Roadmap Thematic Digital Competence Center

Domain Social Sciences and Humanities

Lead authors

Susan Aasman
Henk Wals

Contributing authors

-

NWO secretaries

Alice Dijkstra

Version 1.1, June 2022

1. Summary of Landscape Analysis

Various reports describe the landscape (see the list in the Annex). These analyses provide a good and up-to-date view of the current state of the Dutch research landscape in general and the social sciences and humanities (SSH) in particular. Especially for the humanities, several specific characteristics can be distinguished that mainly emerge from their connection with the heritage domain. Our analysis of the landscape is also partly based on our own experience and involvement with research institutes, research funding bodies, universities, service providers, as well as research and heritage infrastructures and (Roadmap LSI) projects that contributed to this.

For the landscape analysis, we made use of various sources, including an inventory produced by DANS last year that formed the start of a permanent Dutch Open Science Dashboard. In 2022, DANS plans to produce a more detailed monitor specifically for the SSH in order to optimally support the thematic Digital Competence Center Social Sciences and Humanities (TDCC SSH). For example, it is important to know the current position and progress with respect to the percentage of research data available in FAIR form within the domain.

Based on an initial draft of the analysis and ideas detailed below, we held discussions with the Principal Investigators of the two large Research Infrastructures (RIs) in the SSH, CLARIAH and ODISSEI, and with the chair of the Platform Digital Infrastructure SSH. We also submitted the initial draft to a group of representatives from the local DCCs and a think tank put together by us made up of six SSH researchers interested in digital data use, most of whom are not involved in the aforementioned RIs.[1] Besides endorsement, this consultation round led to several useful additions to our proposals that we have incorporated in this document.

In a broader consultation round, we further assessed whether the analyses and plans contained in this document meet the wishes and needs present in the SSH domain. That consultation took place in January and February 2022.

Description of domain

The social sciences and humanities (SSH) domain is very broad and heterogeneous. In the Netherlands, at least 26 academic main disciplines can be distinguished (…) There are 52 SSH faculties at Dutch universities (see Annex), with a total capacity of 16,000 employees. This figure does not include the researchers appointed at the NWO institute Netherlands Institute for the Study of Crime and Law Enforcement (NSCR) nor at the nine Royal Netherlands Academy of Arts and Sciences (KNAW) research institutes. The number of BA and MA students who are registered for courses in the SSH domain has been rising for many years and is currently at almost 150,000 students. Accordingly, the SSH domain serves more than half of all students who are registered at universities.[2]

In the social sciences and humanities, most researchers now make wide use of digital infrastructures. In many cases, this simply concerns searching for and finding information in databanks such as those of Statistics Netherlands and digital corpora Delpher (in Dutch), WieWasWie or the notarial archive of the city of Amsterdam (in Dutch). A growing proportion of SSH researchers also makes use of advanced computational techniques to form, process, analyse, enrich and link complex datasets. Many of them are involved in one of the two large RIs CLARIAH and ODISSEI. Of the three domains, the SSH was the last to start the digital revolution. It can be expected that the use of digital methods and techniques will increase further in the near future. This argues for a review of the compartmentalization of the Roadmap funds, where the ratio between NES, LSH and SSH is currently 45-45-10%. In view of the rapid developments in the SSH and the ratio in student numbers, an equal distribution would be more reasonable, especially with respect to TDCC funding.

[1] The think tank members were: Prof. Anne Beaulieu (University of Groningen), Dr Gijs van Dijck (Maastricht University), Dr Marieke van Erp (KNAW Humanities Cluster), Dr Lidewijde de Jong (University of Groningen), Prof. Chantal Kemner (Utrecht University) and Dr Richard Zijdeman (International Institute of Social History).

[2] From the report of Prof. Mark Bovens, *Samen Sterker. Beeld van het SSH-domein* (in Dutch, 2019). Please refer to this report for further information.

