1. General information

1.1 Grant application title

Social Science and Humanities Open Cloud for the Netherlands (SSHOC-NL)

1.2 Consortium lead and other consortium partners

Erasmus University Rotterdam, Erasmus School of Social and Behavioural Sciences – consortium lead

Consortium partners

- KNAW-Data Archiving and Networked Services (DANS)
- KNAW-International Institute of Social History (IISG)
- KNAW-Meertens Institute (Meertens)
- Statistics Netherlands (CBS)
- University of Amsterdam (UvA)
- Utrecht University (UU)
- VU Amsterdam (VU)

1.3 Main applicant

Name	**Prof. dr. Pearl (P.A.) Dykstra**
Address for correspondence	Erasmus University Rotterdam, Burg. Oudlaan 50, office T19-05, 3062 PA Rotterdam
Email	dykstra@essb.eur.nl
Phone number	+31 10 40 82 076
Website	www.eur.nl/people/pearl-dykstra

1.4 Consortium members

Title, first name, initials, surname	Affil.	Expertise	Reference	Roles[1]
Main applicant				
Prof. dr. Pearl (P.A.) Dykstra	EUR	Sociology	0000-0003-1518-1476	PI
Co-applicants				
Prof. dr. Antal (A.P.J.) van den Bosch	Meertens	Language technology	0000-0003-2493-656X	Co-PI
Prof. dr. Jacco (J.R.) van Ossenbruggen	VU	Human-Centred Data Science	0000-0002-7748-4715	SB
Prof. dr. Julia (J.J.) Noordegraaf	UvA	Digital Heritage	0000-0003-0146-642X	SB
Prof. dr. Daniel (D.L) Oberski	UU	Methodology	0000-0001-7467-2297	SB
Prof. dr. Rens (R.) Vliegenthart	UvA	Communications sciences	0000-0003-2401-2914	SB
Prof. dr. José (J.F.T.M.) van Dijck	UU	Media & Digital Society	0000-0003-0499-9045	SB

[1] PI = Principal Investigator; SB = Steering Board

Prof. dr. Karina (K.H.) van Dalen-Oskam	Huygens	Computational Literature	0000-0001-7697-4512	SB
Prof. dr. Gijsbert (G.J.) Rutten	UL	Language technology	0000-0003-1890-908X	SB
Prof. dr. Chantal (C.) Kemner	UU	Development Psychology	0000-0002-8879-2588	SB
Prof. dr. Clara (C.H.) Mulder	RUG	Demography	0000-0003-0152-2225	SB

1.5 Keywords

Data Infrastructure, Data Standards, Data Linking, Digitalisation, Open Access

1.6 Relevant research field(s)

Code/Field of research:		
37.10.00 (main)	30.55.00	40.20.00
(Software for Humanities)	30.90.00	40.30.00
	31.30.00	40.40.00
14.40.00	31.90.00	40.50.00
21.10.00	32.20.00	40.60.00
23.80.00	32.60.00	41.90.00
23.44.00	32.70.00	42.00.00
23.46.00	32.80.00	43.40.00
27.40.00	32.90.00	44.10.00
27.50.00	35.40.00	44.20.00
27.60.00	37.20.00	45.90.00
27.70.00	37.30.00	46.90.00
27.90.00	37.40.00	47.90.00
28.90.00	37.50.00	48.90.00
30.25.00	37.90.00	49.10.00
30.30.00	38.10.00	49.11.00
30.35.00	38.30.00	
30.40.00	39.90.00	
30.45.00	40.10.00	

1.7 Relevant scientific domain(s)

100% Humanities and Social sciences

1.8 Abstract

SSHOC-NL is a consortium of research infrastructures aimed at creating an ecosystem of services, data and tools for the social sciences and humanities. The consortium is spearheaded by ODISSEI, the national infrastructure for Social Science and CLARIAH, the national infrastructure for the Humanities. Together they will increase interoperability across the domain and allow services, data and tools to be shared, linked, and combined in imaginative and ground-breaking ways. SSHOC-NL will elevate existing services, data and tools through technological and stakeholder readiness levels and ensure that they are mature, stable and widely accessible to the research community at large. SSHOC-NL therefore represents the bold next step in the natural evolution of these two broad and successful infrastructures and in so doing, SSHOC-NL will open up new lines of research in areas of considerable societal interest.

Greater interoperability between cultural and media data on the one hand, and the instruments for collecting public opinion and the dynamics of group behaviour on the other will enable researchers to understand just how deep divisions in our society go, as well as their ability to affect contemporary social debates. Such knowledge will provide a basis for dealing with misinformation and conspiracy theories.

By linking contemporary population registers with historical records, scientists will be able to conduct deep pedigree analyses and better understand how a wide variety of inequalities such as educational or labour market outcomes are intertwined with genetic determinants in these very same outcomes.

By aligning innovations in geospatial research across ODISSEI and CLARIAH, it will be possible to develop new insights into the determinants and consequences of environmental change. These insights require a multilevel, interdisciplinary perspective that allow researchers to zoom in and out to understand the relationship between individuals, society and their environment and develop sustainable, holistic solutions.

SSHOC-NL will also strengthen and intensify the links between SSH and other domains by establishing common standards and vocabularies for the whole of SSH. This enables research infrastructures in other domains to align with the entire SSH domain, collaborating with SSHOC-NL rather than working with a myriad different SSH infrastructures within the domain. In this respect, the potential for SSHOC-NL to increase the social and cultural dimension in broader scientific research is vast and very overdue.

By creating a cohesive, coherent and FAIR ecosystem, SSHOC-NL will enhance the Netherlands position at the cutting edge of social science and humanities scholarship and create fertile new ground for expansive and exciting new lines of research.

1.9 Popular project title and public summary

Dutch

The Social Science and Humanities Open Cloud for the Netherlands

SSHOC-NL is een samenwerking tussen de sociale wetenschappen en de geesteswetenschappen. Het maakt het voor onderzoekers mogelijk om veilig en ethisch verantwoord een grote hoeveelheid data aan elkaar te koppelen en te analyseren, zoals historische gegevens, tekstdata, afbeeldingen, enquêtedata en social media data. Dit helpt hen om de meest urgente maatschappelijke vraagstukken te adresseren, zoals polarisatie, sociale ongelijkheden en het veranderende milieu. Deze vraagstukken zijn complex en om ze te begrijpen moeten verschillende perspectieven op elkaar worden betrokken om oplossingen te vinden. SSHOC-NL bouwt de infrastructuur voor onderzoekers om dat te doen.

English

The Social Science and Humanities Open Cloud for the Netherlands

SSHOC-NL is a collaboration between the social sciences and humanities that is cutting edge and unparalleled internationally. It will make it possible for researchers to securely and ethically link and analyse a huge range of data such as historical records, text data, images, survey data, and social media data. This will help them address some of the most pressing issues that society faces such as polarisation, social inequalities, and our changing environment. These issues are complex and require lots of different perspectives to be brought together in order to find solutions. SSHOC-NL builds the infrastructure for researchers to do that.

2. Large-Scale Research Infrastructure (LSRI)

2.1 General description of the LSRI

The Social Sciences and Humanities (*SSH*) cover economics, sociology, psychology, media studies, linguistics, literature, art history, heritage science, history, communication sciences, political science, and several more. Despite this diversity they share many infrastructural needs that are largely unique to SSH.

Given their focus on people, their interactions, and the products of their minds and culture, SSH share a need for *secure and flexible analytical environments* to analyse sensitive and copyright protected data and artefacts. Tailored infrastructural investments are needed to ensure that research can break new ground without compromising the privacy of individuals or infringing intellectual property rights. This is true regardless of whether the data involves highly detailed health records, an annotated film or TV collection, oral history interviews, granular social network information, or a historical collection of letters.

SSH data is also notoriously complex, sometimes generated by processes which were not designed with science in mind. Data stems from centuries old administrative records, text documents such as newspapers, modern tax systems, broadcasting archives, but also personal information from wearable devices such as fitbits, or terabytes of social media posts that have been harvested from a range of platforms. This complex data requires *bespoke tools* that are methodologically sound, well documented, and transparent so as to ensure that complex and sometimes chaotic data streams can be tamed by transparent, reproducible, open science workflows.

SSH are also exceptionally interdisciplinary, with an emphasis on *bringing together diverse forms of data and methodologies* that offer complementary perspectives on culture and society. To harness this diversity and the opportunities afforded by open science, common standards and vocabularies need to be established and propagated across the domain through the creation of shared FAIR (Findable, Accessible, Interoperable, and Reusable) community guidelines.

To address the needs of the SSH community, we propose the creation of an *ecosystem* consisting of distributed, federated and collaborative infrastructures (see Figure 1 and the Glossary in section 6.1). This ecosystem will facilitate new, interdisciplinary lines of research which draw on the strengths of multiple disciplines across SSH. This proposal brings together a grand coalition of existing SSH research infrastructures including the Open Data Infrastructure for Social Science and Economic Innovations (ODISSEI), the Common Lab Research Infrastructure for the Arts and Humanities (CLARIAH), but also more field specific infrastructures such as the Netherlands Institute for Conservation+Art+Science+ (NICAS), and the European Holocaust Research Infrastructure (EHRI). The scale, precision and exigency of the social sciences and depth, diversity and detail of the humanities must be brought together in order to *understand the exceptionally complex societal problems* that society currently faces. This project raises the SSH infrastructural ecosystem to a new level by making the Netherlands' existing cutting-edge SSH infrastructures truly interoperable, open, and embedded within the broader European Open Science Cloud (EOSC). It will be a Social Science and Humanities Open Cloud for the Netherlands: *SSHOC-NL*.

SSHOC-NL will:

1. Dramatically improve the accessibility, interoperability, and sustainability of SSH research infrastructures and facilitate collaborations that are globally unique and interdisciplinary in nature;

2. Connect and train a community of SSH scholars to jointly work on knowledge development, technology development, technology transfer, and innovation;

3. Create a future-proof and coherent ecosystem that enhances the position of Dutch SSH research infrastructures within the European Open Science Cloud and achieves economies of scale.

SSHOC-NL is a formal consortium of research infrastructures aimed at creating an ecosystem of services and resources that is driven by the FAIR principles and is a magnet for talent across the whole Dutch and international SSH community. Existing infrastructures and services are to be reorientated and scaled up through shared formats, breaking existing silos and enabling new lines of collaboration. The project is structured around five workstreams which represent the different necessary components identifiable in any world-class research infrastructure: Compute, Data, Tools, FAIR, and Expertise. By organising the SSHOC-NL project around these five workstreams, a common

ecosystem will be created that realises the project's three aims and reconfigures the SSH domain so as to achieve ground-breaking interdisciplinary research.

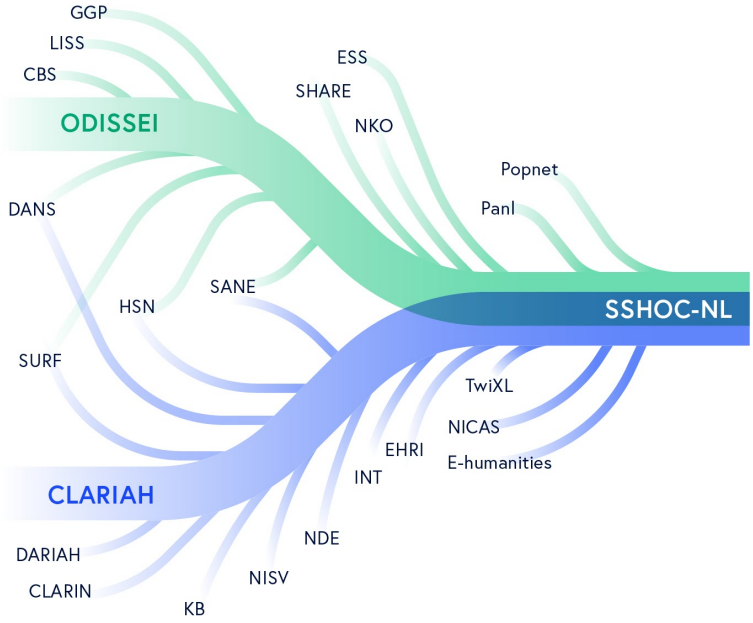


Figure 1 – The SSHOC-NL ecosystem

2.2 Expected scientific innovation and breakthroughs

Many new lines of research will be opened up through the collaborative and interoperable infrastructural ecosystem that SSHOC-NL will deliver. Here we present just three examples; all are linked to the National Science Agenda (NWA) that will provide a substantive focus throughout the project. There are many more research lines, both foreseen and unforeseen, which will be enabled by SSHOC-NL, but these three serve as excellent illustrations of the promise of an interdisciplinary research infrastructure. They are all key societal challenges that are infinitely complex and require the combined strengths of Social Science and Humanities research. They require a simultaneous focus on micro-level, individual-level data and macro-level processes that involve structural, institutional aspects. This interdisciplinarity will generate new knowledge that contributes to the development of more holistic, inclusive and, thus, more sustainable future solutions.

a. Polarisation

The COVID-19 pandemic brought the existence of *parallel worlds* within our society into sharp relief. A section of the population has an intensifying distrust of government, formal institutions and expertise (Roozenbeek et al. 2020). At times such as the current pandemic, mistrust and the spread of *misinformation* are a key danger for public health and social cohesion (Bennett and Livingstone 2018). Such misinformation and distrust can also impact elections, directly influencing the core of our democracy (Brown 2018). Gaining insight into the dynamics of public debates as they develop over time across both traditional media outlets and new digital platforms is of crucial importance for a healthy, democratic society. Only by combining state-of-the-art insights from joint digital humanities and social science research is it possible to truly understand how these debates develop and shape individuals' attitudes, values and behaviour. The social sciences and humanities in the Netherlands together have collected a vast amount of data in this area already. SSHOC-NL will make existing data infrastructures interoperable and create a community of scholars, enabling an analysis of these phenomena from multiple perspectives.

The CLARIAH Media Suite currently provides an interface for accessing audiovisual content, including all digitised items in the Netherlands Institute for Sound and Vision (NISV) public radio and TV collection (Ordelman et al. 2018). It enables automated speech recognition, full transcription of audio and video files, processing videos with computer vision technology including face, object, and motion detection, as well as audiovisual content enriched with manual annotations (Van Noord et al. 2021, Melgar Estrada et al. 2017). This broad functionality allows scholars to perform the equivalent of full-text search on audio and visual content. ODISSEI maintains long-standing instruments that track public opinion, attitudes, values, and media consumption such as the Dutch Parliamentary Election Study (NKO), European Social Survey (ESS), European Values Study (EVS) and the Longitudinal Internet studies for the Social

Sciences (LISS) panel. These social survey instruments provide detailed insights into changes in the public's opinions and perceptions. However, humanities and social science scholars struggle to capture how both traditional and social media are understood, reinterpreted, and shared by the public. It is one thing to measure media content or public opinion, but many exciting new questions could be answered if researchers were able to observe the interaction between media and individuals. To enable this, these existing infrastructural components will be brought together in SSHOC-NL.

A wide array of media content and data will be made accessible and interoperable, among others via the TwiXL project that makes available the majority of the Dutch language Twitter domain, the Dutch web collection of the National Library (KB) of the Netherlands, and the Amsterdam School of Communication Research (ASCoR) online news dataset. These digital media sources will be analysable in a range of analytical environments including the CLARIAH Media Suite, and can be linked with existing collections such as the Newspaper Collection of the KB and the TV collection at the NISV, allowing for cross-media research on both historical and emerging public debates via a single central platform. These expanded media collections will be interoperable with ODISSEI data collections on public opinion, and it will be possible to integrate content into the fielding of Mass Online Experiments.

Mass Online Experiments are at the cutting edge of sociological, psychological and communication science research, allowing participants to interact in real time. In 2021, ODISSEI conducted a pilot experiment where, in a prototype of the Mass Online Experiments tool, approximately 1,500 participants were divided into groups of 96. In some groups, respondents were connected randomly, while in others, respondents were connected to people who held similar political opinions to them, creating artificial polarisation within the experiment. Respondents were then provided with news items of varying veracity and could choose whether to share this with the other respondents they were connected to. The results of these experiments showed that fake news proliferated more broadly in networks that are homogenous (Nunner et al. 2021) (Figure 2). SSHOC-NL will accelerate access to Mass Online Experiments by embedding them within a national panel and providing the technical and methodological expertise to conduct them.

