.g. metadata standards).



Copyright and credit: Ruben Kok

Fig. 1 Diagram explaining collaboration between thematic and local DCCs, as well as connection to infrastructure and generic sources of expertise and support for data and software stewardship. Credit and copyright Ruben Kok.

International landscape

Most if not all research in the domain is inherently international in nature, including many of its digital activities and community structures. On the one hand this originates from the scientific need of using instrumentation and facilities of which there are only one or few in the world – resulting in global communities that organise datasets, software, their meta-data structures, repositories, and interoperability standards. Other projects by themselves collect data from international sources (such as oceanographic and climate data or data from earth observation satellites) and thus build on the datasets and software of global communities to answer new research questions. In both cases, international standards and structures define to a large extent which digital competences are required and available also in the Dutch context, and any national activities in this area should be precisely complementary to these international endeavours.

Divergence between national and global – including European efforts – must be avoided, as any such differences in metadata schemata, preferential data depositions, identifier schemas, or software repositories, would fragment the community. At the same time, these international structures can also have a positive, structuring effect on adjacent research fields (including on those where there is specific national interest): successful standards from one domain can be applied with no or only small adaptations bringing a structuring effect.

The European landscape features a prominent role for the clusters of ESFRI projects, of which many are directly related to the Natural and Engineering Sciences (NES) domain: ENVRI, PANOSC, ESCAPE, EPOS ERIC, ACTRIS, IAGOS, ICOS ERIC, EISCAT_3D, IS-ENES, JIVE ERIC. Besides ESFRI the Copernicus program EU's Earth observation programme is specifically important for earth sciences within the NES domain. Although being service oriented, research plays an important role in usage and further development of Copernicus. All have a strong structuring effect on the use of digital resources (datasets, software, and their descriptions, but also on storage, computing, and networking) in their respective fields of Environmental and Earth Sciences, materials science, and astronomy, astroparticle & particle physics. All of these have 'national' equivalents in the National Roadmap Facilities, with which strong relations exist.

The national thematic DCC activities should complement this by stimulating contribution to this 'ecosystem' of digital services, data, and software – but they should also leverage the work that is being done in the international context and ensure its uptake and adoption within the national context also in fields that are adjacent to the large 'ESFRI' infrastructures and national Roadmap Facilities, but where the standards are equally applicable.

In this ecosystem, both the European Open Science Cloud (EOSC) as well as the (global) Research Data Alliance (RDA) shape the space in which the NES thematic DCC can operate. Where the RDA provides suitable standards for data (such as data formats, interoperable descriptions, and standards for data exchange), the EOSC is pushing for *'an inclusive and federated ecosystem based on FAIR data and other open science outputs, integrating many services such as data visualisation, analysis and physical resources to store and re-use data for open science*'[2]. The EOSC Association has five advisory groups, each consisting of a few task groups. These task groups are charged with identifying strategic gaps and areas for investment which are used for feedback for the development of the European Strategic Research and Innovation Agenda. The task groups focusing on Metadata and Data Quality, as well as Research Careers and Curricula and Infrastructure for Quality Research Software are of particular relevance to the NES domain.

Within the European context, the CESAER network of over 50 research-intensive universities of science and technology in Europe is an important organisation supporting the knowledge dissemination and networking within the NES domain. For example, the CESAER's task force Open Science has published a white paper

[2] Joint Position Paper: *EOSC – a tool for enabling open science in Europe*, by the ESFRI science clusters ENVRI-FAIR, EOSC-LIFE, ESCAPE, PaNOSC, SSHOC, and the European e-infrastructures EGI , EUDAT, GÉANT, and OpenAIRE*;* https://doi.org/10.5281/zenodo.4044009

"Advancing Research Data Management in Universities of Science and Technology"[3], which identifies challenges related to research data and software management, which are specific to the universities of science and technology. The main challenges identified are related to dealing with IPR and data ownership issues, lack of hands-on support for data and software management, insufficient recognition of the role of software, lack of metadata standards and few incentives for researchers to invest in improving their data and software management practices. While some of these challenges are shared also with the two other science domains, some of them, such as lack of metadata standards, is something which could be addressed by a thematic DCC.

The thematic DCC links to these developments both by supporting the inclusive and 'resource-rich' vision of the EOSC, and by ensuring that access to and knowledge of these international developments, such as the EOSC, but equally from RDA for interoperable data standards and from European and global facilities such as the EIROforum organisations (CERN, ESO, ESA, ESRF, etc) and international partners like the (US) NSF, DoE, and NASA agencies) is made more readily available to national research. The National Roadmap Infrastructures, already discussed prior, are a structuring element in making such interaction fruitful and creating sufficient mass for a visible and meaningful Dutch contribution (e.g. through the Roadmap Digital Specialists group).