A recent research report from NWO reveals that in both the social sciences and the humanities, although the vast majority of researchers consulted consider it important to share data, less than half of them regularly or always do that in practice.[3]

Institutions, organisations, infrastructures, public/private parties

The previously fragmented field is now relatively well organised. This is mainly due to the structural effect of a series of NWO investment projects. Two large roadmap infrastructures in the SSH, CLARIAH and ODISSEI, incorporated forerunners such as The Historical Sample of the Netherlands (HSN), CLARIN and Nederlab (in Dutch) and provide an infrastructure for advanced research in disciplines such as history, linguistics, media studies, literature studies and a large part of the social sciences. However, CLARIAH and ODISSEI do not cover the domain entirely: not all researchers from the aforementioned disciplines are involved, and disciplines such as archaeology, heritage studies, art history disciplines, psychology and law have so far remained outside of their scope. If the next LRI Roadmap proposal from a joint consortium associated to the SSH Group is awarded funding, then although the coverage will increase, it will still not be complete.

The SSH domain has strong institutes that have assumed responsibility for parts of the infrastructure. KNAW/NWO DANS and the KNAW Humanities Cluster (KNAW-HuC) emerged from the research field. houses about 160,000 datasets from the SSH. The KNAW-HuC accommodates large projects/datasets such as the Historical Sample of the Netherlands (HSN), Nederlab, REPUBLIC (in Dutch) and CLARIAH.

Furthermore, for the humanities, in particular, a large part of the digital infrastructure is provided from the heritage sector. Examples are the Koninklijke Bibliotheek (Royal National Library of the Netherlands), the National Archives, the Netherlands Institute for Sound and Vision, the Cultural Heritage Agency of the Netherlands and RKD Institute for Art History. NICAS and E-RIHS NL (in Dutch) work at the boundary of research and heritage. The heritage institutions are united in the Dutch Digital Heritage Network, which works closely with CLARIAH for technology development.

In the social sciences, institutions such as Statistics Netherlands, the Planning Agencies and CenterData provide important data sources. They are partners in ODISSEI. DANS is also a partner in ODISSEI and participates in CLARIAH. Another interesting data source is the data register of the Dutch government (in Dutch), which now has more than 20,000 datasets. The social sciences also intersect with the life sciences. For example, a research infrastructure such as Health-RI is also relevant for certain groups of social scientists, and vice versa SSH data is often relevant for life and medical sciences researchers.

Most universities offer their employees the possibility to deposit data in an institutional repository. University libraries play an important role in this. For this, most universities make use of DataverseNL, a service which was initiated at Utrecht University and is now offered by DANS.

For the storage of data and software, SSH researchers also make use of generic international services such as Figshare, Zenodo, GitHub, Open Science Framework (OSF) or Software Heritage.

For supercomputing facilities, researchers can turn to SURF and, in some cases, the universities. The Netherlands eScience Center supports researchers in developing and using innovative software for research.

Finally, we mention the Open Science Communities, which are organised bottom-up from the research field, with a broadening base at Dutch universities. As OSC's are mostly driven by young scientists who are motivated to support open science and the FAIRness of data and software, they offer a support base at the local workshops which is growing fast.. We think that the TDCC can benefit from their views and enthusiasm and that should be at the core of the TDCC networks.

[3] NWO Onderzoek Open Science (2021), 13 en 9.

All things considered, it can be stated that the SSH landscape is well populated with infrastructural activities and services.[4] Even without including the heritage sector, there are dozens of places within the SSH where data for research are available. Furthermore, CLARIAH and ODISSEI are working on portals that are making datasets easier to find. However, despite such positive developments, these infrastructures do not completely cover the domain, still largely work independently of each other, and many things still need to be done. At present, it is still not easy for researchers to find out what data are available or where they can best store their own data. Connections across the entire domain are still lacking.

Local DCCs

The local level is of the utmost importance and TDCC's cannot operate successfully without cooperating closely with the Local DCC's. All universities and many universities of applied sciences now have local DCCs in which data stewards offer researchers support. In principle, local DCCs are not organised in a domain-specific manner. Nevertheless, there are differences in how they are organised; at some universities, DCCs exist at the faculty level, for example, whereas in other cases, they are organised by a university library. The DCCs are largely still in their infancy. The required level of capacity is far from being reached. According to a National Programme Open Science (NPOS) report from 2021, considerable numbers are needed: 'The estimated numbers are dazzling', according to the authors.[5]
The Local DCCss received start-up funding from NWO in 2021. For additional funding of the local DCCs, institutions partly make use of two calls that emerged from the Implementation plan investments digital research infrastructure (in Dutch) and which were organised by SURF (in Dutch). The National Coordination Point Research Data Management (LRDCM) maintains an implementation network. It is currently not yet clear to which extent research institutions are able and willing to assume responsibility for the funding on a permanent basis and expand this. During our contact with the implementation network, it transpired that it is not possible to determine how many and which data stewards specifically focus on the SSH.