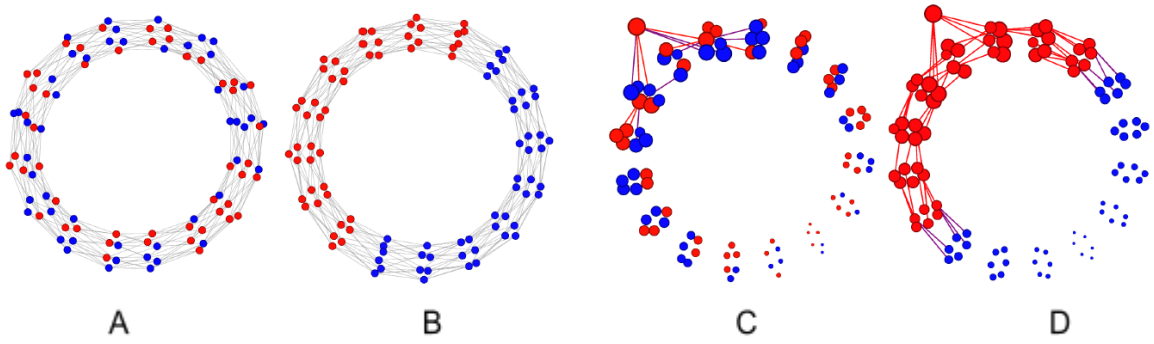


Figure 2 – In the Mass Online Experiments Pilot, subjects were randomly assigned to an integrated (A) or segregated network (B). Red (blue) is conservative (liberal). In A a false message dies out quickly (C), in B it diffuses widely (D).

Moreover, researchers will be able to bring the hugely valuable media and public opinion data that exists in ODISSEI and CLARIAH together with the Mass Online Experiments' tools in innovative new ways. Mass Online Experiments are exceptionally valuable for understanding group dynamics surrounding digital media content because the simulated environment of the mass online experiment mimics that of a social media platform. Through SSHOC-NL, the Mass Online Experiments' tools will be *interoperable* with a broader catalogue of social media and cultural content and it will be possible for Social Science and Humanities scholars to *collaborate* in observing how group dynamics interact with various items from the vast media catalogue, and how these dynamics are influenced by network structures and processes. These are crucial and unaddressed questions for Social Science and Humanities scholars but also for wider society. The interoperability between cultural and media data on the one hand, and the instruments for collecting public opinion and the dynamics of group behaviour on the other will enable researchers to understand just how deep divisions go, as well as their ability to affect contemporary social debates. Such knowledge will provide a basis for dealing with misinformation and conspiracy theories.

To make new strands of research possible, the media catalogue will be expanded (Task 2.3) and made interoperable with public opinion data (Task 4.1), the Mass Online Experiments' tools will be upgraded (Task 3.3) and researchers will be trained and supported in their usage (Task 5.1).

"A cross-media infrastructure that leverages multiple media perspectives is something I long wished for. The dominant reliance on Twitter data risks a biassed perspective on societal phenomena which only an infrastructure like SSHOC-NL can work against."

- Sarah Burkhardt (Media Studies, UvA)

b. Social Inequalities

Social Inequalities are a grand societal challenge because they are systematically *reproduced, reflected and replicated* across society. For example, material inequalities in education and income are reflected in media coverage and cultural representation, and historical inequalities reproduce themselves over many generations. The Social Sciences and Humanities hold data on both historical and contemporary inequalities and taken together can help researchers understand how inequalities are repeated through time and reflected across society.

CLARIAH provides several databases of historical populations, such as the Historical Sample of the Netherlands (HSN). This database includes information on a large representative sample of the Dutch population from 1812 through to 1922, marking births, deaths, marriages, and migrations. HSN is a great example of how historical and sociological research have been brought together to understand long term trends and inequalities in migration (Paiva, Anguita and Mandemakers 2020), fertility (Schellekens and van Poppel 2012), marriage (Jennings and Gray 2017), social mobility (Schulz and Maas 2012) and mortality (Kaptijn et al. 2015). These studies have helped integrate historical evidence into contemporary sociological debates, grounding them in accurate and high-quality analysis. But these studies are only able to make aggregate level comparisons between historical and contemporary trends. In SSHOC-NL, these collaborations will be taken to a new level through linkages between historical and contemporary population databases at the micro-level, thereby enabling research into the choices and mechanisms underlying country-level outcomes.

In 2021, the connection was initiated through direct linkage of the Historical Sample of the Netherlands with the Population Registers maintained by Statistics Netherlands (CBS). This allows for drawing ties with individuals living in the Netherlands today, enabling researchers to consider whether inequalities that were observed many generations ago are still observable within cohorts of descendants today. There is, however, intense interest in improving and widening the historical records that can be linked to the population registers held by CBS beyond a small sample of a single birth cohort within the HSN. In ODISSEI, a powerful new tool for understanding social inequalities has been developed in the form of a whole population network file. This file links all individuals in the Netherlands with all their neighbours, colleagues, family members, and classmates. It provides researchers with a topography of society, with the family connections alone providing an effective family tree for the country (Van der Laan et al. 2021). Using this data, the POPNET project has shown that 87.7% of the Dutch population belonged to the same broader family tree meaning that for any two individuals in this group it is possible to draw a path between them by following multiple parent or child links.

This family tree of the Netherlands was, however, constructed using data records that only stretch back to the 1960s. By extending the data back further, it is possible to see deeper connections and patterns within this family tree and unearth its roots. Doing so will enable researchers to identify which historical lineages are associated with current patterns of segregation in school and work, and the extent to which long-standing social divides persist into the present day. These long-term historical trends also allow sociologists to better understand how a wide variety of inequalities such as educational or labour market outcomes are intertwined with genetic determinants in these very same outcomes, as the linked historical records enable pedigree analysis (De Zeeuw et al. 2019).

SSHOC-NL will ensure that contemporary and historical data collections can be linked (Task 2.2) and that versioning, provenance, veracity and representivity are incorporated into Social Science and Humanities scholars' analytical frameworks (Task 4.3). It will also make the data accessible and interpretable through secure but easy-to-use interfaces that allow researchers to query these complex and rich datafiles (Task 3.2). Finally, it will train researchers on how to use them to answer complex questions on social inequalities (Task 5.2).

"Combining historical and contemporary individual-level data would enable me to examine demographic processes that can sometimes take generations to develop and understand complex patterns at a population level. This collaboration between ODISSEI and CLARIAH would allow researchers to answer entirely new types of questions we did not dare to ask before."

- Willem Vermeulen (Computational Historical Demography, NIDI)

c. Environmental Change

Environmental change has many social and cultural dimensions (Adger et al. 2013). In the face of quickly rising sea levels, there are repeated risks of flooding as well as periods of sustained drought. This is especially urgent in the Netherlands which is physically and socially defined by its water management. Thus far, the problem of *living with water* has been largely approached as a technical, infrastructural problem, in which both the long term developments of water-related practises and their effect on citizens and social groups have not been sufficiently taken into account (Hein 2020). A long-term, socio-cultural perspective is essential to understand the complexity of environmental challenges and for developing *holistic solutions* that deal with institutional barriers, address changes in behaviour and incentives, and consider the role of prevailing power structures (Bressers and Lulofs 2010). To provide this perspective and link individuals with their environment, it is of course necessary to add a geospatial dimension to a wide range of data.

To meet the environmental challenges of the future requires a transdisciplinary perspective and diverse data on the *interdependencies* of natural, technical, social and cultural factors of water and water management as they play out in space and over time (Bucher et al. 2021). New methods of linking and visualising structured and unstructured datasets (Bodenhamer, Corrigan, and Harris 2015, Noordegraaf et al. 2021) allow researchers to detect and investigate correlations between hitherto unconnected phenomena in specific places and times, such as the creation of flood retention basins, the development of housing prices in the area, the income levels of the inhabitants and their experience of the measures. As such, it contributes to the theoretical and methodological framework for applied, analytical history that the complexity of contemporary societal challenges requires (Graham, Milligan, and Weingart 2015).

Simultaneously, the rapid increase in computational power available for the analysis of geospatial data on individuals through the ODISSEI Secure SuperComputer (OSSC) has enabled researchers to produce maps and visualisations of unprecedented resolution (Petrović 2020). This geospatial data not only places individuals in specific places such as the home or work, but also adds a social dimension to geospatial data, indicating population density, poverty, car use, or the installation of solar panels within a given area. For example, ODISSEI has supported the development of the Dutch Opportunity Atlas within the OSSC. A current constraint on progress is that access to detailed *geospatial and planning data* can be subject to complex and often seemingly *impenetrable privacy and intellectual property concerns* that can only be circumvented through sophisticated secure computing architecture and considerable legal and technical expertise. Geospatial data that is linked to individuals is inherently sensitive and so requires highly secure environments. ODISSEI has developed the kind of secure computing system that is vital when considering geospatially detailed data on local neighbourhoods and built heritage, such as street level demographics and risks of flooding or the architectural schematics of historical neighbourhoods (Petrović, Van Ham, and Manley 2022). The result is a detailed atlas of the Netherlands.

SSHOC-NL postulates that by aligning these innovations in geospatial research across ODISSEI and CLARIAH, it will be possible to develop further new insights into the determinants and consequences of environmental change. These insights require a multilevel, interdisciplinary perspective that allows researchers to zoom in and out to understand the relationship between individuals, society and their environment.

The CLARIAH community has made a start with the creation of a geographic information system (GIS) that allows for spatially exploring the historical data on local environmental policy from the early 19thC until the 20thC. Further work is needed to extend this historical GIS back in time to the 18thC and connect it forward to the more recent geospatial data so that researchers can make substantive links across time and space. This requires shared standards to be agreed and mapped between diverse data providers. This historical, macro-level, structural data on environmental policy and the built environment must be linked to micro-level, contemporary data on income levels, housing standards, environmental practises, and personal attitudes towards policies. Scholars will then be able to seamlessly navigate between the macro-level of patterns in environmental change across entire regions and the micro-level of a single source or data point, such as the living labs on biodiversity financed within the NWA. This approach provides an

instrument for researching the complex interplay of local circumstances, technical and policy measures, and the ways in which individuals experience and respond to them.

In SSHOC-NL, an ecosystem will be developed that will not only ensure that common standards are used across the various data collections (Task 4.2) but also that secure compute environments are in place to allow multiple data providers to deposit sensitive geospatial data for research, retain full control over access to that data and how it can be onwardly linked and analysed (Task 1.1). To help researchers navigate the legal constraints on such data, legal and ethical expert support is required in the form of an Ethical Legal Social Implications (ELSI) helpdesk (Task 5.1). Furthermore, once this data has been analysed, easy to use interfaces need to be provided to ensure that only anonymized and aggregated data can be exported for publication (Task 3.2). Achieving these tasks will allow for the social and cultural context of environmental change to be fully understood.

2.3 Coherence of LSRI components

a. The five workstreams

To ensure a future-proof infrastructural ecosystem that supports innovation and interoperability, this proposal will operate across five interdependent workstreams that represent crucial components in the delivery of a world-class SSH research infrastructure.

The *Compute* workstream will work to provide computational solutions for researchers in the social sciences and humanities domain. The domain shares a common challenge in that much of the data analysed must be held securely, either for privacy or copyright reasons. Such security protocols can prevent the use of many generalised compute solutions for handling large or complex data. What is required is a suite of environments that enable intensive data linkage and scaling up, but are also adapted to reflect the sensitive nature of the data. These compute solutions range from distributed processing techniques, which allow for data to be analysed without researchers receiving direct access, to highly secure supercomputing environments for analysing large scale population data.

The *Data* workstream will ensure that high-quality, well-documented and rich data is available to researchers using the infrastructure. The focus of this workstream will be on data sources that are of interest across the domains, including survey data, historical records, text data, experimental data, and multimedia data. This diversity of data is a key strength of the collaboration between the social sciences and humanities and a key aim will be to ensure that these diverse data sources can be brought together in complementary ways to address the challenges set out in section 2.2, such as the need to link historical, structural data on water policy and the built environment to micro-level, contemporary data on individuals' lives in order to understand how a shared history of environmental management shapes individuals well-being and behaviours today.

The *Tools* workstream will ensure that suitable software and tooling solutions are in place to service the community using the infrastructure. The tools are designed to support various elements of the data lifecycle including experimental design, data annotation, automated data enrichment, secure dissemination, and data findability. A key focus in this workstream will be ensuring that innovations are user-friendly and are of use to the broader SSH research community. This encompasses both Technology Readiness Levels and Stakeholder Readiness Levels. Both ODISSEI and CLARIAH have developed cutting-edge tools in recent years and this workstream will seek knowledge exchange between the communities, the cross-pollination of these innovations, and the further methodological canonisation of a core set of tools.

The *FAIR* workstream will help bind the data, computational solutions, and tools together and specifically facilitate the interoperability between data from the social sciences and humanities. It will reflexively navigate between the epistemic rigour of domain and cross-domain FAIRification. The data, computational solutions, and tools developed by ODISSEI and CLARIAH are vast, and for researchers and machines to navigate this landscape will require a highly advanced and ambitious implementation of the FAIR principles. This will be achieved by mapping common standards from the substantive domains and agreeing on shared vocabularies in relation to virtual research environments, data protection, data access, and data quality.

The *Expertise* workstream is where the technical infrastructure meets with the research community at large. Infrastructures are of a socio-technical nature. For researchers to make the most of SSHOC-NL, elements of co-creation and co-design are part of building the infrastructural ecosystem, secured by the research nature of some of the institutions. But, to enable broad adoption, substantial investment is needed in communication, education, training and support. The challenge within this workstream is to simultaneously address students and researchers taking their first tentative steps with digital and computational methods whilst also working with researchers at the cutting edge of their field. Both are vital for the sustainability and vitality of the respective infrastructures.

b. Complementarity of the workstreams

The five components taken together form a coherent and cohesive ecosystem that facilitates the emergence of a new level of SSH infrastructural services based on interoperability of ODISSEI and CLARIAH services, delivering unprecedented economies of scale, diffusing innovations and better enabling interdisciplinary research. The IT architecture of this ecosystem will be developed by the Technical Working Group of SSHOC-NL (see section 4.1) and executed within Task 1.1. This task will bring together components of generic national E-infrastructure providers (such as SURF, the national computing centre for research, and DANS, the national data archive) with specific domain services (such as the ODISSEI Secure Supercomputer or the CLARIAH Media Suite) and make them usable by researchers in combination with their own custom code and configuration. The Social Science and Humanities communities in the Netherlands are teaming with world-class researchers who require such cutting-edge infrastructure to be easy to access and seamlessly integrated in its use. Then the new research lines set out in section 2.2 can flourish.

All five workstreams of the research infrastructure are needed to support the state-of-the-art research conducted by the Netherlands Social Science and Humanities communities. All five workstreams feed and enrich the other four (Figure 3).

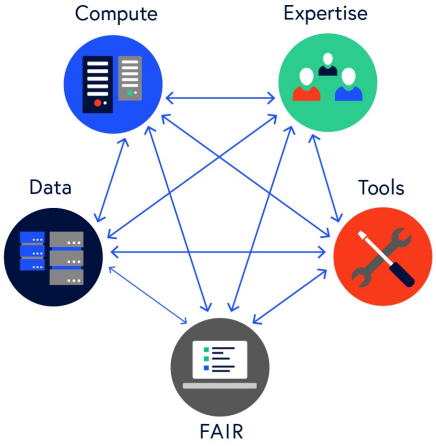


Figure 3 – The five SSHOC-NL workstreams

The solutions developed in the *Compute* environment will open new ways of analysing data and further diversify the data collections available within the data workstream. For example, distributed processing techniques (Task 1.2) will allow for entirely new data sources to be accessed by researchers without infringing data privacy and enable onward linkage using the standards developed in the FAIR workstream (Task 4.2) and enrichment via the annotation tools in the Tools workstream (Task 3.1).