2. Summary of Challenges

The main challenge within the domain of natural and engineering sciences with regards to digital competencies is large fragmentation and diversity in the degree of implementation of reproducible and FAIR research practices. This challenge can be broken down into several more specific challenges:

- Need for improvement of digital competencies within the domain and increased training capacity
- Fragmented landscape - insufficient connections between research communities and existing infrastructures, tools, services and experts (e.g. RDA)
- Insufficient interoperability of data and of workflows: there is a need for improved FAIRness of data, especially the 'I' and 'R' aspects, as well as a need to integrate workflows to put FAIR into practice and bring data to 'life'
- Need for improvement of computational reproducibility practices (e.g. implementation of provenance standards and agreement on the level of detail)
- Insufficient support for long-term preservation and sustained maintenance of research software
- Lack of clarity and guidance about dealing with IP and legal issues
- Insufficient collaboration between disciplines across the domain which could facilitate the adoption of FAIR practices

Some of these challenges, especially those related to software, are shared across the whole of science but are particularly salient in the NES domain where most of the software development experts and methods currently come from. A set of specific challenges are listed below, for the moment in no particular order (i.e. the numbering does not correspond to priority). These challenges were identified by the writing team of this roadmap and are not intended as frozen. Once the TDCC is in place its board will be free to revise the set of challenges.

2.1. FAIR data

Despite numerous funders', publishers' and institutional policies obliging researchers to share research data, the uptake of FAIR data practices across the domain of exact and technical sciences greatly varies across disciplines. Disciplines such as high energy physics or astronomy have largely adopted FAIR practices before the term 'FAIR' came into being. In software research, open source has been the standard for decades. The successful adoption of FAIR practices within these disciplines has been largely driven by the need to collaborate at an international level. The huge amount of data generated by international research infrastructures meant that standardisation and data sharing were essential for effective collaboration. However, other disciplines within the exact and technical sciences still struggle even with the implementation of the "F" (Findability) and "A" (Accessibility) of FAIR. For example, in condensed matter physics, which is a discipline related to and

[3] Mattias Björnmalm, Federica Cappelluti, Alastair Dunning, Dana Gheorghe, Malgorzata Zofia Goraczek, Daniela Hausen, Sibylle Hermann, Angelina Kraft, Paula Martinez Lavanchy, Tudor Prisecaru, Barbara Sánchez, & Robert Strötgen. (2020). Advancing Research Data Management in Universities of Science and Technology. Zenodo. https://doi.org/10.5281/zenodo.3665372

sharing some of the methods and concepts with high energy physics, there are no agreed standards for sharing research data and software[4]. In engineering, and everywhere where research tends to be more applied, the challenges in FAIR implementation are even larger due to the cross-disciplinary nature of research and applications, but also due to frequent problems with balancing the desire for sharing and openness with the need to protect commercially-confidential research outputs.

Altogether, despite the fact that some disciplines within the domain of exact and technical sciences offer best examples of how FAIR could be put into practice to the benefit of the research community as a whole, the uptake of FAIR practices across the domain remains low. A recent survey conducted by the LCRDM has identified 'awareness' as the main challenge to the implementation of FAIR principles. This uptake could be stimulated by providing tools and training that would make it easy to bring FAIR data into practice, or by facilitating collaboration across disciplines for mutual learning.

Improving the uptake of FAIR practices in the NES domain will start from building on existing tools and training developed within local DCCs.

2.2. Sustainable software and eScience

Science as a whole experiences substantial challenges with regards to long-term preservation of research software. Unfortunately, despite recent studies highlighting the importance of long-term sustainability of software[5], among both researchers and research support groups there is insufficient expertise in, and insufficient knowledge of the requirements of, sustainable research software as well as insufficient support to facilitate its long-term preservation (including hosting and maintenance). The eScience Center is focusing its support on training on working with software and collaborating with researchers on software development projects, which are essential and contribute to ensuring that software which is developed is suitable for long-term use, reuse, adoption and preservation.

However, across domains, insufficient structural resources are available to maintain the software, or dynamically respond to changes in its ecosystem or usage. The software life cycle management is an issue which needs to be addressed for long-term preservation.

- Which software is worth preserving or sustaining?
- Can old software be modernised and made "future proof"?
- How do we deal with dependencies and security issues?
- What is already done in the international context and how can we contribute?

Making software sustainable in part implies addressing the challenges of the licence chosen for any given software, as well as the licences of components it depends on, early in the development cycle. This issue is closely related to the question of intellectual property protection. The main goal here is to make sure that licensing issues do not become a blocking point for any intended use of the software. The wrong licence may hinder the use of the software (if licences of dependencies clash), may hinder its integration in other open source projects, or exclude possible commercialisation (the TDCC will of course play no role in decisions on code publication or commercialization). Tools and training to manage licences and IP will be developed and provided by the TDCC, again building on existing material already available within local DCCs and elsewhere (such as for example the licence checker under development by the eScience Center).