International landscape

Just like the research, the RIs in the SSH are involved in a wide range of international developments or form a national branch of these. Important examples for the humanities include ARIADNE, Europeana, CLARIN, DARIAH, EHRI and E-RIHS. A number of the domain-oriented European digital infrastructures offer portals, through which researchers can find and access data. There are promising developments in the work on interoperability. Developments in the field of the Digital Humanities also offer many opportunities for international collaboration outside of Europe, for example through the Alliance of Digital Humanities Organizations (ADHO).
Important data sources for the social sciences are the Dutch branches of the European Social Survey European Social Science Infrastructure (ESS ERIC), the Survey of Health, Ageing and Retirement in Europe (SHARE) and the Generations and Gender Program (GGP). CESSDA, the Consortium of European Social Science Data Archives, has an overarching role.
Other important actors are the international organisations concerned with (FAIR) research data such as CODATA, GOFAIR, the Research Data Alliance (RDA), the World Data System and overarching DataTogether. All told, there are hundreds of Dutch researchers who are actively involved as members, including within the SSH. Finally, we mention the European Open Science Cloud (EOSC), the main umbrella organisation of European infrastructures. A number of Dutch organisations are members of the EOSC association. SURF acts as the national representative for EOSC. All international SSH RIs contribute to EOSC via the SSHOC project.

[4] An overview of the Dutch landscape is offered in the DANS Data Landscape Monitor. For a landscape analysis, see the EOSC-SYNERGY Landscape Gap Analysis and recommendations of the EOSC SYNERGY project.

[5] *Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship* (2021), 7.

2. Summary of Challenges

We made use of several existing inventories that have been produced in recent years. In 2021, *Storage and Availability of Data for Research. From intentions to implementation* was published under the auspices of the KNAW. Although this report stated that the differences per research area are considerable, similarities were also found. These were also reasonably consistent with earlier reports, such as the Final Report on Exploring and Optimizing the Dutch Data Landscape from NPOS published in 2020. Furthermore, in the SSH, the preparation team recently produced an informal and unpublished inventory in the context of the latest Roadmap LSI call in which both authors of this memorandum were involved. Very recently, a new consultation round was held by the Permanent Committee on National Institutes (PCNI). One of the authors of this report participated in the final conference and read the initial draft report. In summary, ten problem areas can be distilled from the existing body of reports and the discussions we have held. Eight of these are general in nature, and two (digitalisation and data collection) are specific to the SSH.

1. Awareness

By far not all researchers are aware of the need for good data management and the sharing of data. Even among those who regularly share data, scarcely half are familiar with the concept of FAIR data.[6] The extent to which that applies to national SSH research is unknown, but there is no reason to assume that the figures there are more favourable. In the European Open Science Cloud (EOSC) report *Six Recommendations for Implementation of FAIR Practice*, awareness training is explicitly recommended. The NPOS report *Professionalising Data stewardship in the Netherlands* makes the same recommendation.

2. Recognition

The recognition and rewards for publishing data and software are insufficient.[7] This has a detrimental effect on the motivation of researchers to publish data or software. Once research has been completed and published, then it is more favourable with respect to a researcher's track record to immediately start on a subsequent article and to leave the dataset and the software that provided the basis for the previous article unpublished (because publishing also costs time). It is a problem that requires attention everywhere and also a problem that is often explicitly stated by researchers. One of the recommendations in the *San Francisco Declaration on Research Assessment* (DORA) is therefore: 'For the purposes of research assessment, consider the value and impact of all research outputs (including datasets and software).' Numerous research organisations worldwide, including NWO and Universities of the Netherlands (VSNU) have endorsed this declaration.

3. Expertise

In university curricula, there is a growing interest for computational methods and techniques and how to deal with data and software. At the same time, overall knowledge about research data management and analytical computational methods is not sufficiently present among current generations of researchers. [8] That is certainly the case in the SSH, where certain subjects attract students whose interest is explicitly *not* in figures and technology. Knowledge of research data management should at least mean that researchers know what FAIR data constitutes and what is necessary to realise its four aspects at an acceptable level. Another blind spot of researchers is developments in the area of (open) linked data. Training programmes that stimulate the digital literacy of researchers are crucial for a thriving infrastructure.