The *Data* workstream identifies new data sources and tackles data challenges that researchers and engineers are wrangling with across the community. For example, the linkage of historical records with contemporary population registers (Task 2.1) necessitates cross-domain standards (Task 4.1) and research environments and tools that can handle historical and contemporary, born digital data simultaneously (Task 1.3).

The *Tools* workstream will develop solutions that are readily usable by the research community and can be deployed across the project to improve analytical insights and data accessibility. For example, the enhancement of annotation tools (Task 3.1) will improve documentation, transparency, and training (Task 5.2) so that an eclectic mix of tools can be applied across a broader range of data in a diverse range of compute environments (Task 1.1).

The developments in the *FAIR* workstream will help guide innovations in annotation and enrichment tools. They will provide controlled vocabularies for data security and access protocols that will allow for scalable compute solutions in handling sensitive or copyrighted data and make these interoperable with the tools workstream. For example, the work on data quality (Task 4.3) will inform the linkage of personal data, in compliance with the General Data Protection Regulation (GDPR), in Data Download Packages (DDP) (Task 1.2), deploy these data and computational solutions on a nationally representative panel (Task 2.2), estimate disclosure concerns of resulting linked data (Task 3.2), and help researchers identify which data is of high quality (Task 5.1).

Finally, none of these workstreams will carry any meaning if they are not orientated towards the research community and if that research community is not actively engaged and supported. The *Expertise* workstream is therefore vital to

the success of the project. For example, through an Ethical, Legal and Societal Implications (ELSI) helpdesk, researchers will be provided with extensive research support in assessing the legal and ethical constraints of accessing particular datasets (Task 2.3), deploying a secure distributed computation infrastructure (Task 1.2), and in handling hundreds of respondents in an experimental setting (Task 3.3).

c. The development framework

To guide the work of the expertise hub on engaging the research community, the project envisions the activities on a matrix along two axes: the Technology Readiness Level (TRL) and the Stakeholder Readiness Level (SRL) (Figure 4). TRL describes the (technical) maturity of a service; SRL reflects the competences of the researchers using the technologies, and therefore the number of researchers able to use the service.

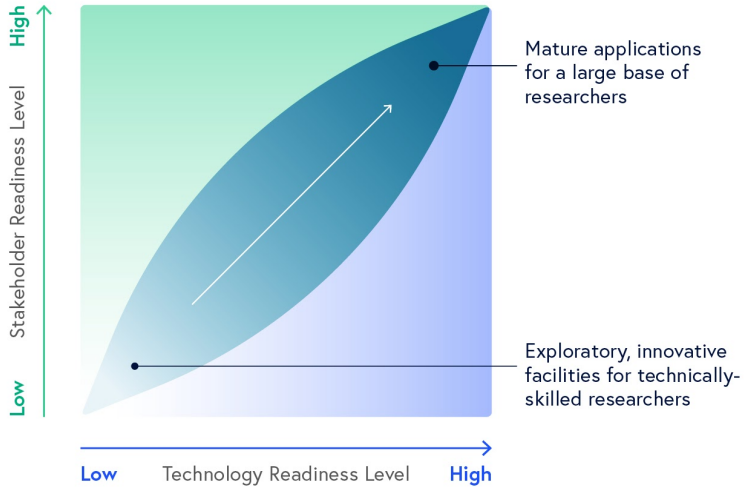


Figure 4 – Technology Readiness Level vs. Stakeholder Readiness Level, and the diagonal on which SSHOC-NL concentrates.

SSHOC-NL concentrates on the diagonal of this matrix. On the one hand, it will create services with a high TRL used by large, heterogeneous communities in a canonical way (high SRL). This is with the aim of elevating services to a TRL of 8 or 9 under the Resource Maturity Classification used in EOSC, such that they can be made available in the EOSC marketplace. On the other hand, ODISSEI and CLARIAH foster a technological and methodological hunger for innovation in the Social Science and Humanities domain (low TRL and low SRL). CLARIAH and ODISSEI will therefore establish a shared infrastructure that facilitates shared innovation in the bottom left-hand corner, and the economies of scale and sustainability in the upper right-hand corner. The innovations must have the potential to benefit both the social sciences and humanities and will be escalated through this matrix to full maturity, and ultimate sustainability and widespread use within the SSH. The matrix can also be used to explain what will not be done. SSHOC-NL will not extensively develop services that are usable for only small or very specific user communities (lower right-hand corner) nor develop standard services (e.g., office-level text and data processors) for which good solutions already exist.

Therefore, the project will also work to develop cutting-edge and ground-breaking innovations in the bottom left-hand corner of the matrix. These are development activities which are highly experimental – with a small user base to start with, and whose journey towards the upper right corner can only be set in motion in the lifetime of this project. For example, the Mass Online Experiments Lab will be developed in this project through a series of ground-breaking pilots led by a small group of world leading experts. The potential of these tools in terms of scale and breadth of application will not be fully realised during this project, but the project will take fledgling innovations and improve their maturity and sustainability for further exploitation after the end of the project. It is also crucial that not only a service is supported in its development towards maturity, but that stakeholders are guided, trained, and supported in its use. Only then can services successfully move to the right of the matrix and full maturity. Beyond the lifetime of the project, these tools will be supported and sustained by ODISSEI and CLARIAH which are anchored in their respective research communities.

d. The five workstreams and new research lines

Section 2.2 detailed three of the new research lines that will be more comprehensively opened through the collaboration of CLARIAH and ODISSEI. These new research lines can, however, only be fully explored by making use of all five workstreams, in an integrated and seamless workflow.

If a researcher studying polarisation in public debates wants to utilise the CLARIAH Media Suite and analyse multimedia data captured by the ODISSEI Media Content Analysis Lab in a secure Virtual Research Environment (sVRE), the SSHOC-NL ecosystem will facilitate the navigation of those two *environments (Compute workstream – Task 1.1).*

If a researcher wants to generate audio data from a representative sample of the Netherlands and link this with survey data, SSHOC-NL will make this possible through the integration of *diverse data* collection workflows *(Data workstream – Task 2.1)*.

If a researcher is conducting Natural Language Processing (NLP) to explore contemporary and historical accounts of environmental impact, they should be able to bring together and combine *tools* from CLARIAH and ODISSEI and not invest large amounts of time and resources on developing and validating new tools *(Tools workstream – Task 3.1)*.

If a researcher is studying trends in criminal behaviour from 1900 to the current day, it should be possible for them to do so using *common vocabularies and ontologies* that are agreed across the domain *(FAIR workstream – Task 4.2)*.

Researchers utilising the services of ODISSEI, CLARIAH, NICAS, or EHRI will not see five workstreams. They will be guided through a myriad of services by *experts*, not diverted or altered by divergent standards across disciplines or differing service agreements *(Expertise workstream – Task 5.1)*.

It is only when these barriers to interdisciplinary research are lowered and the cross-utilization of infrastructures is possible that researchers can leverage all the scientific knowledge at their disposal and be fully empowered to address the infinite complexity in the grand challenges societies are facing.

2.4 Applying consortium, researchers, research groups and other partners involved

Suitability of the consortium

The SSHOC-NL consortium spans the social sciences and humanities and encapsulates a vast breadth and depth of research expertise. The scientific direction of the project is set by the eleven members of the project Steering Board (see section 4.1c), composed of leading SSH scholars from across the domain. They consist of five women and six men, from seven universities from across the Netherlands.

Prof. dr. Pearl Dykstra, the Director of ODISSEI, is Professor of Empirical Sociology at Erasmus University Rotterdam (EUR). From 2016 to 2020, Dykstra held the position of Deputy Chair of the Group of Chief Scientific Advisors to the Cabinet of European Commissioners. She is also an elected member of the Royal Netherlands Academy of Arts and Sciences (KNAW). In 2010, Dykstra was elected as fellow of the Gerontological Society of America, and in 2016 as fellow of Academia Europaea. In 2012 she received an European Research Council (ERC) Advanced Investigator Grant for the research project “Families in context”.

Prof. dr. Antal van den Bosch is Professor of Language and Artificial Intelligence at the University of Amsterdam, has been the director of the Meertens Institute since 2017 and Director of CLARIAH since 2020. He was elected a member of the Royal Netherlands Academy of Arts and Sciences in 2012. Van den Bosch is an esteemed researcher in Artificial Intelligence, memory-based learning, text mining, and language variations. Van den Bosch is a guest professor at CLiPS, the Computational Linguistics and Psycholinguistics Research Centre at the University of Antwerp.

Prof. dr. Jacco van Ossenbruggen is Professor of Human-Centered Data Science at Vrije Universiteit Amsterdam. Van Ossenbruggen is a renowned researcher who led the Human-Centered Data Analytics group at Centrum Wiskunde & Informatica (CWI) in Amsterdam. Van Ossenbruggen is currently active as scientific director of the Civic AI and Cultural AI ICAI labs and is a member of the ODISSEI and CLARIAH management boards, responsible for the implementation of semantic technologies across the SSH domain.

Prof. dr. Julia Noordegraaf is Professor of Digital Heritage in the department of Media Studies at the University of Amsterdam, where she leads the digital humanities research programme and lab ‘Creative Amsterdam: An E-Humanities Perspective’. Noordegraaf currently coordinates the realisation of the Amsterdam Time Machine and is Vice President of the European Time Machine Organisation that employs digital heritage and Artificial Intelligence with the aim of building a simulator for 5,000 years of European history.

Prof. dr. Daniel Oberski is Professor of Health and Social Data Science, with a joint appointment at Utrecht University’s Department of Methodology & Statistics, and the department of Biostatistics and Data Science at the Julius Center,

University Medical Center Utrecht (UMCU). Oberski was awarded with a VENI grant from NWO in 2019. Oberski is a member of the Young Academy of the Royal Netherlands Academy of Arts and Science, and was also a visiting professor at the Joint Program for Survey Methodology (JPSM) at the University of Maryland, teaching experimental design.

Prof. dr. Rens Vliegenthart is Professor of Media and Society at the Amsterdam School of Communication Research (ASCoR), University of Amsterdam. Vliegenthart is also the director of ASCoR and held visiting positions at the University of California, Irvine, and the University of Southern Denmark. Vliegenthart is co-chair of the political communication division of the Netherlands Flanders Communication Association (NEFCA). He is a former member and chair of the Young Academy of the Royal Netherlands Academy of Arts and Science. His research is funded by grants from the NWO (e.g. VENI, VIDI, NWA). He is currently involved in several large infrastructure projects, including the H2020 project OPTED that focuses on the analysis of political texts.

Prof. dr. José van Dijck has been Professor of Media and Digital Society at Utrecht University since 2017 and led the Media Studies department. From 2015 to 2018 she was the first female president of the Royal Netherlands Academy of Arts and Sciences (KNAW). In 2019 Lund University awarded Van Dijck an honorary doctorate for her scientific merits and contributions to the social aspects of digitalisation. Van Dijck is a recipient of the prestigious Spinoza Prize in 2021 with the committee stressing the societal impact of her work and her expertise and clear vision for science and society in the age of digital media.

Prof. dr. Karina van Dalen-Oskam is Professor of Computational Literary Studies at the University of Amsterdam. She is also head of the Department of Literary Studies at Huygens ING (KNAW) and member of the Huygens ING management team. From 2015-1029 she was chair of the international digital humanities organisation ADHO (Alliance of Digital Humanities Organisations). She was PI of the project 'The Riddle of Literary Quality, investigating whether literary quality can be measured with computational means'.

Prof. dr. Gijsbert Rutten is Professor by special appointment of Historical Socio-linguistics of the Dutch language at the University of Leiden. Rutten, who also works as Director of Education at the Leiden University Center for Linguistics (LUCL), conducts research on the historical relationship between French and Dutch, Language variation, and Historical and contemporary sociolinguistics.

Prof. dr. Chantal Kemner is Professor of Biological Developmental Psychology, in particular developmental psychopathology and principal investigator of Youth of Utrecht (YOUth). Chantal is scientific director and main applicant of the Consortium Individual Development (CID). In 2007, Kemner received the VICI grant from the NWO. She is also a member of the Advisory Board for the Dutch National Think Tank foundation.

Prof. dr. Clara Mulder is Professor of Demography and Space at the University of Groningen (RUG). From 2011 to 2021, Mulder was the Chair of the Department of Demography and head of the Population Research Centre. Mulder received the VICI grant from the NWO in 2004 and an ERC Advanced Grant for a project on Family Ties in 2019. She also holds an Honorary Doctorate from Umeå University. She was elected member of the Royal Netherlands Academy of Arts and Sciences (KNAW) in 2017 and is a member of the Board of KNAW since 2022.

The technical implementation of the project will be conducted by a large team of experts and listed and discussed in detail in section 5.1.

2.5 Strategy and embedding

a. The SSH Domain

Much of the Netherlands can only exist because communities and their 'water boards' work together to create a system of water drainage. If done well, this system produces exceptionally fertile land, such as the *polders*. In Dutch, 'to polder' is a verb meaning to ensure broad and inclusive collaboration through interdependent cooperation. In SSHOC-NL, we will polder the SSH landscape and collaboratively create innovations to manage the flow of data, open up new fertile research areas and support a sustainable ecosystem for the whole of SSH. For this to succeed, the project must be inclusive of the whole of SSH, and SSHOC-NL is a unique proposition, representing the collaboration of all LSRI in this domain.

The Open Data Infrastructure for Social Science and Economic Innovations (ODISSEI) is the national infrastructure for social sciences in the Netherlands. It consists of 43 paying member organisations including all social science faculties in the Netherlands. In addition, economics faculties, CBS, public research agencies (Planbureaus), the Dutch Central Bank, and Royal Academy (KNAW) and Research Council (NWO) Institutes are member organisations. This ensures

that ODISSEI is a collaborative infrastructure where cutting-edge social science research is oriented towards answering pressing societal questions.

The Common Lab Research Infrastructure for the Arts and Humanities (CLARIAH) is a world-leading infrastructure in the field of humanities and incorporates Royal Academy Institutes (KNAW) and Humanities faculties from across the Netherlands. It services a diverse range of sub-fields including linguistics, media studies, socio-economic history, and literature. CLARIAH delivers cutting-edge tools and data to this diverse research community and represents an active hub of innovation in the Dutch and European humanities. CLARIAH is coordinated by the KNAW Humanities Cluster (HuC) which was created through the integration of the digital infrastructure of Huygens, Meertens and IISG institutes.

In addition to these national infrastructures, SSHOC-NL also includes the Netherlands Institute for Conservation+Art+Science+ (NICAS) which brings together researchers from the disciplines of conservation, art history, physical science, and computer science to develop a cohesive research programme, with a focus on the origin and life of an object through time. The ultimate goal is a more complete understanding, enriched presentation and optimal preservation of cultural heritage.

Finally, SSHOC-NL includes the European Holocaust Research Infrastructure (EHRI) which aims to integrate the archives and collections of partners from across Europe on items related to the Holocaust. EHRI contributes its generic focus on interoperability in metadata and data of diverse types of archives that each contribute different pieces of complex puzzles.

SSHOC-NL represents a unique, radical and inclusive approach spanning a broad range of research sciences running from economics and psychology through to heritage science and art history. In this respect, this proposal can be said to be truly representative of the domain. It is, however, not a closed club, as SSHOC-NL will be an open and diverse infrastructure with room for engagement with sub-fields which are less well represented in the implementation phase. The principles of FAIR and open science will make SSHOC-NL fertile ground for the broader field, and for interactions with other domains looking to consider a social or cultural dimension for their research.

b. The European Context

The SSHOC-NL project takes its name from the European Cluster Project for the Social Sciences and Humanities which aims to integrate the domain with the European Open Science Cloud. The Social Science and Humanities Open Cloud (SSHOC) is implemented by 22 participating organisations, and also represents collaboration between European Research Infrastructure Consortia (ERICs) in the SSH domain such as CLARIN (Common Language Resources and Technology Infrastructure for Europe), CESSDA (Consortium for European Social Science Data Archives), SHARE (the Survey of Health, Ageing and Retirement in Europe), and DARIAH (Digital Research Infrastructure for the Arts and Humanities). A third of these participating organisations have their administrative seat in the Netherlands and many of these are also represented in SSHOC-NL. This is because the Netherlands is a leader in the development of SSH infrastructure and an unprecedented hub for innovation in the domain. We detail this leading role in terms of the main SSH ERICs.