We note that sustainable software requires structural support (i.e. people). When software has to be maintained, the effort needed to do so increases over time. The TDCC will play a role in more clearly formulating the sustainability problem, bundle and propagate the expertise present within the domain, and bring it to the attention of funding agencies as something that should be supported structurally across domains.

---

4 https://www.nature.com/articles/d41586-021-00954-8

5 https://doi.org/10.5281/zenodo.4543569

2.3. Connections to the international activities

As sketched above, there is a lot happening in international projects and it will be important that any TDCC efforts and initiatives take the international landscape into account. Specific challenges to address:

- Monitor international activities and make the information available to NES researchers.
- Take the above information into account when planning TDCC projects, avoid repeating existing/planned international efforts.
- Identify which NL groups are involved in the international activities and what expertise they might contribute (or are contributing) to TDCC activities.
- Identify best practices in international activities and how these can be leveraged at the NL level.
- Identify expertise available in NL which could be contributed to international activities.

2.4. Long term data archiving

Data collected or generated in the context of for example climate research or astronomy remain of value decades to centuries into the future. This presents the challenge of how to archive such data so that it can be re-used straightforwardly in the future. Questions to address:

- Which data is worth preserving and which data can be easily recomputed, regenerated or resimulated in the future?
- What software should be preserved along with the data in order to keep the data readable/interpretable? This ties into the sustainable software challenge above.
- Can old data be modernised and made “future proof”?
- What is already done in the international context and how can we contribute?

There are many parties tackling these issues but the effort needs to be guided by the needs of the research community. Advice from the community can only be provided in a useful way if the researchers have a good understanding of the challenges and existing or planned solutions. The TDCC will take the lead in collecting the archiving and sharing needs, the existing or planned solutions, and then develop training for researchers.

Addressing this challenge does not imply that the archiving necessarily takes place at a Dutch facility (although this is not excluded of course). Instead the aim is to provide NL researchers with the knowledge needed to, for example, contribute to discussions within international collaborations on the archiving and data sharing questions.

2.5. Locate computing capacity close to the storage

There are initiatives to “bring the processing” to the data in many fields. For example the availability of a Jupyter notebook (and relevant data processing/analysis libraries) attached to a data archive, allowing for fast and efficient data processing and/or data exploration. Such solutions are relevant if one wants to explore interactively very large data sets (such as from the ESA Gaia mission which provides complex astrophysical information in the form of many 100s of distinct data items on about 2 billion sources). If properly set up, such a facility would also enable FAIR as a natural part of the process. The TDCC will play a role in increasing the expertise among researchers so that (as in the previous challenge) they can properly steer any initiatives in their own field.

Again, addressing this challenge does not imply that the implementation has to happen at a Dutch facility.

2.6. Human capital

As noted in the report ‘Topwetenschap vereist topinfrastructuur’ from June 2017, there is an overall need to improve the ‘digital literacy of all researchers’ and to ‘stimulate local support [...] to assist researchers in understanding their digital needs and options’. Both require investment in training and capacity building, both nationally (within the thematic DCC), as well as locally, within institutional DCCs and individual research groups.

In addition more people are needed who can provide all the support functions (IT, library, software engineering, etc). While increasing the digital literacy of all researchers is useful, they cannot be expected to provide support functions in parallel to their scientific careers (at least not without a substantial change in the way one's work is recognized and rewarded). There is an urgent need to increase the capacity of expert support professionals such as data stewards and research software engineers (RSEs) who could join research teams and provide hands-on support with research data management and sustainable software management practices.

While the NES TDCC recognises the acute need for sustainable appointment of data stewards and RSEs in order to facilitate the implementation of FAIR data and FAIR software practices, the budget allocation for TDCCs is insufficient to drive this change and the TDCC is not a funding body (is not in a position to fund capacity building). Sustainable increase of these capacities has to be supported by structural funding. That said, the NES TDCC is committed to advocate for the recognition and rewarding of essential professionals such as data stewards and RSEs.

2.7. Cross-field collaborations

Within the NES domain, the TDCCs will facilitate/stimulate collaborations across fields. For example the Einstein Telescope will not only observe gravitational waves but also collect and process lots of seismic data as a matter of course, e.g. during its design and commissioning. The TDCC will bring together physicists/astronomers/geophysicists to discuss commonalities in the collected data and how to leverage those. Another example is quantum research, where multiple research disciplines come together (such as electrical engineering, quantum physics, materials science) and where agreement on standards for reproducible research practices are needed. Again, TDCC will play a role in helping communities identify best practices and implement them in cross-disciplinary collaborations.
Across the three domains the NES TDCC will facilitate/stimulate software infrastructure and tools that improve research in all areas where large quantities of data or software play a role. The NES TDCC will contribute to the development of a community-owned and community-driven ecosystem of digital services by identifying and stimulating the tools used to create such services, such as database systems design or programming toolboxes.