4. Support

A researcher should possess basic knowledge, but the help of experts is needed for in-depth analyses and excellent research data management. That is initially provided by software developers and data stewards who are employed at a local DCC and, beyond that, in-depth expertise that is offered, for example, at a national level by engineers/researchers from the eScience Center, technicians from SURF or data experts from DANS.

---

[6] *Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship* (2021), 8.

[7] National Plan Open Science (2017), 25.

[8] *Storage and Availability of Data for Research. From intentions to implementation*, 27.

Data Stewards can fall back on these experts. Although research institutions are working on setting up DCCs, as discussed above, researchers still experience an acute lack of domain specific support.[9]

5. Sensitive data

A peculiar aspect of dealing with data is privacy legislation and how it is applied at many locations. Many researchers are of the opinion that the interests of the research unnecessarily suffer heavy damage. The social sciences, in particular, struggle with this, but it also increasingly applies to the humanities, where social media data are an increasingly important source for research. The necessary legal knowledge and support are often lacking and, insofar as it is present, university legal advisers often prefer to take no risks, as a result of which entire research projects must sometimes be modified or even called off. Dealing with personal data often requires special digital environments, to which not all researchers can easily obtain access, insofar as these are present in the first place. Another problem is found in the copyright legislation, which can throw up expensive obstacles in the case of media studies, art history or literature studies, for example. Our consultations revealed that various universities are working on digital tools that state whether an action is permissible. The TDCC could probably play a role in helping to realise collaboration and alignment in this area.

6. FAIR data and software

The availability of optimally FAIR datasets and software forms a necessary condition for the proper functioning of RIs such as CLARIAH and ODISSEI, or the SSH-wide *Macroscope* that is described in the new Roadmap for Large-Scale Research Infrastructure. And that applies to all digital RI. However sophisticated the tools, powerful the computers and organised the RIs are, incomplete data will yield limited research results.

FAIR data and software form the basic condition for good science, but this basic condition is still not satisfied. As the reports referred to above reveal, researchers struggle with the availability, findability, accessibility and interoperability of data. That mainly applies to data produced in previous, already published research. In the case of software, the state of affairs is probably even worse. Although institutions like Software Heritage and GitHub offer adequate services to deposit software, the vast majority of built research software remains on local servers or laptops of researchers, unfindable and inaccessible for others. That is not cost-efficient because the programming of research software is often associated with considerable investments.

This limited reusability of research data and software is not only inefficient but also hinders reproducibility and verifiability and therefore harms the quality of the research. There is a lack of sufficient incentives, resources and organisation to make the bulk of research data produced available.

Although it is not conceptually complex, the challenge of realising an acceptable level of availability, accessibility and interoperability of research data and software is gigantic.[10] In practice, the production of FAIR data boils down to the fact that datasets and software must be provided with standardised and adequate metadata from which it is clear (also for machines) how that data arose and what a researcher can and may use it for. That last aspect is also essential to satisfy the privacy and copyright legislation: FAIR data is not automatically open data. Permanent findability must be guaranteed through the attachment of a persistent identifier (PID). Knowledge graphs are also needed to be able to search for and link data. These matters form the foundation. Without standards, high-value metadata, knowledge graphs and PIDs, other researchers cannot easily find data and software, and accessibility and interoperability will be problematic.

The TDCC SSH will carry out (or commission) a study into the current situation regarding FAIR data and software in the SSH (see bottleneck projects). What we currently know based on the DANS Data Monitor applies to the national level. [11] Of the potentially available Dutch research data, only a fraction is currently stored in a reliable and certified environment under controlled conditions and is optimally findable for the long

[9] *Storage and Availability of Data for Research. From intentions to implementation*, 28.

[10] "(…) making data fully FAIR is difficult, if not impossible". *Storage and Availability of Data for Research. From intentions to implementation*, 22.