CLARIN is coordinated from Utrecht University and is led by Executive Director Franciska de Jong. Investments and innovations in CLARIAH have supported the development of CLARIN over the last decade and the realisation of the vision that all digital language resources and tools from all over Europe and beyond are accessible through a single sign-on online environment for the support of researchers in the humanities and social sciences.

Similarly, the Netherlands plays an important role in DARIAH. The Netherlands was the coordinator of the preparation phase of DARIAH (since 2014) and remains a core player with Alice Dijkstra (NWO) currently co-chair of the General Assembly, Andrea Scharnhorst (DANS-KNAW) as Chief Integration Officer, to manage the Joint Research Committee; and Richard Zijdeman (KNAW Humanities Cluster – HuC, IISG, CLARIAH) acting as DARIAH National Coordinator. Many more Dutch researchers and infrastructure staff participate in the DARIAH Working groups.

The European Research Infrastructure Heritage Science (E-RIHS) was added to the European Strategy Forum on Research Infrastructures (ESFRI) roadmap in 2018 to give the international community of heritage researchers access to the infrastructure of international laboratory facilities, research data and the network of knowledge and expertise. The European E-RIHS consortium comprises various national consortia under formation, including E-RIHS Netherlands. As coordinator of E-RIHS.nl, the Cultural Heritage Agency is committed to establishing a national consortium before 2023. NICAS is a key partner within this national consortium and will ensure that the SSHOC-NL project is aligned with and supportive of the development of E-RIHS-NL.

The European Holocaust Research Infrastructure (EHRI) was added to the ESFRI Roadmap in 2018 and aims to integrate the archives and collections of partners from across Europe on items related to the Holocaust. The

infrastructure is hosted by the Netherlands and coordinated by the National Institute for War, Holocaust and Genocide Studies (NIOD). For EHRI, the national research data archive (DANS) manages metadata standards and the integration of metadata from the collections of the different institutions.

The Dutch social science community played a crucial role in establishing the SHARE governance structure at the European level, acting as the first legal and administrative seat of SHARE-ERIC. Centerdata developed the SHARE survey instrument using Blaise, a world leading survey software produced by CBS. Since then, Centerdata has developed a range of innovations that are utilised by several European social surveys such as the Translation Management Tool, the Sample Management Software and the Question Coder. Within the European Social Survey (ESS) project, the Netherlands plays a key role in refining its methodology and enlarging its scope. Finally, DANS has a prominent position in the Consortium for European Social Science Data Archives (CESSDA) which brings together social science data archives across Europe, with the aim of promoting the results of social science research and supporting national and international research and cooperation.

The Netherlands plays a central role in the development of the Generations and Gender Programme (GGP) research infrastructure, as the Netherlands Interdisciplinary Demographic Institute (NIDI) is the coordinator and host of the central hub of the GGP since 2009. The GGP became an ESFRI project in 2021. The Netherlands has also played a pivotal role in the European Values Study (EVS) since its inception. The project is coordinated by a team at Tilburg University and their work over the last three decades constitutes a large part of why the Netherlands holds a reputation of excellence in the field of comparative social surveys. This role is echoed in the prominent and continued collaboration with the World Values Survey. Erasmus University as the coordinating institution of ODISSEI is also a participating organisation in the preparatory phase project for SoBigData Europe, which was added to the ESFRI Roadmap in 2021. SoBigData shares many of the same ambitions as ODISSEI in delivering the tools and resources needed to conduct cutting-edge computational social science research.

In SSHOC-NL further collaboration with international research groups will be actively pursued, especially in the context of the EOSC. As the European Open Science Cloud develops further, tools will be developed in SSHOC-NL with a focus on broad interoperability and the use of international standards advocated by EOSC and SSHOC such that they can be made available to the wider European research community via the EOSC Marketplace. In preparation of this proposal, the SSHOC consortium was notified and was supportive of the proposals aims and ambitions. SSHOC-NL therefore aims to help realise the vision of the European Open Science Cloud by investing in Europe's leading hub for SSH Research Infrastructure.

c. Strategic Embedding

SSHOC-NL is well embedded within several broader national and international developments. ODISSEI, CLARIAH, and NICAS have contributed to the consultation on the National Platform for Open Science (NPOS). Through SSHOC-NL, they will be able to coordinate the transition to open science and to disseminate its importance throughout the SSH domain.

SSHOC-NL will also be closely aligned with the development of the new Thematic Digital Competence Centre (TDCC) for SSH coordinated by DANS. DANS is an important participating organisation of both CLARIAH and ODISSEI and the proposed activities, particularly within the Expertise workstream, are developed to complement the current aims and objectives of the TDCC.

The project will also align with international FAIR implementation efforts including GoFAIR, FAIR Impact and FAIRCORE4EOSC. The FAIR workstream has collaborated closely in the development of the GoFAIR implementation network and development of FAIR implementation profiles.

SSHOC-NL will also seek to develop and support recognition systems for the implementation of open science and develop sustainable and attractive career trajectories for research software engineers, data stewards and data producers. These efforts are aligned with the Recognition and Rewards initiative of the Universities of the Netherlands (UNL) and its objective of supporting FAIR and open science through talent development. To support this, Early Career Researchers and Technicians within the project will be supported in attaining recognition for their contributions and creating a portfolio of open science and FAIR implementation that supports their career development.

3. Impact

3.1 User groups, access policy, and capacity

The community that will be served with the SSHOC-NL infrastructure consists of all researchers in Social Sciences and Humanities in the Netherlands, KNAW, and NWO institutes such as NCSR (Netherlands Institute for the Study of Crime and Law Enforcement), IISG, NIDI, Huygens ING (Huygens Institute for the History of the Netherlands), Meertens Institute, and policy research institutes such as SCP (Netherlands Institute for Social Research), CPB (Bureau for Economic Policy Analysis), PBL (Netherlands Environmental Assessment Agency), Nivel (Netherlands Institute for Health Services Research), and RIVM (National Institute for Public Health and the Environment). The size of the community is considerable. In 2020, over *13,000 scholars* conducted research or taught in the fields of social sciences and humanities at one of the 14 Dutch universities alone, and this number has been steadily increasing over the past decade[2]. The scholarly community outside universities consists of approximately 2,000 further researchers who will be served by the infrastructures in the consortium. The community includes researchers in various seniority levels with diverse scientific interests and methodological approaches.

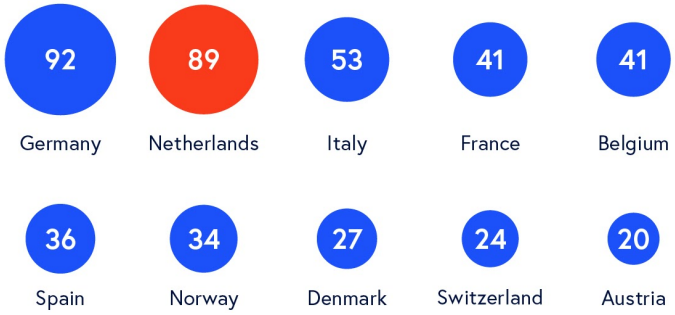


Figure 5 – Absolute number of individual ERC grants in SSH – 2018-2020, top 10

This community is consistently, highly successful in acquiring competitive individual European grants. Between 2018-2020, researchers at Dutch organisations in the SSH domain were granted over 57 ERC Starting grants, 25 ERC Consolidator Grants, and 7 ERC Advanced grants[3]. This impressive achievement means that in this period the Netherlands received the *second most ERC grants in SSH of any European Union country* (see Figure 5). This is in absolute numbers, not in per capita terms. The figures clearly illustrate that the Netherlands is *the hub* for SSH research in Europe and demonstrates the tremendous potential of the research community to create impact and conduct cutting edge research. The scope of granted projects is also very broad: in humanities they cover aspects such as critical study of the texts, images, and moving images produced by tech companies (Dr. Niels Niessen, Radboud University – RU, ERC Stg 2019) or critical discourse oriental Christianity in Europe (Prof. dr. Heleen Murre-van den Berg, RU, ERC Adv 2019). In social sciences, a few examples include research into career development in a changing labour market (Dr. Thijs Bol, UvA, ERC Stg 2020), the well-being of platform workers (Prof. dr. Clara ten Hoeven, EUR, ERC Con 2020) and the influence of news feedback loops on citizens' beliefs and societies (Dr. Damian Trilling, UvA, ERC Stg 2020). The community is also very successful in acquiring prestigious NWO grants. In the period between 2018-2021, SSH scholars were awarded 180 VENI grants, 70 VIDI, and 31 VICI grants.

There has been a recent increase in the number of funded multi-disciplinary SSH projects, made possible by the increasing opportunities afforded by combined use of SSH infrastructures. One example is the Using machine learning to improve the measurement of the European fertility intentions project, a joint effort by NIDI and Meertens Institute (both KNAW) with the University of Groningen. This innovative project deploys natural language processing and machine learning techniques (CLARIAH tools) to analyse the data from open-ended and closed-ended questions from a unique large population-based survey (LISS panel, ODISSEI). The project is expected to contribute to the emerging field of computational demography.

Another example of exciting cross disciplinary collaboration includes Stepfamilies then and now, led by *Prof. dr. Matthijs Kalmijn (NIDI)* and *Dr. Richard Zijdeman (IISG)*. The project focuses on the current and historical influence of disruption in the family on children's lives from the child's point of view. The open access database that is developed in this project facilitates historical and comparative demographic research into the development of family structures, by combining different datasets in different time periods: one at the beginning of the 20th century (HSN/LINKS) and

[2] VSNU 2020
[3] ERC

one at the end (OKiN, in collaboration with CBS). Study of intergenerational reproduction and family structure can be expanded beyond the traditional, contemporary settings.

A major development in interdisciplinary research has been the SCOOP *(Sustainable Cooperation: Roadmaps to a Resilient Society)* project which was awarded €18.8 million in 2015 as part of NWO's gravity program. SCOOP is a research and training centre (for students and practitioners alike), dedicated to the interdisciplinary study of sustainable cooperation in resilient societies. The project connects researchers from sociology, psychology, history, philosophy, public administration, research methods, and statistics and is an excellent example of cross-pollination between the disciplines. Applying an interdisciplinary perspective, the project aims to find solutions for care, inclusion and work. This approach breaks with the tradition to focus on a single domain of cooperation at a time because interventions that secure cooperation sustainability need to take into account the implications that they have for cooperation in other domains. Finally, the project aims to integrate those insights to contribute to theory formation.

SCOOP research projects are intensive users of existing SSH research infrastructure with examples of such projects including: Caring communities: Integrating newcomers into the labour market; Sustainability of public goods in a changing society examines the conditions for the sustainability of public good provisions under changing group compositions, by specifying the role of social norm formation and social norm conflict; Information sharing and social identity establishes how the social mechanisms that drive peer pressure and group fragmentation impact on the rationality of information exchange; Bipolarization and the new media investigates how and why the new media (email, the internet, social media; online communication more generally) may have increased socio-political conflict and bipolarisation reducing the chances of sustainable cooperation in discourse and decision-making using or informed by these new media; A historical lens on family firms and gender equality in the Netherlands, 1900-2020 tests whether family businesses are more conducive to gender equality in the workplace than companies without a family background, using a historical perspective to develop new intervention tools that can be used in promoting gender equality in non-family businesses; Group norms, intrinsic motivation and sustainable energy consumption examines whether more sustainable and cooperative patterns of energy consumption can be driven by group concerns that are internally and/or intrinsically motivated transcending the need to impose (socially and economically) costly forms of incentives or surveillance.

The Consortium on Individual Development (CID) is a forerunner in interdisciplinary research, and received a grant in the first round of the prestigious Gravitation Programme in 2012. All Principal Investigator's in CID have unique and relevant expertise in different areas of developmental research, including behavioural genetics, developmental (neuro)biology, psychiatry, neurocognition, developmental psychology, pedagogical sciences, communications science, mathematics, and neuroimaging, and work together in a series of close-knit, interdisciplinary, projects. CID focuses on the question of why some children thrive and others do not. It is organised in four work packages: brain development, the effect of interventions on child behaviour, generational transmission in families, and the final work package complements all studies with advanced mathematical modelling and animal research.

Open Science plays an important role in CID. The YOUth Cohort has been a trailblazer for open science since its kick-off in 2013. Being a large-scale, longitudinal cohort following children in their development from gestation until early adulthood, YOUth collects a vast amount of data through a unique variety of research techniques and multiple platforms. In order to foster appropriate use of its data, YOUth has invested heavily in setting up an extensive data storage and sharing infrastructure. This infrastructure enables YOUth to produce high-quality, FAIR data while safeguarding the privacy of participants, and to provide safe and sustainable storage and access to the data (described here). The YOUth data infrastructure is the product of intense collaborative efforts of researchers, data managers, IT departments, and the Utrecht University Library, and has been specifically designed to be used broadly. YOUth was at the heart of the development of YODA Social Sciences shared environment for research data that is now being used across Utrecht University, but also (inter)nationally.

The Dutch SSH domain has a vibrant digital and computational community working on using cutting-edge infrastructure to deliver new scientific insights. Efforts to digitalise historic data income panels (HIP-NL: an Historical Income Panel for the Netherlands, dr. Auke Rijpma, dr. Paul Puschmann (RU), prof. dr. Jan Luiten van Zanden (UU)), stock exchange markets (Gauging past performance. Creating financial metadata for economics, finance, and economic history in the Netherlands, prof. dr. Herman de Jong (RUG)), or citizen science based efforts to reconstruct the entire population of Curacao and Suriname (prof. dr. Jan Kok (RU)) are an invaluable source of historic data. Examples bridging history and demography disciplines include the project on Enriching historic death certificate data with other information already available in Open Link Data Format (dr. Sanne Muurling (RU) and dr. Tim Riswick (RU)), and Giants of the modern world that offers an unique longitudinal and intergenerational perspective on the history of heights and health in The Netherlands, making the study of height in relation to the marriage market, careers, and in reproduction possible. IISG in collaboration with Webmapper.nl (a commercial partner) each year aligns the CBS codes for new municipalities to the longitudinal generic Amsterdam Code to ensure that changes in size and names of

municipalities are correctly mapped over time. The open access tool facilitates historical (geo) analysis and is being used by social organisations, such as the Dutch Electoral Council ('Kiesraad'), for their time series on election results.

Lawnotation, an initiative by the Digital Legal Lab led by prof. dr. Gijs Van Dijk (University of Maastricht – UM) makes legal data (legislation and court decisions) and annotation schemes accessible for annotation and analysis, enabling researchers in the field of law, social sciences, and humanities to systematically analyse legal documents. Digital data donation infrastructure (dr. Theo Araujo (UvA)) will deliver a data donation platform allowing individuals to donate their digital trace data to academic research in a secure, transparent, and privacy-protecting manner, expanding the possibilities to study causes, contents, and consequences of (online) communication, behaviour and cultural production and consumption within social media platforms. The Panl initiative (prof. dr. Martin Tanis (VU)) develops an affordable, sustainable, and secure online participant/annotator recruitment platform for Netherlands-based academic researchers. Its applications are numerous, from online experiments or surveys, development of tests, and measurement instruments, to annotations of textual data, artefacts, and images, all relying on the knowledge of Dutch language and cultural context, which is unattainable in international platforms.