3. Role of the NES Thematic DCC

The NES Thematic DCC, similarly to other thematic DCCs, is a network organisation. The NES thematic DCC will bring together researchers, local DCCs, infrastructure providers, knowledge and expertise networks, as well as other strategic initiatives active within the domain. The thematic DCCs complement the work of local DCCs by addressing issues which are discipline- and domain-specific (e.g. metadata standards), which cross institutional borders and are more effectively addressed in collaboration.

The NES thematic DCC aids this existing mesh of experts, organisations, and structures to connect research communities to digital infrastructures for research, to link with the Roadmap Facility experts and 'digital side' of the European Research Infrastructures and ESFRI clusters, and to ease communications through joint activities, collaborations, and supporting projects. Together, these build a network where the TDCC is supportive, but – intentionally – not the deciding factor in itself. Decisions rest in the domain, through the contributions of its members that collectively shape the TDCC joint work programme. The TDCC facilitates and supports, giving a push here and there to make computing, software, and data work better together.

The role and the main tasks of the NES Thematic DCC are:

- Collaborating with the local DCCs to achieve 'economies of scale' in identifying, connecting and supporting the needs of domain-specific research, where solutions to problems extend beyond institutional (and national) borders (e.g. interoperability standards, metadata).
- Helping research communities connect to existing infrastructures, tools, services and expertise networks to stimulate mutual learning and develop a more coherent ecosystem of digital services, data and software across the domain.
- Connecting national activities and endeavours to international developments (EOSC, RDA, global facilities and international partners) and leverage the work done internationally.
- Consolidating the field by stimulating collaboration and exchange of best practices across disciplines.
- Identifying joint needs (versus needs specifically related to a single discipline) and translating them into project proposals submitted to NWO.
- Increasing the expertise among researchers so that they can properly leverage national initiatives in international collaborations.
- Based on domain-wide observations, providing continuous feedback to NWO on research support needs that could help inform NWO's (future) resources distribution on digital competence topics.
- Facilitating the connections and collaboration with the two other thematic DCCs to establish effective ways of supporting cross-disciplinary research.
- Continuously monitoring the challenges related to digital competencies within the domain and proposing strategies to address them within the programme of activities.

4. Investment plans: network organisation

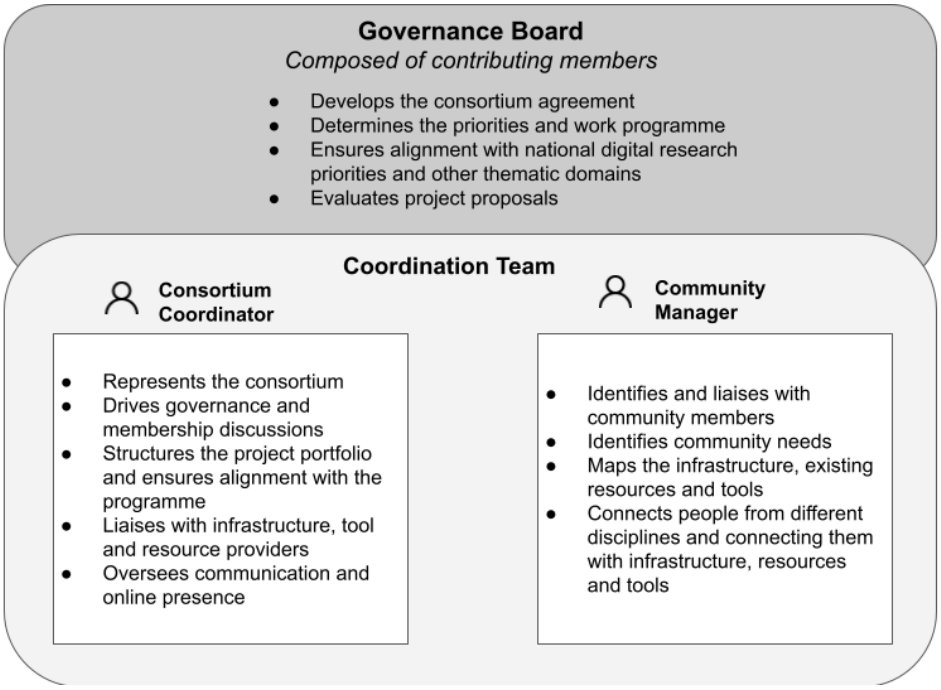
Description of organisation and activities

The thematic DCC for NES brings together more than data: it builds a network of organisations and people. It is a node in a federated network that augments the existing entities along its thematic line - forging links and promoting exchange of tooling, innovation, and ideas from across the NES domain, between the local LDCCs, SURF, the eScience Center, the GWI Roadmap digital experts, ESFRI infrastructures and clusters, and the institutional digital support teams and research software engineers. It will be organisationally lean: resources are available to support two people in a network hub, and it will both rely on the existing distributed expertise as well as encourage collaboration through targeted projects that together build a common 'NES TDCC' work programme.