[11] See: DANS Data Monitor 2020-2021

term as FAIR data. [12] In a large number of cases, persistent identifiers are not allocated to data. [13] Most repositories use standards for the allocation of metadata, but the overall qualitative level that would at least make the data reasonably FAIR has not yet been reached. A lot also depends on how completely and consistently metadata fields are filled out in practice.[14]

7. Digitalisation

Although digitalisation in the heritage sector is gradually progressing, the vast majority of the historical collection of the Netherlands is not yet digitally available, which forms a barrier for research into Dutch history and culture. Digitalisation is expensive, and it is difficult for heritage institutions to acquire the funding necessary to make fast progress. Once the material has been digitalised, then in many cases, it is met with considerable interest as revealed by the success of the previously stated newspaper archive Delpher or of digitalised population registers. Although the quantity of digital-born material is rapidly increasing, archiving and granting access to this still forms a technical and legal challenge.

8. Data collection

When asked about their wishes, many SSH researchers state a specific data collection that they would like to have access to for their research. In the SSH domain, data collection is often tedious manual work and therefore expensive. Examples are sociological surveys, tabling data from population registers, printed statistics and accounts, or manual transcriptions of handwritten documents. The cost for surveys in the social sciences have risen explosively.

9. Funding

In the Netherlands, RIs are largely funded on a project basis, whereas the management and maintenance costs are structural and have risen considerably in recent years. In many cases, institutions are required to guarantee that after the project, the management and maintenance costs will continue to be borne for a given period of time. That is often realised by institutes outside of the university world, such as KNAW institutes. This funding problem is particularly present in the SSH.[15] It often concerns data collections or facilities for which termination after such a period has ended would mean an enormous loss of capital, and so these collections are continued and further developed. It can be foreseen that the cumulative costs of the increasingly larger infrastructure will eventually become unbearable if the structural funding remains the same.

10. Flexibility and growth

Developments in digital infrastructure are struggling to keep pace with the increasingly faster developments in society and science. Infrastructure orders and facilitates but must also be open to new developments: new types of data that are emerging (e.g. digital-born data or 3D objects) or the need for new tools and computational methods require a high degree of flexibility to prevent an infrastructure emphasising too much status quo and hindering innovation.
In addition, as noted earlier, in the SSH domain a strong increase is expected in the coming period in the use of digital methods and techniques. How can future growth be facilitated and accommodated?

---

[12] It is not known what percentage of the research data produced has been deposited in a data repository. This is probably a minority of the data. According to the DANS Data Monitor 2020-2021, 18 of the 62 Dutch repositories were certified in 2020.

[13] According to the DANS Data Monitor 2020-2021, 63% of the data repositories in the Netherlands assigned a PID.

[14] According to the DANS Data Monitor 2020-2021, 51 of the 62 repositories use a standard (such as Dublin Core, DDI or DataCite) to describe the data managed. DANS did not investigate the completeness and quality of the (standardised) metadata entered, but it is suspected that it is unsatisfactory in many cases.

[15] 'A lack of structural investments in the digital infrastructure will hinder these projects or make it impossible to safeguard the results. And, in turn, that will hinder efforts to further build upon these projects. In that case, Dutch humanities and social sciences researchers will miss the connection with other countries.' Rapport Topwetenschap vereist topinfrastructuur. Adviesrapport nationale digitale infrastructuur voor wetenschappelijk onderzoek, 27

3. Investment plans: network organisation

Consortium

CLARIAH, DANS and ODISSEI currently form the basis of the consortium.

Description of organisation and activities

Organisation
For an optimal effect, a TDCC can best focus on fundamental issues that are important for large parts of the domain and that are not already tackled via a different avenue. That last aspect means that the TDCC should primarily be aligned with existing infrastructures and organisations in the scientific domain but also national and international developments and infrastructures, especially the Local DCC's. In the Dutch SSH domain, the two large collaborative RIs have already emerged: CLARIAH and ODISSEI. New LRI Roadmap proposals in the pipeline seek to further expand the domain infrastructure. A TDCC will need to work closely with the Local DCC's and these infrastructures to yield the maximum benefit. Where a Local DCC offers every-day expert support for researchers on the shop floor, the Thematic DCC's support the Local DCC's by bringing in specific, up-to-date knowledge about domain-oriented standards, practices and services. In addition, a LDCC can formulate and identify local projects to be granted through the TDCC, for instance on the FAIRification of datasets or software, or for raising awareness and expanding knowledge of researchers about standards and tools used in a scientific domain.
In relation to the RI's, the added value of a TDCC lies, amongst other things, in its structural character and the fact that it works across the domain with expert knowledge on the one hand while it is not linked to certain disciplines or a limited set of scientific questions. The TDCC serves as a domain community hub, where connections are being made and maintained, coordination and alignment are being facilitated and projects will be carried out to address issues that hitherto have not received enough attention.