Cross-disciplinary collaborations are expanding also towards other fields. Prof. dr. Daniel Oberski, social data methodologist, was recently appointed full professor of data science for healthcare, jointly between the Utrecht University Faculties of Medicine and Social & Behavioral Sciences. Economist dr. Bastian Ravesteijn (EUR) will study how a child's development during pregnancy and the first two years of life affect health and development throughout (Children and (future) parents, supported by prediction and professionals in prevention, to improve opportunity). In co-creation with (future) parents and professionals in midwifery and preventative youth healthcare, a data-driven support instrument will be built and evaluated to support professionals in their shared decision-making with (future) parents.

a. Attracting new users

SSH in the Netherlands can therefore be described as exceptionally vibrant and a leader in Europe and globally. SSHOC-NL aims to help coordinate and strengthen this community through the further integration and interoperability of research infrastructures. This will make the Netherlands SSH community *a magnet for talent* and further strengthen its position as a world leader. Due to the many new possibilities it offers for innovative research and its wide accessibility, SSHOC-NL will help develop a new generation of researchers who are familiar with innovative, interdisciplinary methods at a very early stage, thus creating the best potential for achieving scientific breakthroughs and exciting career opportunities. An important necessary condition for creating such a pool of avant-garde researchers is offering support at every stage of the research process. The *Expertise workstream* delivers dedicated computational and educational support that will help researchers answer novel substantive questions on the boundary of Social Science and Humanities. Training materials – from beginner to advanced level – that bring together a wide variety of disciplines represented by the SSHOC-NL user community will also be realised. The joint effort will be directed to implementing digital tools and data in university teaching and research in collaboration with University Digital Competence Centres, and nurturing cross-fertilization of methodologies, which will further strengthen the community and attract new users.

b. Capacity & access policy

SSHOC-NL will elevate the uptake of the facilities and their impact on conducting cutting-edge research. Services made available through this investment fall into two separate categories based on the degree to which there are limits on the number of access units. For those services, such as data access or the use of new analytical tools, where there are no limits on the number of access units a *wide-access model* will be used by which any authenticated researcher may use these services. In certain instances, access to such services may be restricted to those who can be authenticated as researchers, and constraints placed on their use of the data or tools provided, but there will be no further cost or affiliation restrictions placed on access. There will also be no delineation between users from inside or outside the consortium.

For services with a limited number of access units, such as Microdata Access (Task 1.3) or research software engineering support by the Netherlands eScience Centre (NLesC) and ambassadorships (Task 5.1), SSHOC-NL will support excellence-driven access and it will be allocated through open calls to researchers. These cover a broad range of access units ranging from time on a representative panel (Task 2.2), the time of a research software engineer (Task 5.1), or access to secure data environments (Task 1.3). Calls are overseen by the Grant Coordinator (Task 5.1) and administered through a standardised call process. To reduce burden on researchers and reviewers and to focus only on the quality of the proposal, the Grant Coordinator will ensure that applications are brief (1-2 pages) and in adherence with the DORA principles. The evaluation of applications will involve a clear and transparent process with all calls disseminated across the whole SSH community. Only researchers who are affiliated with a Dutch-based SSH faculty or knowledge institution are eligible to apply for and receive access grants. The call administration process will be managed in alignment with NWO best practises as is currently the case. The projects that are ranked highest for

scientific excellence will be granted use of the facility. The Project Board (see section 4.1) has ultimate deciding power over the allocation of grants.

The Grant Coordinator will adapt and modify calls to ensure that they are adequately targeted to early career researchers, researchers from underrepresented disciplines, researchers working on interdisciplinary projects, or researchers from disadvantaged or underrepresented backgrounds.

3.2 Services to scientific users and impact on the field

SSHOC-NL is a distributed infrastructure that serves all Social Science and Humanities scholars in all research phases in the Netherlands. It promises to become the foremost example for digital infrastructures in Europe and beyond, by showing what an overarching SSH ecosystem can achieve. The reach of SSHOC-NL is unprecedented in the Netherlands, and the depth of the collaboration across the domain is unparalleled internationally. Thus, SSHOC-NL provides researchers in the Netherlands a cutting-edge advantage on an international level, it will advance Dutch Social Science and Humanities research in ways that were impossible before, and it will prove to be a model for other countries aiming to bolster and broaden their scientific infrastructure in the SSH.

a. Secure computational services

One of the major challenges in both Social Science and Humanities research is the processing of the wide variety of available datasets in a secure way. To meet this challenge, SSHOC-NL will offer researchers state-of-the-art computational facilities with virtual research environments that *guarantee Secure, Advanced, Large-Scale Analysis of Data* (SALAD). This is crucial for research that harnesses privacy sensitive data or copyrighted data, but also those who combine a variety of datasets in their research (Task 1.1).

The European General Data Protection Regulation (GDPR) complicates the use of personal data by individual social scientists. SSHOC-NL helps researchers to use new ways of dealing with data in compliance with the GDPR. Through SSHOC-NL, existing secure environments will be made interoperable with PORT software (Boeschoten et al. 2022), enabling researchers to analyse personal data securely and link it to other datasets (Task 1.2).

Furthermore, with the ever-expanding datasets, it is necessary for researchers to make use of a secure supercomputing environment. Through SSHOC-NL, researchers will be able to make use of the extended functionalities of the ODISSEI Secure Supercomputer to conduct analysis on complex linked data including the wide range of administrative data held at CBS. SSHOC-NL will enable a wider variety of research projects to use the secure supercomputing environment for their heavy analyses (Task 1.3).

b. Rich data collections that span generations

The quality of historical and longitudinal data collections in the Netherlands is already a highly valuable asset for the scientific community, but combining datasets from social sciences and humanities as SSHOC-NL will offer scholars the opportunity to *revolutionise the perspective on generational development through distinguishing critical periods, continuity, and secular change*. SSHOC-NL will ensure the expansion of existing datasets and it will enable novel research through the linking and enrichment of datasets. Moreover, the combined Social Science and Humanities perspective will guarantee novel ways of using existing data collections.

A prime example of data collection in the Netherlands is the LISS panel, a true probability sample from households drawn from the Dutch population register. Through SSHOC-NL, researchers will be able to use *LISS panel data* thanks to open calls that are aimed at projects that span Social Sciences and Humanities. Thus, the LISS panel that is already largely used by social scientists, will be employed for *answering new questions evolving around the intersection of SSH* (Task 2.1).

The datasets that SSHOC-NL develops and connects will furthermore stimulate researchers to *connect questions of the present day with a historical perspective* on socio-economic development. Researchers can extend their research either into previous or future generations, by using the enriched Historical Sample of the Netherlands that will be linked to the administrative records at CBS (Task 2.2).

Researchers can make use of the facilities that will be built to *find and access multimedia data* much more efficiently, in a system that will be broadly tailored to be used by both social scientists and humanities researchers (Task 2.3).

c. Tools to transform research

For the challenges of the expanding number of datasets, the increasing variety of types of research data, and the security concerns that become ever more complicated, it is imperative to develop tools to serve SSH research.

Therefore SSHOC-NL will deliver a suite of data enrichment tools to *support ethical and accurate usage* of automatic annotation tools for the entire SSH (Task 3.1).

SSHOC-NL will further enhance the existing POPNET tool that enables *secure research into complex networks* based on CBS administrative records, by making it interoperable with a wide range of Social Science and Humanities data (Task 3.2).

The Mass Online Experiments Suite offers software to *conduct large-scale, real-time, dynamic experiments* with more than 100 participants and will be scaled up and generalised to cater to a wide SSH community (Task 3.3).

d. Data and software

Cooperation in SSH will lead not only to traditional publications, but also to datasets, software, and open-source reusable code, which then becomes available to the open science research community (Task 4.1). SSHOC-NL encourages the users and project partners to work towards open science standards: 'as open as possible, as secure as necessary', and to comply with the FAIR principles. The selection of FAIR resources to be adopted will be declared in a FAIR Implementation Profile, so as to make the community itself FAIR and accelerate convergence across communities in the SSH, but also with other domains.

Data collected within the project is wholly the responsibility of the executing party. SSHOC-NL itself does not hold or process research data, to minimise risks (also see section 5.2a). Data collected by *project partners* during the project must be deposited at a trustworthy digital repository with a CoreTrustSeal (e.g., the DANS SSH Data Station), and its metadata catalogued and published in an agreed machine-readable standard. The FAIR workstream will support intensive implementation of the FAIR principles and enrich existing metadata where possible, and promote the use of controlled vocabularies and domain-wide standards (e.g., European Language Social Science Thesaurus – ELSST).

Scientific outputs from SSHOC-NL will be made freely available, with those in receipt of access grants required to publish all outputs as open access, at least in its green route (i.e., the author makes a pre- or post-print version freely available on a repository). The software that SSHOC-NL builds will be compliant with FAIR Software standards, open source and therefore widely available. In line with the recommendations, source code and derivatives of software will be published via Github or equivalent repository, selecting appropriate licences which promote reusability and linking to archiving services (e.g., Zenodo, Figshare, Open Science Framework – OSF) to assign persistent identifiers (e.g., DOI).

Furthermore, all users of the SSHOC-NL facilities will be required to comply with a *user agreement* that asks them to work towards making their data FAIR. This user agreement will be based on the existing ODISSEI user policy and adapted at the outset of the project to reflect the needs of the Humanities community and partners, and approved by the Project Board. The user agreement will provide clear guidelines on how to make the data FAIR, and researchers will be supported in this by the FAIR workstream. Researchers are required to publish their analytical code in a FAIR way and ensure that it is stored in a repository with a persistent identifier, and reference to both the specific data to which it is applied and any resulting outputs (publications or datasets) that it results in.

e. Expertise to support and inspire

To ensure that researchers from all career stages will be able to use all facilities that SSHOC-NL offers, the infrastructure will have an extensive and easily accessible *SSH Support Centre*.

All researchers working in SSH in the Netherlands will be able to consult the *Social Data Science (SoDa) Team* for free advice on the execution of their cutting-edge projects. All researchers in SSH can moreover apply for grants that will support teaching and research, and consult with an ELSI helpdesk (Task 5.1).

Researchers can make use of the wide range of training materials that will be developed with a focus on interdisciplinary topics and methods. To further *develop their knowledge*, researchers can attend the workshops and events that aim to bring together Social Science and Humanities research (Task 5.2).

Through regular *communications*, researchers can stay up to date with the latest developments in the digital SSH community (Task 5.3).

All technical work within SSHOC-NL will be overseen by the Technical Working Group (section 4.1 & 5.2a).

3.3 Impact on other scientific fields and societal and economic impact

a. Other scientific fields

SSHOC-NL spans the SSH domain, but its influence and impact will be far broader.

There is already close collaboration between the life sciences and the SSH domain. For example, Amsterdam University Medical Center is currently an observer member of ODISSEI as it provides a means to link health data to administrative data held at CBS within a high-performance computing environment. Several projects with a biomedical dimension have already been conducted in ODISSEI such as Genome-Wide Association Studies (GWAS) on social outcomes observed in linked CBS data (de Zeeuw et al. 2021) and numerous LISS projects examining complex health behaviours. ODISSEI has close collaborations with the National Cohort Consortium, resulting in coordinated efforts in survey data collections and data linkage. ODISSEI also aligns FAIR implementation and data linkage efforts with the work of the Consortium on Individual Development (CID). The consortium has developed a strong infrastructural basis that enables the unison of data on social and emotional development from tens of thousands of children and adolescents collected by all existing and two new cohorts in the Netherlands: TRAILS, Generation-R, RADAR, The Netherlands Twin Register, L-CID, and the YOUth Cohort. Their integration offers a rare but exceptionally powerful critical mass that puts the Netherlands at the forefront of child developmental research worldwide. Accrediting the value of this integration, the Social Sciences and Humanities Council awarded additional funding to increase the synergy even further through the development of a metadata system.

Several studies supported and maintained within ODISSEI also participate in these initiatives and operate as bridges between the biomedical fields and social science research. The humanities have also illustrated the potential for spillover in the biomedical domain through projects such as Homo Medicinalis conducted by *Prof. Henk van den Heuvel*, which utilises Natural Language Processing to codify and interpret audio recordings of medical consultations.

There are also large potential spillovers from SSH into the environmental sciences domain. As alluded to in section 2.2, the SSH domain carries vast and diverse data on energy-saving behaviours, attitudes and impacts that are of exceptional value to scholars of the Earth and its resources. This ranges from opinions and values expressed in surveys such as the European Social Survey, the European Values Study or the LISS panel, through to energy-consumption data available via CBS. The environmental sciences are typically not as well acquainted with handling highly sensitive personal data, and the SSH infrastructure can facilitate these linkages, improving the social dimension within their research. Through secure interfaces that provide geospatial social indicators that are linkable with environmental data, SSHOC-NL will allow many researchers studying the Earth and its resources to fully incorporate social and cultural processes into more holistic models and better understand the interdependency between social and environmental systems that is central to achieving a sustainable future and adapting to a changing climate.

The links between SSH and other domains will, however, be intensified and strengthened simply due to the consolidation of the SSH domain itself. This enables research infrastructures in other domains to align with the entire SSH domain by collaborating with SSHOC-NL rather than working with a myriad different SSH infrastructures within the domain. In this respect, the potential for SSHOC-NL to increase the social and cultural dimension in broader scientific research is vast and very overdue.

b. Societal relevance

The themes that can be investigated in novel ways by combining Social Science and Humanities' forces fit seamlessly into core questions of the National Science Agenda, including questions on the way that old and new media impact individuals and society (110), the challenges of inequality (41), and the effects of the built environment on health and wellbeing (78). Furthermore, the research that this infrastructure will enable is essential to achieve the Missions of Horizon Europe, aimed at tackling climate change and achieving sustainable development and the monitoring of the Sustainable Development Goals. By combining forces and stimulating cross-pollination, the ambitions of this infrastructure also underpin the goal of the SSH sector plan: stronger together.

Innovations and investments in the Dutch top sectors will be greatly stimulated by the combined insights from social sciences and humanities that only this large-scale infrastructure can facilitate. The cutting-edge digital methods that combine the enormous datasets from the social sciences with the rich, long-view perspective data of the humanities will elevate our understanding of the social structures in which the Top Sectors operate. From the 'Creative Industry' sector to the 'Health Holland' sector, this SSH infrastructure will thus bolster development to further the Top Sector Missions for the Future. The research made possible via SSHOC-NL will support many aspects of the Future Missions

of the innovation agenda including with regards to water management and environmental change as outlined in the research example in section 2.2, through to research aiming to optimise social earning capacity by revealing previously unseen interdependencies, inequalities and inefficiencies using intensively linked data.

c. Finding and responding to opportunities

To achieve maximum impact, the infrastructure works with a broad variety of societal partners and stakeholders. Contacts with national research institutes, such as the Dutch planning offices, CBS, and government bodies for policy advice are already well established in the social sciences. The humanities, on the other hand, have strong links with institutions in the field of culture, heritage, and media. SSHOC-NL offers the unique opportunity to develop the infrastructure that can bring together ground-breaking research of SSH scholars with relevant stakeholders of an unprecedented variety.

To ensure that stakeholders and societal partners will benefit from the knowledge generated by SSHOC-NL researchers, the infrastructure will form three working groups around its three key research themes (Polarisation, Social Inequalities, and Environmental Change), in which both scholars and representatives of societal partners will meet annually to share, discuss and establish ways in which the outcomes will benefit parties outside of academia. Through working groups, wide-ranging collaborations and consultations, SSHOC-NL will position Social Science and Humanities research at the centre of society, unlocking the potential that cooperation between the social sciences and humanities hold, and promising to become the international example of how to develop an impactful SSH infrastructure.

Finding new opportunities for impact will be an important priority in the development and implementation of SSHOC-NL. The academic output of researchers that use the infrastructure will be monitored by the Project Coordination Team who will guide researchers from research design and stakeholder consultation, through to disseminating results and engaging in discussions with the general public. This will ensure that the Coordination Team is well placed to identify impactful research for further use and dissemination. Once every two months, the Coordination Team will have a dedicated 'Current Affairs' meeting in which wider societal developments will be discussed in relation to the infrastructure, to identify new potential partners and new avenues to meet societal demands. The results of these meetings will be discussed by the Project Board, which will decide on new directions. Events will be organised for researchers to discuss societally relevant research with a wider audience, including monthly SSHOC-Society Seminars and an annual public conference. Furthermore, SSHOC-NL will develop the SSHOC-cast: a podcast aimed at a wide, public audience where researchers discuss their new insights. Each episode will host one researcher from the social sciences, and one from the humanities, who together comment on recent research or societal developments.

4. Organisational and financial aspects

4.1 Organisation and governance

SSHOC-NL governance consists of a Steering Board, a Technical Working Group, a Project Board, a Project Coordination Team, and Task Leaders (Figure 6). These will be set out in a consortium agreement based on DESCA and adapted to ensure compliance with NWO guidelines at the outset of the project. This consortium agreement will remain in place for the duration of the project. SSHOC-NL will not create new permanent governance structures or new legal entities. It is a collaborative project incorporating existing research infrastructures. SSHOC-NL's host partner is Erasmus University Rotterdam (EUR) who acts as the host of the Project Coordination Team and secretariat to all governance bodies.