The NES TDCC consists of a coordination team, and of a governance board that sets the thematic priorities and defines the long-term work programme of the TDCC. The coordination team is responsible for ensuring coherence of the consortium activities, alignment and supervision of the projects within the TDCC, facilitating

the connections between the research communities, institutional DCCs, and national and international bodies. The coordination team comprises at least a consortium coordinator and a community manager (Fig. 2), bringing with them a thorough understanding of research practice.

Fig.2. Diagram outlining the structure and responsibilities of the NES TDCC Governance Board and the Coordination Team.



The coordination team will meet regularly with the coordination teams of the two other thematic DCCs to align the work programme and other collaboration areas and to avoid duplication of efforts. Moreover, to ensure coherence of communications, facilitate the findability of information, as well as alignment of activities, resources of all three thematic DCCs will be findable through a shared website/portal.

One of the first tasks of the coordination teams will be to create an overview of all communication channels and stakeholders which have to be involved for the successful implementation of the NES thematic DCC.

The *coordinator* represents the consortium, and takes care of ensuring continued engagement of the consortium members and the institutional DCCs, and monitoring progress of the ongoing projects, and maintaining links with infrastructure, tool, and resource providers about the domain needs (e.g. long-term software preservation, interoperability standards, big data storage, and the tools and infrastructure for data preservation and reproducibility). The coordinator also plays a key role as the linking pin between the institutional and thematic DCCs.

The *community manager* has a similar role, in the heterogeneous and diverse domain of the natural and engineering sciences. The community manager engages directly with research groups and domain specialists, identifies existing and emerging solutions that can enhance the digital competence of research domains, and supports research groups and institutions in connecting with peers where collaborations could be established. The community manager also supports research groups and institutional DCCs in formulating joint challenges that can be turned into concrete projects. The main mechanisms used by the community manager include organisation of workshops, training, and in-person information sharing (e.g. through workshops).

The *governance board* plays a role not only in ensuring the TDCC remains aligned with the national digital research priorities, but also uses its broad composition and background to drive the TDCC work programme from which concrete projects derive. The TDCC will establish a rolling work programme that sets the priorities for the upcoming two-year period, and which forms the basis for inviting concrete proposals to address the

identified bottlenecks. The composition of the NES TDCC governance board will reflect the diversity of the field, and include expertise and infrastructure organisations, knowledge institutes, institutional DCCs, as well as explicitly stay aligned with the peer TDCCs in the LSH and SSH domains. The work programme prioritises projects that support 'horizontal' digital competences (i.e. where the NES domain can both support and benefit from concurrent activities in the LSH and SSH domains) and projects with outcomes applicable across multiple disciplines within the NES domain. The consortium coordinator at the TDCC will support the governance board in ensuring representative, domain-wide, input into the work programme through periodic stakeholder consultation (with the institutional DCCs and e.g. the boards of research organisations and universities).

The NES TDCC governance board is also responsible for evaluating the projects submitted to the NES TDCC for funding - following a formal eligibility check by NWO. For this, the governance board can seek input from external expert reviewers, and the process will emphasise effectiveness and timeliness in awarding. The key metric for project proposals is their structuring impact on the field, and the extent to which they engage multiple scientific disciplines. The projects must fit the work programme of the NES TDCC, and pro-actively engage with the existing digital infrastructures for research in the Netherlands or abroad in order not to duplicate work. No organisation or consortium is a-priori excluded from applying for project funding from the NES TDCC: it is embedded in the domain, through this governing board with all contributing members This board is also the principal group responsible for the strategic evolution of the TDCC, endorsing the outlook of the future needs and challenges of the domain based on the periodic stakeholder consultations.

Although the 'project' concept could give the impression that each activity has a defined end-date, it is envisioned that some activities have such a strong structuring impact on the NES community as a whole, that any project end-date should be seen more as an evaluation moment rather than as a terminal point: successful projects that have such a structuring impact should be extended, and their awarding prioritised in the work programme of the NES TDCC. This includes the continuation of projects identified at the time of establishment of the TDCC as 'bottleneck projects'.

The NES TDCC governance board is appointed by all consortium partners contributing materially to the NES TDCC (e.g. through co-funding of activities), and takes into account the recommendations of the coordinators of the institutional DCCs, the relevant NWO domains, national infrastructure organisations and expertise centres (such as the Netherlands eScienceCenter, SURF, 4TU.ResearchData) and the research organisations and universities. Its composition thereby reflects the diverse nature of the NES domain, and it will choose its convener on a rotating basis.