The TDCC SSH will be set up as an open network organisation that organisations, units of organisations and individuals can join. There will be room to accommodate new developments and new initiatives will be welcome (and even stimulated) to join. Together they will form a members' council that will meet once a year in what can be seen as the main annual event of the TDCC. We envision the membership council to be a large body with hundreds of members, where the main objectives, strategies and results of the TDCC can be laid out and discussed. Membership is for free, but members will be asked to sign a manifesto, in which they commit to the objectives of the TDCC and subscribe to the principles of Open Science and FAIR research objects.[16] Members are invited to summit project proposals that contribute to these objectives.

The network will be governed by a board consisting of the main user-communities, contributors to research infrastructures and persons with relevant expertise. It is important for the network to include domains which are not yet represented in the two main SSH research infrastructures.

- The main RI in the Social Sciences (ODISSEI)
- The main RI in the Humanities (CLARIAH)
- The main SSH-service provider (DANS)
- Universities and research institutions
- The Local DCC's
- The Digital Heritage Community (Netwerk Digitaal Erfgoed)
- The Platform Digital Infrastructure SSH
- Libraries and other heritage institutions
- Planning agencies
- Open Science Communities
- The two other TDCC domain organisations
- Other stakeholders and experts (e.g.: GDPR expert/copyrights expert)

[16] This is a working membership model, copied from the Netwerk Digitaal Erfgoed.

The board will be responsible for approving and monitoring the strategic course of the TDCC. The board will gather twice a year, or more if necessary.

A small, three-headed executive board will be responsible for setting up the strategy and supervise the implementation of it. A network coordinator and a project manager will conduct the daily management under responsibility of the executive board. The executive board will meet monthly, in most cases joint by the managers. The Executive Board consists of the directors of CLARIAH, DANS and ODISSEI, the main parties responsible for building and maintaining the research infrastructure of the SSH-domain.

Under the responsibility of the executive board, a network coordinator and project manager will jointly form the daily management. The chairing of the (executive) board will rotate annually under the executive board members. The chair will maintain contact with NWO.

The executive board will draw up a rolling multiannual programme to be annually approved by the board after being discussed in the members council. The executive board will actively support and stimulate the SSH community in jointly defining and prioritizing its TDCC needs. The executive board will appoint the network coordinator and project manager, who will report to the executive board. Each year, the board will approve the budget and the annual accounts.
The network coordinator is responsible for the organisation of the network and the community management. He/she will recruit, encourage and involve the members and organise their meetings. The project manager will support the board, draw up budgets, annual plans and (with the help of the financial department of the institution that houses the administration of the TDCC) the annual accounts. He/she will lead the realisation of the non-project-based programme aspects of the TDCC and shape the collaboration with other parties. Furthermore, together with the chair, the network coordinator will maintain contact with NWO on behalf of the TDCC.

The project manager is responsible for all processes with respect to the community and the projects, including coordinating training programmes, administrative processes, monitoring the progress, completion, et cetera. He/she will be supported by the financial department and the secretariat of the institution where the TDCC is administratively accommodated.

We estimate the support to be funded by NWO as 2 FTE at salary scale 12 for an initial period of five years, to be co-funded by the TDCC-network After successful evaluation, it should be continued for at least another 5 years.

Activities
The activities of the TDCC SSH will focus on increasing the use and reuse of data and software. We see five clusters of activities to facilitate this.

1. Increasing the percentage of findable, accessible, interoperable and reusable (in short: **FAIR**) research data and software.

2. Raising **awareness** of researchers about the importance of FAIR data and software.

3. Enhancing the level of **knowledge** in the area of digital data used in the SSH and the usage of SSH-oriented tools, aimed at researchers as well as data stewards.

4. Addressing a number of **pressing issues** related to the collection and usage of data, like privacy and copyright legislation and the high costs of collecting and producing digital data.

5. Building and maintaining an open and inclusive **network organisation**, involving all relevant parties, as a vehicle for making 1-4 happen.