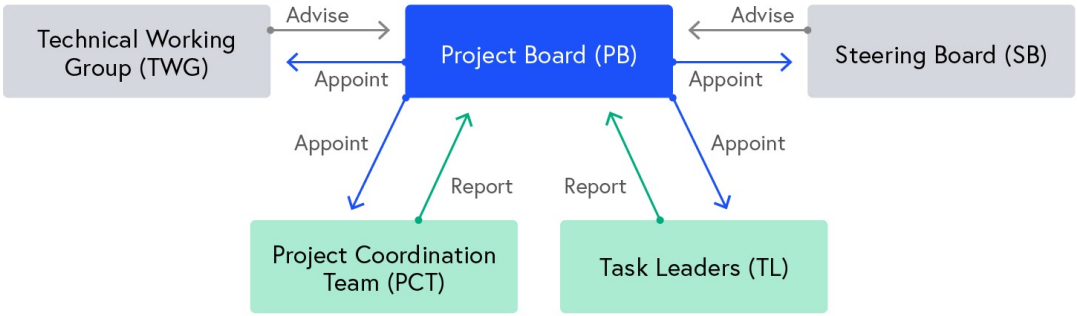


Figure 6 – SSHOC-NL governance structure

a. Project Board (PB)

The Project Board (PB) is the project's highest governance body and is fully mandated by the project partners to operate SSHOC-NL. The group represents the diverse project partners and ensures effective management and operations. It is composed of leading scientists, senior managers, and technical leads. The Principal Investigator (PI) chairs the Project Board. The Project Board has six members who represent the project partners from both the social sciences and humanities:

- Pearl Dykstra (EUR – PI, ODISSEI)
- Antal van den Bosch (HuC – Co-PI, CLARIAH)
- Tom Emery (EUR – Deputy Director ODISSEI)
- Susanne Shatanawi (HuC – Programme Manager CLARIAH)
- Lucas van der Meer (EUR – CTO ODISSEI)
- Roeland Ordelman (NISV – CTO CLARIAH)

The Project Board meets ten times a year and:

- Holds responsibility for the implementation of SSHOC-NL;
- Drafts the annual work plan, annual budget, and annual report;
- Approves the design reports of products and services and allocates budget for further development;
- Approves production versions of products and services before launch and allocates budget for operation and further development;
- Can reallocate budget and tasks in accordance with funding guidelines;
- Can appoint partners who complete tasks on behalf of SSHOC-NL;
- Determines expenditures from SSHOC-NL operating budget and is responsible for procurement;
- Implements funding decisions;
- Manages risk in line with tolerances identified in the project plan and defines risk tolerance for the task leaders, including financial setbacks, delivery delays, quality of partner participation.

When there is a budget issue or time overruns, the Project Board will be responsible for adjusting and approving the original project plan. If disputes arise between partners, the Project Board will be tasked with reconciling differences. To protect the individual interests and autonomy of all parties, unanimity in decision-making is required. The current work programme in this proposal has been approved by all stakeholders and, in the case of further disputes and lack of consensus on how to amend the work programme, the work programme as set out here provides the default parameters for the project and project partners. If disputes or disagreements remain unresolved, It is possible for a joint meeting of the ODISSEI Supervisory Board and the CLARIAH Supervisory Board to be arranged to resolve the dispute.

In addition, the Project Board meetings are attended by the Project Manager from the Project Coordination Team (see next section) who will report on progress across the workstreams, budget overruns or technical delays to the execution of the project, and -where necessary- recommend remedying actions including the potential reallocation of tasks and budgets within the terms of the grant.

In the event that the Project Board identifies a breach by a project partner, the host partner will give formal notice to such Party. If a breach is substantial and is not remedied within 30 days, the Project Board may propose to declare a

Party to be a defaulting party and the consequences thereof, which may include termination of participation in the execution of this proposal.

b. Project Coordination Team (PCT)

The Project Coordination Team (PCT) is responsible for supporting, facilitating, and monitoring the work conducted by the task leaders. This includes coordination, finance, and the secretariat. Its work is included within the SSHOC-NL Expertise workstream. To facilitate synchronisation across the SSH domain, PCT employees will be primarily spread across EUR and the HuC.

The Coordination Team is run by a *Project Manager* who coordinates the implementation of the SSHOC-NL work programme. The Project Manager is appointed by the PB in coordination with the host partner, Erasmus University Rotterdam. The Project Manager oversees the implementation of all tasks across the project, ensures they are adequately coordinated and ensures that the Project Board is provided with sufficient administrative, technical, and financial oversight to fulfil its obligations. The Project Manager is situated at EUR, acts as the secretariat for all project-level meetings and represents the Expertise workstream in the project.

The Project Manager is supported by three senior staff members who address specific areas of operations. They will be recruited at the outset of the project and report directly to the Project Manager.

The *Communications Manager* coordinates the external relations of SSHOC-NL. Many SSHOC-NL constituency organisations have existing work programmes which overlap with the aims and goals of SSHOC-NL such as the use of new methodologies and support for computational social science. The Communications Manager will be responsible for liaising with communication and educational officers across all member organisations, and organising events to raise awareness of SSHOC-NL (also see Task 5.1 – 5.3). The Communications Manager will also be responsible for developing a communications strategy in the first year of the project.

The *Training Manager* will be in charge of coordinating the educational programme to ensure that the infrastructure is usable and used (see section 4.1g, Education & training of users), coordinating training across SSHOC-NL.

In addition to the Project manager and three senior staff members, the Coordination Team will consist of *Coordination Support Staff* to help conduct day-to-day operations. The Coordination Support Staff will support all SSHOC-NL staff and governing bodies with the execution of their work.

c. Steering Board (SB)

The Steering Board consists of nine members and meets at least once a year to evaluate, amend and approve the program of work. The Steering Board directs the Project Board on content-related issues and identifies potential synergies and alignment between the different SSHOC-NL workstreams. It can do so on its own initiative or upon request by the Project Board. Members of the Steering Board are appointed by the Project Board based on their expertise, notwithstanding their affiliation, residency, or nationality. The Steering Board is responsible for the strategic development of SSHOC-NL and looking for opportunities to facilitate innovation and increase cooperation with other initiatives, especially in an international context. Currently, the Steering Board consists of Prof. dr. José van Dijck, Prof. dr. Julia Noordegraaf, Prof. dr. Karina van Dalen-Oskam, Prof. dr. Gijsbert Rutten, Prof. dr. Daniel Oberski, Prof. dr. Chantal Kemner, Prof. dr. Clara Mulder, Prof. dr. Rens Vliegenthart, Prof. dr. Jacco van Ossenbruggen. Meetings of the Steering Board will also be attended by the PI and Co-PI Prof. dr. Pearl Dykstra and Prof. dr. Antal van den Bosch, as well as the Project Manager. The Steering Board makes decisions by simple majority vote.

d. Exploitation & management

All five workstreams are made up of three tasks. Each task has a designated partner as Task Lead who is responsible for executing the work, within the risk tolerance as set out by the Project Board. The task leaders report to the Project Board. On a quarterly basis, the Project Coordination Team will meet with the task leaders from each workstream to discuss the progress of work, and ensure the smooth running of the project. Task leaders are appointed by the Project Board. The task leaders are supported in their daily work by the Project Coordination Team, who are the central coordinating point for all cross-task-related issues. The task leaders are identified in section 5.1.

All partner organisations executing work within this proposal will sign a multilateral consortium agreement, based on the widely used DESCA standard. In it, they agree on matters like the governance structure (as described in this section), ownership (each partner organisation is accountable for performing the work as outlined in this proposal and for procuring its own materials), and intellectual property (the project partners remain owner of their work, but all of that work should be freely available to the SSHOC-NL constituency). No third-party intellectual property is brought into the project. Commercial activities within SSHOC-NL are not permitted.

e. Technical Working Group (TWG)

Each task will include dedicated technical staff and will be responsible for the technical implementation of their work. A Technical Working Group (TWG) oversees the technical implementation of products and services across SSHOC-NL and meets once every quarter. The TWG includes senior representatives of existing national E-infrastructures (SURF, DANS, NLeSC, CBS, HuC-DI), is chaired by CLARIAH's Chief Technical Officer (CTO), and deputised by ODISSEI's CTO. The TWG also includes DANS' CTO and DANS' Data Station Managers for the Social Science and Humanities.

The TWG sets out the direction for the project's IT architecture (section 5.2), ensures alignment between the project and existing software components, creates synergies between the tasks and workstreams, and develops contingency plans. The TWG interacts with individual tasks, supports the iterative development of services, and confirms requirements, programming (connection) standards, specifications, and acceptance tests for the deliverables. The TWG works with each task to amend and update the architecture and individual component requirements and specifications, based on project and user feedback.

The Project Board is advised by the TWG before approving stages of product and service development (see section 5.3).

f. Embedding SSHOC-NL in the domain LSRI's

SSHOC-NL is a collaboration between existing infrastructures, the two largest of which are CLARIAH and ODISSEI. CLARIAH and ODISSEI will assume the responsibility of ensuring the sustainability of data, services and tools developed in the project.

ODISSEI is a consortium of 43 member organisations which include faculties, CBS, public research agencies, and research institutes who are collaborating with the explicit aim of uniting the social sciences and creating a common, national infrastructure for research. ODISSEI was launched in 2016. All ODISSEI member organisations contribute to the operational budget of ODISSEI and ensure its financial sustainability and the continuity of services. ODISSEI is a federated research infrastructure with diverse components. The Microdata Services at CBS offer researchers access to rich administrative data. For computationally intensive projects, this data can be analysed via the ODISSEI Secure Supercomputer (OSSC) which is hosted by SURF. ODISSEI also supports a diverse array of data collections including the elements of the Survey of Health, Ageing and Retirement in Europe (SHARE), the Generations and Gender Programme (GGP), the European Social Survey (ESS), the European Value Survey (EVS), and the EuroCohort project (GUIDE). In addition, ODISSEI provides access to the LISS panel which is a high quality, open access panel where researchers can conduct experiments and link the data ex-post with administrative data at CBS. To support researchers' use of the infrastructure, ODISSEI has developed the ODISSEI Portal which allows researchers to search for data across all ODISSEI data providers. ODISSEI also operates the Social Data Science Team (SoDa) which is a team of computational social scientists who aid researchers in using the infrastructure.

CLARIAH, the Common Lab Research Infrastructure for the Arts and Humanities, is a distributed research infrastructure, i.e. the data, tools and services are distributed across various locations and institutes, connected and made interoperable by a reference architecture CLaaS (CLARIAH as a Service). User-friendly access is offered through a public portal called 'Ineo' (launched spring 2022). CLARIAH provides researchers with access to large collections of digital data and to innovative and user-friendly applications for the processing and analysis of these data. The sustainable backbone of the CLARIAH infrastructure is formed by a group of institutions (KNAW Humanities Cluster (HuC, represented by Meertens and Huygens), National Library (KB), The Institute for the Dutch Language (INT), The national research data archive (DANS), and the Netherlands Institute for Sound and Visions (NISV)). All co-develop components of CLARIAH as part of their mission and are committed to providing services for a period of five years following the end of SSHOC-NL. All CLARIAH developing partners are connected and accountable to the community of Humanities scholars in the Netherlands though the national e-humanities.nl network.

Both have existing and similar governance bodies, based on the recommendations in the OECD Global Science Forum report on the sustainability of Research Infrastructures and the InRoad project recommendations of the European Strategic Forum on Research Infrastructures. The relevant governance bodies in both ODISSEI and CLARIAH unanimously approved the submission of the SSHOC-NL proposal.

g. Education & training of users

The SSHOC-NL project places a great emphasis not only on technical infrastructure and data but on ensuring that the infrastructure is usable and used. The project will be a magnet for talent and in doing so will become a stimulating community of users, with an extensive training, education, and support programme. The Educational programme will be coordinated from Task 5.2 and will be overseen by a dedicated Training Manager who will be a member of the Project Coordination Team. The training programme will be explicitly targeted at the cross-pollination of skills between disciplines and will be executed by project partners familiar with the technical infrastructure and its

potential. Task 5.2 incorporates additional budget for partners who can deliver such training by developing training materials (tutorials, video lectures, workshop formats) and organising events (workshops, summer and winter schools), in addition to their budget for executing their task. All training materials and events will be free to use or attend for all researchers in the Netherlands, capacity permitting and some limited support for travel and accommodation will be provided to ensure that early career researchers and underrepresented groups are able to participate.

h. Key performance indicators

To support the periodic evaluation and for monitoring and reporting on the technical and financial realisation of the project, the following KPIs have been set out (Table 1). These KPIs are developed by the Project Board and approved by the Steering Board. They are monitored and assessed by the Project Coordination Team. Many of the specific targets have yet to be set but these will be added, along with further indicators at the outset of the project. The targets are a distinct mix of qualitative and quantitative measures.

Table 1 – Key Performance Indicators

Indicator	Target	Current
Development		
Support of all SSH Faculties for sustainability	All Social Science and Humanities faculties to be committed to sustainability plan by 2028	All social science faculties are members of ODISSEI
International collaborations	Formal collaboration and alignment with similar initiatives across Europe	None
Access		
Number of users of services	1,000 unique users by 2028	None
Excellence-driven access units provided	50 per year	None
Number of visitors to the SSH Data Station	2,000 visitors per year by 2028	None
Interdisciplinary projects supported	20 per year by 2028	None
Training & Education		
Number of attendees at training events and workshops	600 unique attendees at events per year	None
Number of faculties & DCCs cooperating in educational program	All SSH faculties & DCCs	None
Number of researchers supported by ELSI helpdesk	20 per year	None
Science & Technology		
Number of scientific papers describing the infrastructure	10 per year over 5 years	Nove
Number of software citations derived from SSHOC-NL services	No target set, for monitoring only	None
Number of data citations for data made available via SSHOC-NL	No target set, for monitoring only	None
Impact for Society		
Number of attendees at engagement events	500 unique attendees per year	None
Number of website visits	10,000 visits per year	None
Press mentions in the Netherlands	No current target set	None
Communication		
Number of social media subscribers	Total of 5,000 by 2025	1,723 combined
Number of newsletter subscribers	Total of 8,000 by 2025	None
Attracting New Talent		
Number of international researchers recruited by SSHOC-NL	No current target set	None
Number of projects (such as MSCA/NWO/ERC) using SSHOC-NL	No current target set	None

4.2 Financial aspects

4.2.1. Financial overview of the project

The budget is financed by two streams of revenue (Table 2). First, €14,901,505 is requested from NWO as part of this proposal. The funds will be spent over five years (2023-2027), however, certain tasks operate over variable timeframes within that period due to task specific contingencies and dependencies. The costs are split into the five workstreams and spread more or less equally over the years. The Compute, Data, Tools, and FAIR workstreams account for the capital costs; the Expertise workstream accounts for the running costs. Secondly, for the years 2028-2032, key partner organisations have committed to the continued operation of the core services of the ecosystem generated by the SSHOC-NL project, as in-kind contributions. These in-kind contributions, totalling to €7,076,467, are supported by written affirmations in the annex of this proposal. No cash costs are applicable.

All the costs have been estimated and collated in cooperation with SSHOC-NL partners and are based on provisional estimates regarding the tasks outlined below. The estimates have been assessed and approved by financial controllers at EUR and adhere to the guidelines of the call. Most costs are detailed at a reasonably granular level (Table 3) and are identifiable in the task descriptions outlined in section 5.1 under the corresponding task number. There are three notable exceptions to this rule: CLARIN-ERIC membership fees over 2.5 years (€1.5 million), DARIAH-ERIC membership fees over 10 years (€500k) and Survey Data collections (including ESS, GGP, SHARE, EVS, etc.) (€1.6 million). These costs are large singular material costs that represent the participation of the Dutch SSH community in international research programmes.

A breakdown for the personnel and material costs is provided in Table 4 and 5, respectively.