Main hub of organisation

The NES Thematic DCC will be a consortium organisation, governed by consortium partners. Consortium partners will develop and agree on a consortium agreement, which will set out the rights and obligations of the consortium members.
Given the stated wide diversity of the NES domain, no single thematic organisation exists within the Netherlands that on its own targets this broad range of disciplines - whereas for the 'SSH' and 'LSH' domains such an organisation is more easily identified. Yet a successful NES TDCC has to have an equally vibrant, action-oriented, and firmly established home base, which is embedded and respected in the Dutch digital landscape. 4TU.ResearchData (TU Delft)[6] is willing to act as the legal entity of the consortium.
4TU.ResearchData is a data and software repository for researchers from science, engineering and design domains. It is at the moment the only organisation related to digital competencies that connects engineering research, whose activities touch almost all of the natural sciences, and whose strategic programme specifically aims for an open partnership approach targeting the domain of science, engineering, and design. 4TU.ResearchData already brings along a lively and diverse community of researchers and research professionals which are at the forefront of advancing the implementation of FAIR practices. The experience in community building, knowledge of successfully running a multi-institutional consortium organisation, as well as the domain relevance, make 4TU.ResearchData an ideal candidate to act as the legal entity for the NES Thematic DCC.

[6] TU Delft is the legal entity behind 4TU.ResearchData

No other Dutch research organisations that are scoped specifically to science or engineering have been identified that are cross-disciplinary, have such a well-established background in digital repository and accessibility activities, and would be capable and willing to take on this role.

Consortium

Although the consortium will have an enumerable set of partners that materially contribute to the TDCC, the mission of the TDCC is to support digital competences throughout the natural and engineering sciences. This applies both for the activities of the coordination team as well as for the awarded projects. The coordination, community management, and training activities of the NES TDCC are therefore open to all Dutch researchers in the domain as well as their international collaborators, and are supplied on the same terms to all.

Materially contributing partners complement the TDCC by assigning experts to work on joint NES TDCC activities, either in the coordination function of the TDCC or as part of specific projects. The full cost of these experts, regardless of the original source of funding for these experts, will constitute the co-funding (matching) of the effort that is granted by NWO to the TDCC, as long as their activities support or directly implement the agreed work programme of the TDCC and are cross-institutional in scope and impact.

The materially contributing partners in the TDCC also designate the members of the governance board, and thus define the long-term work programme of the TDCC.

Any formal (NWO) co-funding requirements on the TDCC are to be calculated over the ensemble of the coordination team and all running projects, i.e., surplus co-financing available for the coordination team may be used to alleviate the need for co-financing at the project level, which is specifically relevant for projects that have members with necessary specific expertise and capabilities that are not already part of the NES TDCC and that cannot customarily accommodate co-funding requirements.

5. Investment plans: bottleneck projects

There are several issues which have to be tackled first in order to address the domain challenges outlined in chapter 2. These will be addressed through so-called bottleneck projects:

1. Community building and establishing online presence
2. Creating a Training Hub
3. Facilitating long-term software preservation and sustainability
4. Increasing interoperability and integrating tools and workflows for FAIR

Each of these projects is briefly explained below.

The bottleneck projects are not aimed to address *all* the challenges listed in chapter 2. They provide the mould for the next decade. The TDCC is set up to evolve in line with the needs of the domain, driven by its own open governance structure - however challenging that may be given its scope and diversity. The evolving work programme implements the strategic outlook and focus on emergent and future needs by defined pre-agreed project targets to ensure coherence of the results within this common work programme.

The bottleneck projects listed below are the most urgent needs within the domain, as identified during stakeholder consultations. The list below is not exhaustive and not all challenges will be addressed from the start (the remaining challenges will be addressed through project funding and regular calls).

5.1. Community building and building the national expertise map and thematic survey

This bottleneck project is directly addressing the issues described in challenge 2.7. Cross field collaborations.

The goal of the community building is to facilitate collaboration and exchange between researchers from various disciplines and connecting them with tools and infrastructure providers, as well as expertise centres in order to improve FAIR practices. The community manager within the NES thematic DCC will take the lead on implementing this work. Facilitating the community building process requires both establishing an online presence (where appropriate collectively for all TDCCs), and the organisation of networking workshops and events.

Online presence will consist of a platform (a website or tool) for the NES thematic DCC, where community members could find one another (a 'national expertise map') and to interact with one another. But also infrastructure itself has to be included in this thematic survey, complementing the service portfolio activities of the infrastructures and existing organisations: making the services themselves 'FAIR'. This will also be the place to find information about the upcoming workshops and events.
The funding requested will be needed to compose the national NES expertise map and identify the relevant capabilities and services, and to develop a community platform (a website or tool) to make this information itself findable and accessible. Making information about the existing resources, expertise and services already available will also help to avoid duplication of efforts and help manage the limited resources available to the TDCC.