Explanatory:

- We consider the FAIRification of research data and software (1) as the key-issue. Infrastructures do not work properly without it. Awareness raising (2) and knowledge enhancement (3) can be looked upon as instruments to advance FAIRification, although of course their implications go beyond this context. The fourth cluster combines a number of more specific problems encountered by certain research communities. The last cluster is about organisation-building and can be considered as a precondition for cluster 1-4.

- With this list of activity-clusters, most of the challenges mentioned in Paragraph 2 are being addressed, except for recognition, digitization, local support and funding, which lay beyond the scope of the TDCC.

- To kick-start the five activities, in Paragraph 4 we will propose a number of *bottleneck-projects*, each tied to one of the activity-clusters.

- Carrying out these activities will be a collective effort of the TDCC-network, in which Local Digital Competence Centres, Research Infrastructures, Research Performing Organisations and Service Providers will join forces with more informal research communities and individuals representing them.

Main hub of organisation

DANS is prepared to serve as the main hub. DANS is part of the legal entity KNAW, is jointly governed by NWO, and falls under the collective labour agreement NU. It has about 50 FTE of employees and has the facilities to accommodate the board, committees and members of the TDCC and to provide administrative and secretarial support. DANS is independent of any discipline and has considerable experience with realising, administrating and providing accountability for large and/or European projects. DANS has a wide support base in the SSH. It is not only closely involved in the development of CLARIAH and ODISSEI, but is also well-connected in disciplines that are not part of either one, such as archaeology.
DANS will just serve as a host-organisation and will have no special authority in the network. In the contracts with the TDCC employees, it will be recorded that they will be managed by the board of the TDCC.

Contribution of consortium

For the first two years of the TDCC, NWO will not require any co-funding. This period will be used to make agreements about co-funding in the form of personnel deployed.

4. Investment plans: bottleneck projects

This report has identified and explained a series of challenges and has linked five clusters of activities to these. Based on this list of activities, we present a series of bottleneck projects that should be set up in the short term. The primary objective of the bottleneck projects is to get the TDCC SSH up and running and to create a strong support base in the SSH. We have chosen a show and then tell approach in which, from the start, we will show that, in practice, researchers and institutions can benefit from the TDCC. Fast and tangible results on the work floor – and the showcasing of these online and at events – will lead to appreciation, growth and active participation in the network and ensure a flying start. In addition, we will realise bottleneck projects that tackle a number of problems that stand in the way of a structural realisation of FAIR data and software practices, such as basic knowledge among researchers and domain-specific knowledge among data stewards, the FAIRness of data in repositories, privacy and copyright legislation, interoperability of data and software, and the FAIRness of software. To realise this, we are considering the following projects. The amounts stated are still very tentative and initial and merely serve to indicate an order of magnitude.

Bottleneck projects

Activity cluster 1: FAIRification of data and software

- **Intelligence** - Carrying out a study into the current situation regarding FAIRness of datasets in the domain. [20 k€]

- **Low-threshold, small-scale FAIR data projects for researchers and data stewards** – The executive board will actively support and stimulate the SSH community in jointly defining and prioritizing the TDCC needs of FAIR data projects of researchers and data stewards. Besides creating support and appreciation for the TDCC network, this bottleneck project will lead to an expansion of knowledge and will quickly yield a large quantity of FAIR SSH datasets. [300 k€]

- **Improving FAIRness in repositories** - The setting up of a funding programme that offers repository services and other organisations the opportunity (based on proposals submitted) to improve the FAIRness of the datasets entrusted to them and to set up their work process in such a way that datasets still to be deposited will also be optimally FAIR. If so desired, this could lead to repositories being eligible for a CoreTrustSeal certification. This support could (partially) be provided in the form of expertise comparable with the approach of the eScience Center. [300 k€]

- **Interoperability** - An increasingly important desire for data-driven research is interoperability within the domain and across the boundary of domains. That primarily requires that the dozens of classification systems, thesauruses, vocabularies et cetera used in the domain are identified, and made findable and searchable, possibly in a similar manner to what the Dutch Digital Heritage Network did with its Network of Terms. This is a typical bottleneck for realising interoperability. Interoperability also forms an important point of attention in the SSH LRI Roadmap proposal, which means that consultation needs to take place with CLARIAH/ODISSEI […]