Table 2 – Total projects costs (k€)

	Capital investment (k€)	Running costs (k€)	Total (k€)
Requested NWO contribution	11,585.993	3,315.512	14,901.505
In kind contribution consortium	-	7,076.467	7,076.467
Cash contribution consortium	-	-	-
Total contribution consortium	-	7,076.467	7,076.467
Total project costs	**11,585.993**	**10,391.979**	**21,977.972**

Table 3 – Project costs per task (k€)

	Personnel (k€)	Material (k€)	Total (k€)
Task 1.1	684.153	640.673	1,324.826
Task 1.2	299.564	180.000	479.564
Task 1.3	425.841	400.000	825.841
Task 2.1		950.000	950.000
Task 2.2	249.637	740.000	989.637
Task 2.3	228.356	984.484	1,212.840
Task 3.1	298.089	1,490.938	1,789.027
Task 3.2	586.368	200.000	786.368
Task 3.3	499.274	100.000	599.274
Task 4.1	936.850	250.000	1,186.850
Task 4.2	638.762	282.090	920.852
Task 4.3	420.914	100.000	520.914
Task 5.1	973.163	515.000	1,488.163
Task 5.2	456.102	250.000	706.102
Task 5.3	820.984	300.263	1,121.247
Total	**7,518.057**	**7,383.448**	**14,901.505**

Table 4 – Personnel costs
The sustainability costs will be spent 2028-2032, after execution of the work programme.

Description	Contri-butor	Capital inv. (k€)	Running costs (k€)	Total (k€)	Yr(s)
Task 1.1 – EUR – Developer 0.5 FTE, NWP WO	NWO	228.051		228.051	5
Task 1.1 – Meertens – Developer 1 FTE, NWP WO	NWO	456.102		456.102	5
Task 1.2 – UU – Project Liaison 0.2 FTE, Senior Scientific	NWO	49.927		49.927	3
Task 1.2 – UU – Research Engineer 1 FTE, Senior Scientific	NWO	249.637		249.637	3
Task 1.3 – CBS – Project Liaison 1 FTE, Senior Scientific	NWO	425.841		425.841	5
Task 2.2 – IISG – HSN Linkage 1 FTE, Senior Scientific	NWO	249.637		249.637	3
Task 2.3 – UU – Project lead 0.92 FTE, Senior Scientific	NWO	228.356		228.356	3
Task 3.1 – VU – Data enrichment 0.7 FTE, Senior Scientific	NWO	298.089		298.089	5
Task 3.2 – UvA – Coordination 1 FTE, Senior Scientific	NWO	249.637		249.637	3
Task 3.2 – UvA – Development 1 FTE, Senior Scientific	NWO	336.731		336.731	4
Task 3.3 – UU – Coordination 1 FTE, Senior Scientific	NWO	249.637		249.637	3
Task 3.3 – UU – Developer 1 FTE, Senior Scientific	NWO	249.637		249.637	3
Task 4.1 – Meertens – Model dev. 1 FTE, Senior Scientific	NWO	425.841		425.841	5
Task 4.1 – VU – Model development 1.2 FTE, Senior Scientific	NWO	511.009		511.009	5
Task 4.2 – DANS – Data Officer 1.5 FTE, Senior Scientific	NWO	638.762		638.762	5
Task 4.3 – UU – Coordination 0.25 FTE, Senior Scientific	NWO	84.183		84.183	4
Task 4.3 – UU – Data linking 1 FTE, Senior Scientific	NWO	336.731		336.731	4
Task 5.1 – Meertens – Coordination 1.2 FTE, NWP WO	NWO		547.322	547.322	5
Task 5.1 – UU – SoDa Member 1 FTE, Senior Scientific	NWO		425.841	425.841	5
Task 5.2 – Meertens – Training Coord. 0.5 FTE, NWP WO	NWO		228.051	228.051	5
Task 5.2 – UvA – Training Coord. 0.5 FTE, NWP WO	NWO		228.051	228.051	5
Task 5.3 – EUR – Communications Man. 0.9 FTE, NWP WO	NWO		410.492	410.492	5
Task 5.3 – EUR – Project manager 0.9 FTE, NWP WO	NWO		410.492	410.492	5
Sustainability – Service usage 0.5 FTE, NWP WO	DANS		228.051	228.051	5
Sustainability – Principal Investigator 1 FTE, Senior Scientific	EUR		425.841	425.841	5
Sustainability – Project manager 1 FTE, NWP WO	EUR		456.102	456.102	5
Total contribution	NWO	5,267.808	2,250.249	7,518.057	
Total contribution	DANS	0.000	228.051	228.051	
Total contribution	EUR	0.000	881.943	881.943	
Total Personnel costs		**5,267.808**	**3,360.243**	**8,628.051**	

Table 5 – Material costs
The sustainability costs will be spent 2028-2032, after execution of the work programme.

Description	Contri-butor	Capital inv. (k€)	Running costs (k€)	Total (k€)	Yr(s)
Task 1.1 – EUR – NISV developer – PartnerHours	NWO	340.673		340.673	5
Task 1.1 – EUR – SURF Architecture dev – ICT	NWO	300.000		300.000	3
Task 1.2 – UU – Software Development – ICT	NWO	150.000		150.000	1
Task 1.2 – UU – Use Cases – Consumables	NWO	30.000		30.000	1
Task 1.3 – CBS – Server Dev – ICT	NWO	200.000		200.000	1
Task 1.3 – CBS – Workshops – Calls	NWO	50.000		50.000	5
Task 1.3 – EUR – SURF – ICT	NWO	150.000		150.000	1
Task 2.1 – EUR – LISS Panel – PartnerHours	NWO	950.000		950.000	5
Task 2.2 – EUR – Survey Infra – Outsourced	NWO	740.000		740.000	2
Task 2.3 – EUR – KB GLAM Collections – PartnerHours	NWO	407.000		407.000	1
Task 2.3 – EUR – KB Integration – PartnerHours	NWO	190.050		190.050	5
Task 2.3 – EUR – NICAS Heritage integration – PartnerHours	NWO	212.921		212.921	5
Task 2.3 – EUR – NISV Integration – PartnerHours	NWO	164.513		164.513	5
Task 2.3 – UU – Material costs – ICT	NWO	10.000		10.000	1
Task 3.1 – EUR – CLARIN Fee – Outsourced	NWO	774.761		774.761	3
Task 3.1 – EUR – INT Developer – PartnerHours	NWO	298.089		298.089	5
Task 3.1 – EUR – INT Project lead – PartnerHours	NWO	298.089		298.089	5
Task 3.1 – EUR – Licences & Software – ICT	NWO	25.000		25.000	5
Task 3.1 – EUR – Training Events – Consumables	NWO	80.000		80.000	4
Task 3.1 – VU – Licences & Software – ICT	NWO	15.000		15.000	5
Task 3.2 – UvA – CBS support – Consumables	NWO	200.000		200.000	4
Task 3.3 – UU – Experiments – Consumables	NWO	100.000		100.000	1
Task 4.1 – Meertens – Licences & Software – ICT	NWO	150.000		150.000	5
Task 4.1 – VU – Licences & Software – ICT	NWO	100.000		100.000	5
Task 4.2 – DANS – Licences & Software – ICT	NWO	150.000		150.000	5
Task 4.2 – EUR – DARIAH Fees – Outsourced	NWO	132.090		132.090	3
Task 4.3 – EUR – SURF Engineer – ICT	NWO	90.000		90.000	2
Task 4.3 – UU – Licences & Software – ICT	NWO	10.000		10.000	1
Task 5.1 – Meertens – Grants and Fellowships – Calls	NWO		500.000	500.000	5
Task 5.1 – UU – Support costs – Consumables	NWO		15.000	15.000	5
Task 5.2 – UvA – Training materials – Consumables	NWO		250.000	250.000	5
Task 5.3 – EUR – Communication costs – Consumables	NWO		300.263	300.263	5
Sustainability – Essential administrative data	CBS		3,705.000	3,705.000	5
Sustainability – Service maintenance	DANS		54.303	54.303	5
Sustainability – Research capacity	Huygens		200.000	200.000	5
Sustainability – Service usage and maintenance	Huygens		364.100	364.100	5
Sustainability – Research capacity	IISG		50.000	50.000	5

Description	Contri-butor	Capital inv. (k€)	Running costs (k€)	Total (k€)	Yr(s)
Sustainability – Service usage and maintenance	IISG		63.732	63.732	5
Sustainability – Digitisation of research content	KB		500.000	500.000	5
Sustainability – Software engineering capacity	Meertens		200.000	200.000	5
Sustainability – Usage, storage, licences and maintenance	Meertens		336.750	336.750	5
Sustainability – Digitisation of research content	NISV		109.464	109.464	5
Sustainability – Software Engineering capacity	NISV		383.124	383.124	5
Total contribution	NWO	6,318.185	1,065.263	7,383.448	
Total contribution	CBS	0.000	3,705.000	3,705.000	
Total contribution	DANS	0.000	54.303	54.303	
Total contribution	Huygens	0.000	564.100	564.100	
Total contribution	IISG	0.000	113.732	113.732	
Total contribution	KB	0.000	500.000	500.000	
Total contribution	Meertens	0.000	536.750	536.750	
Total contribution	NISV	0.000	492.588	492.588	
Total Material costs		**6,318.185**	**7,031.736**	**13,349.922**	

4.2.2. Financial feasibility and sustainability

SSHOC-NL will be developed between years 1-5. During this time period the five workstreams will make significant investments in the improvement and integration of the SSH infrastructural ecosystem. After this initial period of development, during years 6-10 SSHOC-NL partners have formally committed to maintain core services that sustain the infrastructure via in-kind contributions for running costs, supported by written affirmations in the annex of this proposal[4]. These affirmations come from the core central providers within the ecosystem including EUR, HuC, CBS, and DANS and total €7,076,467. These institutions are firmly committed to the continuity of ODISSEI and CLARIAH and the integrated services that the SSHOC-NL project will deliver. The lifespan budget is provided in Table 6. In years 6-10, SSHOC-NL intends to continue investing in further developing the infrastructure. Since these costs are not firmly committed by the partners, they are excluded from Table 6.

The ODISSEI and CLARIAH communities are formal representations of their research community and will be responsible for sustaining the services developed. The Project Board of SSHOC-NL consists of the respective Directors of CLARIAH and ODISSEI and requires unanimity as it is vital that both research infrastructures are committed and invested in the services developed, and willing to assume responsibility for them at the end of the project. Currently, ODISSEI has 43 member organisations that make annual contributions for the maintenance of such services. To increase the number of member organisations and the financial contributions that they make, it is necessary that ODISSEI delivers services that meet the needs and requirements that their research community requires and is willing to pay for. This therefore is the strategic focus within the project and means by which the services in SSHOC-NL are to be accountable and sustainable.

CLARIAH is currently pursuing a governance structure similar to that of ODISSEI through the expansion and formalisation of the e-Humanities network which encapsulates university faculties, knowledge institutions and academy institutes from across the humanities. This formalisation will be completed within the timeframe of SSHOC-NL in order to ensure that together ODISSEI and CLARIAH represent the entire SSH domain. The commitment of the community to supporting such services is currently made through the implementation of CLARIAH and the KNAW Humanities Cluster and represented through their extensive in-kind contributions. By years 11-15, it is not expected that capital investments will be needed and that services will be sustained by the ODISSEI and CLARIAH communities. The SSHOC-NL partners will also contribute directly to the maintenance of the core elements of the infrastructure.

[4] SURF will maintain elements as part of its mandate by the Dutch higher education institutes; therefore, a formal affirmation is not needed.

Since SSHOC-NL will not be building a physical infrastructure but a federated digital and human one, disentanglement, legacy and decommissioning costs are not applicable.

Table 6 – Lifespan costs (k€)

Description	Contributor	Yr 1-5 (k€)	Yr 6-10 (k€)	Yr 11-15 (k€)	Total (k€)
		Requested	Committed	Projected	
Requested NWO contribution	NWO	14,901.505			14,901.505
Capital investments			0.000	0.000	0.000
Personnel	Multiple		0.000	0.000	0.000
ICT	Multiple		0.000	0.000	0.000
Material: Calls	Multiple		0.000	0.000	0.000
Material: Consumables	Multiple		0.000	0.000	0.000
Material: Outsourced	Multiple		0.000	0.000	0.000
Material: PartnerHours	Multiple		0.000	0.000	0.000
Membership fees (cash)	Not applicable		0.000	0.000	0.000
Maintenance	Not applicable		0.000	0.000	0.000
Housing Costs	Not applicable		0.000	0.000	0.000
Dismantling	Not applicable		0.000	0.000	0.000
Running costs			7,076.467	7,076.467	14,152.935
Project manager	EUR		456.102	456.102	912.204
Principal Investigator	EUR		425.841	425.841	851.682
Service usage	DANS		228.051	228.051	456.102
Service maintenance	DANS		54.303	54.303	108.606
Essential administrative data	CBS		3,705.000	3,705.000	7,410.000
Software engineering capacity	Meertens		200.000	200.000	400.000
Usage, storage, licences and maintenance	Meertens		336.750	336.750	
Research capacity	IISG		50.000	50.000	
Service usage and maintenance	IISG		63.732	63.732	
Research capacity	Huygens		200.000	200.000	
Service usage and maintenance	Huygens		364.100	364.100	728.200
Software Engineering capacity	NISV		383.124	383.124	766.248
Digitisation of research content	NISV		109.464	109.464	218.929
Digitisation of research content	KB		500.000	500.000	1,000.000
Total costs		**14,901.505**	**7,076.467**	**7,076.467**	**29,054.439**

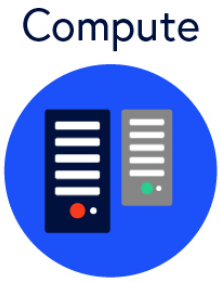
5. Technical aspects

5.1 Technical feasibility

The workstreams

SSHOC-NL will organise its activities across five closely interconnected workstreams. The workstreams are designed to be complementary but independent. Each workstream is divided into three tasks, each representing a distinct program of work. The task leader is responsible for setting out the work programme in conjunction with the Project Board, and ultimately its execution. Project execution is managed using an *agile approach* by which the Task plan is adapted on a regular basis, reflecting developments and newly identified synergies with other tasks. Task leaders report on progress to the Project Board on a biannual basis, with the Project Board amending the task plan and budget if needed. The Technical Working Group (see section 4.1e) will also consult on each task plan to ensure a coherent strategy across the project. The five workstreams and the fifteen tasks are as follows:

Workstream 1. Compute [Roeland Ordelman – NISV & Lucas van der Meer – EUR]



Task 1.1. SALAD [Roeland Ordelman – NISV & Lucas van der Meer – EUR]

In this task, the core computational facilities for the SSH research ecosystem are defined in terms of a suite of secure virtual research environments (sVREs) that ultimately enable Secure, Advanced, Large-scale Analysis of Data (SALAD). SALAD integrates existing distributed generic infrastructure components and services already being used in the SSH ecosystem or national E-infrastructures such as SURF and DANS, into easy-to-use sVREs. Moreover, they facilitate the integration of, or connection to, evolving specialised services and tools within the ecosystem.

SSHOC-NL will build two sets of sVREs, representative for a large number of user needs, taking deliberate advantage of available ingredients in order to "serve the salad that suits a menu" best. In technical terms, containerisation is used to ship the ingredients to the respective sVREs. The two sets sVREs are (1) with a strong emphasis on computationally intensive (e.g., HPC) or secure processing of sensitive or copyright data (e.g., Secure Analysis Environments – SANE), and (2) geared towards browsing, annotation and reflection (e.g., Media Suite, Nederlab).

This task puts the FAIR datasets and tools in a computational perspective, integrating them with provisioning services such as authorization and authentication components, secure access environments, and computational resources. The way in which the building blocks of the shared infrastructure operate and interoperate with existing infrastructure components such as Dutch E-infrastructures or European (cloud) infrastructures – in technical terms, the architecture of SALAD – will be designed, developed, and implemented in close collaboration with SURF under the direction of ODISSEI and CLARIAH. Because of the integration of many components, it is a crucial link between all workstreams (see section 5.2).