Networking workshops and events will help researchers and other stakeholders within the domain to find one another and to start collaborating. The topics of networking workshops and events will be directly addressing the community needs, and can focus on areas such as standardisation and interoperability, workflows and tools for FAIR, computational reproducibility etc. The topics will be agreed in consultation with the community, including the local DCCs.
The funding requested will be needed to support the work of the events and communication coordinator to organise the events, as well as to contribute to the running costs of the events (venue, catering, AV, technical support, video and photo production etc).

5.2. Creating a Training Hub

This bottleneck project addresses challenges described in point 2.1 FAIR data, 2.2. Sustainable software and eScience and 2.6 Human capital

One of the biggest roadblocks to improvement of FAIR practices within the NES research communities is large fragmentation with regards to skills required to design and implement workflows for effective data processing, computing, computational reproducibility and to work with and produce open, sustainable and robust research software (FAIR software). Skills are necessary to put the principles into practice (see Six Recommendations for Implementation of FAIR Practice[7]). To facilitate the uptake of digital competencies within the NES domain and to create a Training Hub we need a:

- Training programme for the NES domain
- Coordination with existing training networks and communities

Training programme for the NES domain will focus on developing and delivering training on computational reproducibility and managing data, software and computational workflows, as well as associated IP aspects within the NES domain, including discipline-specific data carpentry workshops. A training programme will be developed, discussed and agreed with relevant training networks and initiatives in the field, such as RDNL, Research Software Training Network, DCC Implementation Network (representing local DCCs), LCRDM, NPOS, etc.. The role of the network will be to ensure complementarity, facilitate knowledge and experience exchange between trainers and to avoid duplication of efforts.

Funding will be provided to support network coordination efforts (people costs, but also the costs of organisation of meetings and workshops), as well as the delivery of the training programme for the NES domain (external trainers and workshops - e.g. eScienceCenter, RDNL, as well as trainers appointed by the NES thematic DCC to deliver workshops and courses).

5.3. Facilitating long-term software preservation

This bottleneck project directly addresses the issues described in chapter 2.2. Sustainable software and eScience

A dedicated project will be launched to facilitate long-term preservation and reusability of software (the issue of the quality of the software is not addressed here). The need to improve software management practices to make software more suitable for long-term preservation and re-use is a significant challenge within the NES domain (see chapter 2). Therefore, it requires dedicated attention and investment to be effectively tackled. Ensuring that software is sustainable in the long-term also requires a lot of time and resource investment. As outlined in the report "Research software sustainability in the Netherlands: Current practices and recommendations"[8]: "*not all software is stable or suitable enough to be reused and maintained. Some software will only be preserved for reproducing research. Researchers should therefore carefully consider which of the research software they develop is suitable for reuse by others, and which part may simply be shared or archived.*"

To tackle these issues, the following streams of work will be pursued:

- Development of guidance for researchers on research software quality and sustainability. This will include advice on curating software for sustainability, and developing an efficient plan for the long term.

[7] 'Six Recommendations for Implementation of FAIR Practice' - report by the FAIR in Practice Task Force of the European Open Science Cloud FAIR Working Group, October 2020, doi: 10.2777/986252

[8] Saskia van Eeuwijk, Tom Bakker, Maria Cruz, Vera Sarkol, Barbara Vreede, Bart Aben, Patrick Aerts, Gerard Coen, Bob van Dijk, Peter Hinrich, Lena Karvovskaya, Meta Keijzer-de Ruijter, Jacko Koster, Jason Maassen, Miriam Roelofs, Jan Rijnders, Alfons Schroten, Laurents Sesink, Chris van der Togt, … Petri de Willigen. (2021). Research software sustainability in the Netherlands: Current practices and recommendations. Zenodo. https://doi.org/10.5281/zenodo.4543569

- Provision of specific courses and instructional materials in areas such as software publishing (archive & repositories), software citation and software licensing & copyright.
- Provision of consultancy support for researchers who wish to improve the sustainability of their research software.

All streams of work will be conducted in collaboration with the eScience Center, the institutional DCCs, as well as with the other thematic DCCs.
The funding will be used to fund capacity within the NES thematic DCC, and for expert support of the eScience Center.

5.4. Support for metadata standards and vocabularies for FAIR data

This bottleneck project directly addresses the challenges described in chapter 2.1 FAIR data.

As indicated in Chapter 2, one of the key bottlenecks to improving FAIR practices within the research communities is the lack of interoperability (insufficient use of/absence of standards). However, as indicated in the aforementioned report *Six Recommendations for Implementation of FAIR Practice* for successful adoption of FAIR practices, it is essential that such efforts are community-driven. It is therefore essential to 'fund development, adoption, and maintenance of community standards, tools and infrastructure' at a community level. As demonstrated by the extremely successful Open Science Fund organised by the NWO, smaller grants awarded at a community level can stimulate the community to actively engage with open science practices[9]. A similar funding scheme will be used to increase the interoperability and to integrate tools and workflows for FAIR within the NES domain. Specifically, we will fund efforts aimed at supporting the development/implementation of metadata standards and vocabularies across the disciplines within the NES domain or services focused on the improvement of services/interoperability of specific file formats (e.g. Hdf5).