- **Software** - The problem of FAIRness of software requires special attention. In the Netherlands, the eScience Center, in particular, is drawing attention to this, but this has not yet led to the development of deposition practices that already exist for data. This problem also exists in other domains, and the other TDCCs will also develop projects to tackle this. This is something we can work together on in collaboration with the eScience Center. Examples are the development of guidelines for researchers concerning sustainability of the research software, the development of specific courses and instruction material in areas such as publishing software (archive and repositories), software citation, and software licences and copyright. In the project, consultancy practices will be developed to support researchers who want to improve the sustainability of their research software. […]

Activity cluster 2: Raising awareness

- **Awareness campaign** - This will be deployed for the first (peer-to-peer) awareness campaign, in which several projects will be funded. This campaign will be set up in close cooperation with the LDCCs, which will bring in their knowledge of the local situation, while the TDCC provides domain-related showcases and best practices for curating and depositing data and software. [100 k€]

Activity cluster 3: Knowledge enhancement

- **Training for researchers and data stewards** - An important bottleneck lies at the basis of the research. Researchers in the SSH expressed the need for low-threshold, basic training and workshops in the area of **Research Data Management** and **software carpentry** that are aligned with their own research practice. In a bottleneck project, we will design a programme for this in collaboration and consultation with local DCCs, the RIs, the Research Data Netherlands Network (RDNL) training network and the research schools working in the domain. We will also involve a group of SSH researchers from the Open Science Communities to ensure sufficient focus on the perspective of the (young) researcher. The programme will consist of modules that will be tested in practice and adjusted, if necessary. In addition, we want to set up a programme in which data stewards are trained in supporting domain-specific research practices. Ultimately, the training programme will become a part of the standard TDCC practices. [100 k€]

- **Re-use of training material** - Just like data and software, training material should also be findable, accessible and reusable. This is not the case yet. In a project, we will sort out how this best could be duoachieved in the SSH-domain. In cooperation with the other TDCC's and RDNL we will explore how aggregation at the national level can take place.

Activity cluster 4: Addressing pressing issues

- **Privacy and copyright legislation** - In certain areas of the SSH, researchers struggle with limitations that arise from privacy and copyright legislation. There is a need to exchange knowledge, experiences and practices and to record these in a knowledge bank that can be consulted in the future. The problem of ownership also fits in this bottleneck project. In the bottleneck project, we will investigate together with a group of researchers involved how this can best be realised and take steps towards its realisation. Obviously, this will be done in collaboration with the other TDCCs. [100 k€]

- **Legislation tools** - The consultation rounds showed that various universities and other institutions are working on tools that indicate what is and is not permitted by law when working with sensitive datasets. In this bottleneck project, these endeavours are inventoried and it is examined whether it is efficient to combine the efforts. [...]

Activity cluster 5: Building a network

- **Online presence** - The realisation of an online presence via the web and social media. [50 k€]

- **Kick-off meetings** - The organisation of national kick-off meetings and a series of smaller physical events on location in collaboration with local DCCs and Open Science Communities. Giving presentations at events organised by third parties. [30 k€]

- **Signalling, coordinating, informing and connecting**
  There is a clear need in the SSH domain to be well informed as a community of users regarding the most important developments in the area of digital data use. The TDCC SSH will signal important

national and international trends, and encourage and coordinate activities that increase awareness and expertise in the area of digital data. That can concern training sessions in data management, computational methods or cross-disciplinary research projects.

5. Data hubs

A description of the main data hubs has already been included in the section Institutions, organisations, infrastructures, public/private parties.

Annex: Sources

DANS Data Monitor 2020-2021 (2021)

Final Report on Exploring and Optimizing the Dutch Data Landscape van het Nationaal Programma Open Science (2020)

Durven delen. Op weg naar een toegankelijke wetenschap (2016)

National Plan Open Science (2017)

NWO Onderzoek Open Science (2021)

- Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship (2021)
- Samen Sterker. Beeld van het SSH-domein (2019)
- Storage and Availability of Data for Research. From intentions to implementation (2021)

Topwetenschap vereist topinfrastructuur. Adviesrapport nationale digitale infrastructuur voor wetenschappelijk onderzoek (2017)

Van inzicht naar impact. Sectorplan Maatschappij- en Gedragswetenschappen 2020-2025 (2020)