Competitive funding calls will be launched to stimulate community initiatives which will aim at:
- Connecting with existing expertise (e.g. RDA),
- Adapting or developing interoperability standards,
- Consolidating the community and networking activities facilitating the adoption/development of metadata standards.

Funding will be used not only to fund the awarded projects, but also to provide administrative support for project review and administration, remuneration of the reviewers and to promote case studies, tools and workflows resulting from these projects.

[9] Funders need to credit open science, Hans de Jonge , Maria Cruz & Stephanie Holst,https://doi.org/10.1038/d41586-021-03418-1

Addendum - Consultation approach for the NES thematic DCC

The NES-domain is large and diverse: not only with regards to the multitude of disciplines covered within the domain (from astronomy to civil engineering), but also with regards to its fragmentation and diverse needs for data and software within the domain.

The writing team of the NES-TDCC was tasked with identifying challenges which are common to the entire domain and also with proposing a way for effectively tackling these challenges. Given the limited time for preparing the roadmap (September - December 2021), the writing team, supported by the NWO, had to be very strategic in ensuring that the needs of a broad stakeholder group are taken into account in the proposal.

This has been addressed at three different levels:

- Composition of the writing team
- Consultations during Roadmap preparation
- Consultations on the draft Roadmap

Composition of the writing team

The initial composition of the writing team, as suggested by NWO, already ensured representation from both natural and engineering sciences. David Groep from Nikhef represented subatomic physics (natural sciences), Anthony Brown from the University of Leiden represented astronomy (natural sciences) and Marta Teperek from 4TU.ResearchData represented the technical sciences.

To better reflect the diversity of the domain of exact and natural sciences and diversify the perspectives, four more members have been invited to join the writing team: Irene Dedoussi from TU Delft (engineering), Vera Sarkol from CWI (natural sciences) and Wim Som de Cerff from KNMI (natural sciences). The fourth invited member was from the University of Twente (engineering), but was unable to join the writing team due to the time constraints.

Consultations during the Roadmap preparation

During the process of Roadmap drafting, the writing team, as well as NWO, took a proactive stance for seeking input from the broader community. Input from members of the 4TU.Federation (technical sciences) was gathered regularly and at various levels (university rectors, faculty deans and research directors, library directors, individual researchers and data stewards). Feedback from the astronomers and KNMI was ensured through direct representation in the writing team. There were also several consultations organised with the NWO institutes.

In addition, to reach out to all the research performing organisations with natural or engineering sciences faculties, a dedicated consultation was organised to which representatives from all local DCCs were invited. During this session, the writing team presented the vision for the Roadmap and received feedback from the participants, which directly shaped the Roadmap.

Given that the thematic DCCs are network organisations, they need to rely on infrastructure provided elsewhere. Therefore, the writing team had several bilateral conversations with colleagues from SURF. In addition, members of the writing team spoke on several occasions with colleagues from the eScience Center to

align on support for sustainable research software and also training involving research software and computational reproducibility.

Finally, given the industry component particularly prominent within engineering sciences, the topic of the NES thematic DCC was also discussed in a workshop on FAIR data for the TO2 group organised by the EZK Ministry.

In addition to domain specific consultations, there were also ongoing alignment meetings between the writing teams of all three thematic DCCs. Updates on the thematic DCC process were also regularly shared with members of the NPOS FAIR Data Tafel.

Consultations on the draft Roadmap

In order to ensure comprehensive and in-depth consultation on the Roadmap itself, the team, supported by NWO, has gathered a list of 46 advisors who were asked to provide direct input on the Roadmap in January 2022.

Among the advisors were representatives of researchers from a multitude of disciplines in the natural and engineering sciences, data stewards, directors and innovation leaders of various national and domain specific infrastructures and service providers (including large research infrastructures), as well as experts on research software.

The list was gathered by members of the writing team who directly requested the research performing organisations and infrastructure providers names of their representatives who could provide feedback on the Roadmap. In addition, colleagues from NWO have approached all the deans of science faculties in the Netherlands as well as the seven disciplinary consultation committees (‘Round Tables’) within the NWO Domain Science for additional name suggestions. Finally, some advisors pro-actively approached NWO and asked to be added to the list.

All advisors have received the draft Roadmap and were invited to provide their input on the Roadmap in writing and/or by attending dedicated online consultation meetings.

The comments that could be addressed within the scope and publication time frame have been incorporated into this revised version of the Roadmap. All comments that could not be addressed given the timeline, will be addressed by the TDCC NES through its work programme. Comments that were important, but out of scope for the TDCC NES, have been highlighted as such within the Roadmap.