> *⇒ The work in this task will be completed through 1.5 FTE for 5 years to be based at NISV and EUR, with a material budget of €640,674*

Task 1.2. Secure Distributed Processing [Laura Boeschoten – UU]

The General Data Protection Regulation (GDPR) grants all persons within the EU the right to an electronic copy of their personal data (e.g., WhatsApp messages or bank statements), in a so-called `Data Download Package' (DDP). Citizens can therefore now consent to donate relevant information derived from their digital traces to trusted researchers, in a manner that is legally mandatory for data processing companies. The PORT software (see Figure 7), developed by Utrecht University, allows researchers to securely analyse DDP data. It achieves this by distributing the processing of individuals' data to the data subjects' own personal computer, where pseudo-anonymized summary statistics can be extracted and collated centrally with thousands of others. This preserves the privacy of research participants (Boeschoten et al. 2022) who retain control over their personal data. First, PORT will be made interoperable with other secure environments such as SANE and the ODISSEI Secure Supercomputer (OSSC, Task 1.3), requiring the implementation of the controlled vocabularies on data security and data access (see Task 4.2), and integration with SURF Research Drive, SURF Research Cloud and SURF Authentication and Authorisation Infrastructure (AAI, Task 1.1). Second, PORT will be extended for use in exploratory analysis, in combination with vaults and extensive de-identification procedures. Third, PORT will be integrated with LISS and other data collections so that participants in these studies can also donate their data (see Task 2.1).

⇒ The work in this task will be completed through 1.2 FTE for 3 years to be based at UU, and a material budget of €180,000

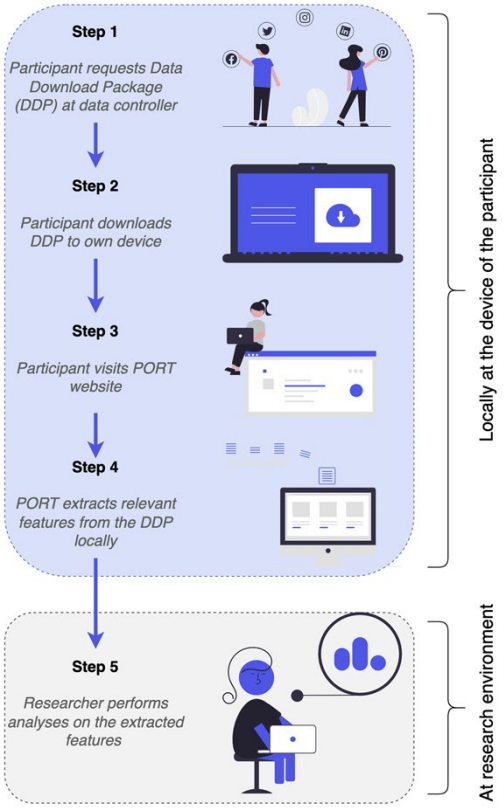


Figure 7 – PORT Software: Step-by-step illustration of the workflow that allows for a privacy-preserving analysis of data download package

Task 1.3. Secure Supercomputing with Administrative Data [Fatima el Messlaki – CBS]

Some of the highest standards of data security in the Netherlands are held by CBS, which operates sVREs for researchers to access the population registers and administrative data on the whole Dutch population. In recent years, the demand for starting new projects within these sVREs has grown enormously and is now nearly 1,000 projects per year, illustrating the exceptional potential CBS Microdata have for the SSH community. Despite the fact that CBS is continuously working on improving processes and user experience, the demand remains high and CBS has to reject projects due to a lack of capacity or because projects are technically or legally infeasible. Improvements across three dimensions are necessary to ensure scalable operations of the sVREs in a safe, secure, and ethical manner.

(1) Providing technical support and implementation to continuously improve the user experience, provide support to researchers who need more computing performance through the ODISSEI Secure Supercomputer (OSSC)[5], automate the application of a microdata project, implementing user authentication and authorization in collaboration with SALAD (Task 1.1), translating the Dutch metadata to English, and providing a way to reuse the result of microdata projects (data products) via a FAIR data repository.

(2) Providing more support on legal aspects by identifying the legal requirements (following the GDPR), implications, and obligations for using the microdata by different kind of organisations, and defining and streamlining legal procedures needed for using Microdata Services that are easy to follow by (non)experienced users. This work will be done in close alignment with the Ethical Legal Social Implications (ELSI) helpdesk in Task 5.1.

(3) Further leveraging the potential of CBS Microdata, by organising more community meetings about the content of data, workshops about working in the CBS sVREs, aiming particularly for cross-pollination projects between SS and H, and making CBS Microdata more findable via the several portals included in SSHOC-NL.

[5] The ODISSEI Secure Supercomputer (OSSC) allows the creation of sVREs which comply with CBS high levels of data security on the Dutch national supercomputer at SURF.

⇒ The work in this task will be completed through 1 FTE for 5 years to be based at CBS, and a material budget of €400,000

Workstream 2. Data [Richard Zijdeman – IISG]

Task 2.1. Online Panel [Joris Mulder – Centerdata]

The LISS panel is at the heart of ODISSEI and provides SSH researchers with the opportunity to field questions to a high quality and precise, representative sample of 7,500 individuals and conduct ground-breaking experiments that improve the understanding of how individuals react to societal changes. In this task, the panel will be maintained and access to the panel will be extended to the entire SSH domain. There will be annual open calls for researchers to conduct interdisciplinary experiments within the LISS panel in cooperation with the Expertise workstream (Task 5.1). For example, there will be specific calls for Audio modules in LISS in which respondents are recorded during the interview and their responses are transcribed and analysed. This creates a representative corpus of natural language for linguists and a wide range of new insights into respondents for survey researchers and sociologists. Further, calls will be specifically targeted at projects involving artefacts or cultural material. In the LISS panel it is possible to ask respondents to evaluate images, text or objects that are commonly found in humanities collections, which provides researchers with exciting new insights into the public perceptions of their collections or their understanding of an object's significance or underlying meaning. This is already common in political and communication science and suggests a fertile ground for collaboration. Researchers can also deploy experiments that see respondents interact anonymously in Mass Online Experiments (Task 3.3), enrich traditional survey data with passively or actively measured sensory data (e.g. accelerometers or voice recorded answers in surveys) and link their research data to the registry data at CBS. All data collected in the LISS panel are made available in the LISS Data Archive[6].

⇒ The work in this task will be completed through a material budget of €950,000 for Centerdata

Task 2.2. Longitudinal Life History Data [Richard Zijdeman – IISG]

SSHOC-NL offers the opportunity to create links between historical and contemporary data collections. This task sets out a highly ambitious work programme to link diverse world leading data collections at the individual level. The 1900-1922 birth cohort in the Historical Sample of the Netherlands (HSN) was linked to the general administrative data on contemporary populations held at CBS in the fall of 2021. This allows researchers to see the social outcomes of the descendants of those in the HSN and track inequalities observed in the early 20th century through to the current day. This task lays the foundation for this work to be rapidly expanded to cover the whole historical population and not just the 78,000 individuals in HSN. In collaboration with the Center for Family History (CBG), the IISG will accelerate the digitization of crucial historical records that document the key life events of Dutch residents in the early twentieth century and make these linkable with the administrative records at CBS. This would enable the current rich pedigree data on family lineages held at CBS to be extended back by several generations, opening up new avenues for research into socio-genetics, social mobility, socio-economic history, and persistent inequalities across generations. What's more, as the linkages, unlike those in HSN, would be population wide they can help enrich life course history research for key sociological studies such as the Survey of Health, Ageing and Retirement in Europe (SHARE)[7] and the Generations and Gender Programme (GGP)[8]. This task will ensure that these longitudinal data streams can continue to serve as a key source of data for the social sciences enabling the understanding of how individuals' life courses unfold over decades, in a cross-national context. But further, it will link these collections to deep family lineages provided through these enriched record linkages. All data generated will be integrated into the ODISSEI Portal and will be linkable with administrative data held by CBS and accessible via relevant sVREs (Task 1.3).

⇒ The work in this task will be completed through 1 FTE for 3 years to be based at IISG, and a material budget of €740,000

[6] The LISS Data Archive (LDA) offers access to researchers worldwide to download research data collected in the LISS panel. The LDA is certified with a core level certification (CoreTrustSeal) of Trusted Digital Repositories.

[7] SHARE is a long standing survey of later life and has been operating since 2004. The Netherlands was a founding member of SHARE and plays a pivotal role in the operation of the survey as much of the technical operations are operated out of Centerdata. Innovations in the Dutch survey research have therefore had a direct deployment pipeline into international studies, and developing into gold standards in the field.

[8] The GGP is entirely operated out of NIDI in the Hague, making the Netherlands the host country for this cross-national study which was admitted to the ESFRI roadmap in 2021. The GGP includes data dating back decades to when it was part of the World Fertility Survey, allowing researchers to understand how people's decision to have children has changed over the last few decades.

Task 2.3. Heritage and Cultural Data [Martijn Kleppe – KB & Jasmijn Van Gorp – UU]

By means of Heritage and Cultural Data, researchers are able to understand societal processes and human behaviour in their historical context. These multimedia data can be textual, structured and/or audio-visual, and range from (historical) ledgers, registrations, and census data to (historical) collections of 'cultural artefacts' such as photographs, films, audio-tapes, newspapers, books, online blogs, and television programs. Some of the materials are open access and freely available and accessible, while others are protected by diverse and dispersed copyrights, portrait rights and related privacy issues. These collections are challenging to use and investigate for researchers, both in terms of Intellectual Property Rights (IPR) and GDPR. The aim of Task 2.3 is threefold: (1) to allow researchers to find heritage and cultural data that currently are available in the dataset registry of the Network Digital Heritage (NDE), and the EHRI Portal, (2) to create access to datasets by developing methods and workflows to enable a generic and straightforward process for making datasets available in the SSHOC-NL ecosystem, (3) enable reuse of newly generated datasets derived from research on collections of cultural and heritage data. Task 2.3 will build generic solutions and protocols that can be applied for future datasets in the SSHOC-NL ecosystem. To create generic solutions, the set-up will be tested in two pilot studies that will take the user-perspective of a researcher as a starting point. The pilot studies will focus on two heritage and cultural data sets: one historical collection from the European Holocaust Research Infrastructure (EHRI) and one yet to be determined collection (e.g., from the National Archive, EYE Film Institute Netherlands, etc). The use cases will be carefully selected to ensure diversity in data forms and to target collections with high added value for interdisciplinary research that spans the social sciences and humanities.

⇒ The work in this task will be completed through 0.92 FTE for 3 years to be based at KB, NICAS, NISV & UU, with a material budget of €984,677

Workstream 3. Tools [Julia Noordegraaf – UvA & Antske Fokkens – VU]

Task 3.1. Machine Learning and AI for Methodologically Sound Data Enrichment [Antske Fokkens – VU & Jesse de Does – INT]

Annotation and enrichments of (raw) data are a prerequisite for many forms of analytics – they are often needed to make data discoverable or explorable on levels otherwise inaccessible, or may be used just to add reflective notes to data. Data enrichments are essential for a huge range of scientific tasks. This task *catalogues, integrates,* and *enhances* machine learning, AI and manual tools for data enrichment. Its main purpose is to provide a suite of data enrichment tools supporting *ethical and accurate usage* of annotation tools across the SSH domain. The task is divided into two closely connected subtasks. Subtask A primarily works on integrating automatic enrichment methods in the infrastructure, facilitating the use of the various tools on data from the *Data workstream* and improving the technological and stakeholder readiness levels of automatic enrichment tools from CLARIAH, ODISSEI, NICAS, and at the NLeSC. We prioritise integration based on (1) low integration effort; (2) high expected utility; (3) interoperability with the environments in the *Compute workstream* and the standards developed in the *FAIR workstream.* In addition, the question of how to deal with the fast changes and advances of the field and trade-off between offering stability and best available tools will be addressed and support to researchers on ethical and legal issues will be closely aligned with the work of the ELSI helpdesk (Task 5.1). Subtask B focuses on manual annotation and enrichment and, in particular, its role in evaluation, validation and training of machine learning methods. It particularly aims at providing straight-forward ways of integrating machine output in a manual enrichment environment for exploring and validating results. The two subtasks together greatly enhance the possibilities of using automatic enrichment tools in a methodologically sound manner for SSH researchers with different levels of technical skills.

⇒ The work in this task will be completed through 0.7 FTE for 5 years to be based at VU & INT, with a material budget of €1,490,940 (inclusive of CLARIN membership fee)

Task 3.2. Assessing disclosure risk in sensitive data [Eelke Heemskerk – UvA & Frank Takes – UL]

With an increased reliance on non-research data for scientific research, it becomes increasingly difficult to balance data access and data protection, both from a legal and ethical perspective. The most advanced current use case of data disclosure risk assessment is the analysis of network data extracted from CBS administrative records. This is because the data is highly sensitive, but also exceptionally complex and intensively linked, making disclosure risks particularly hard to assess with conventional approaches common in the areas of statistical disclosure control. Nevertheless, POPNET[9] has broken new ground in devising disclosure risk assessment methods that ensure a low risk of data disclosure whilst preserving the high scientific value of the data. An initial version of the POPNET tool has been developed to assess the disclosure risks within a specific dataset rapidly and at scale. POPNET collaborates

[9] POPNET handles large and complex network population data from CBS, one of the most secure data analysis environments in the Netherlands.

intensively with CBS to assess the data disclosure risk at multiple points in the data lifecycle and ensures that these risks are managed and minimised, whilst still providing scientists with the ability to analyse the data through a flexible and user-friendly interface. In SSHOC-NL, the POPNET tool will be made *interoperable* with a wider range of Social Science and Humanities data from the *Data workstream*. It will be built using *controlled vocabularies* of data access and data disclosure risk in the *FAIR workstream* (Task 4.2) and will be closely aligned with the ELSI helpdesk (Task 5.1). Functionality will be increased with exploratory *visual data analysis*. The tool will be able to integrate with environments developed in the *Compute workstream* and applied to a wider range of datasets that include sensitive information, including historical population records or complex social media networks.

⇒ The work in this task will be completed through 2 FTE for 3.5 years to be based at UvA and Leiden University, and a material budget of €200,000

Task 3.3. Mass Online Experiments [Rense Corten – UU]

A key challenge in social science is to understand how macro-level social phenomena – e.g., social inequality, social cohesion, or the evolution of culture – emerge as the often unintended consequences of the actions of interacting individuals. Mass Online Experiments – experiments in which more than 100 people interact via an online interface – offer a solution as they allow the study of both micro-level behaviour and the emergence of social phenomena at scale in a controlled fashion. However, such experiments are difficult and expensive to run[10]. This task supports the maturation of a permanent suite of tools for the deployment of Mass Online Experiments and eliminates the fixed costs associated with such experiments for Dutch researchers. The suite will consist of three components: (1) recruitment of participants from i.a. the LISS panel (see Task 2.1), Panl, and from respondents samples drawn from CBS, RvIG or through online advertising, (2) A code library of reusable building blocks that allow participants to interact in a variety of experimental designs programmed in Elixir, Empirica, oTree and Z-tree Unleashed, (3) The durable provision of technical support for researchers in developing experiments by establishing a pool of developers with relevant expertise, (4) The procurement of hardware for experiments. ODISSEI has already supported the piloting of Mass Online Experiment Tools but the existing solutions need to be scaled up and generalised so as to improve accessibility to the wider SSH community.

⇒ The work in this task will be completed through 2 FTE for 3 years to be based at UU, and a material budget of €100,000

Workstream 4. FAIR [Tobias Kuhn – VU]

Task 4.1. Build SSH FAIR & Linked Open Data Model [Ronald Siebes – VU & Thomas Vermaut – HuC]

This task will ensure the implementation of the FAIR principles across the social sciences and humanities by creating an SSH wide vocabulary registry, a registry between entities[11] in the datasets, alignment with tooling for annotation and search (see Task 3.1 on annotation), and Application Programming Interfaces (APIs) for entity management.

Linked data will be the central means to achieve interoperability due to its underlying principles of reuse of existing vocabularies[12]. This task will develop, maintain and map a Linked Open Data model that spans the entire SSH domain, using a framework such as SEMAF[13] or nanopublications. The Linked Open Data model connects the most important SSH vocabularies on scientific concepts, provenance, versioning, accessibility, licensing, and multi-linguality. This will improve the functionality of a range of existing services including the ODISSEI Portal, Ineo and the DANS SSH data station. Access and interoperability of the vocabularies, thesauri and reference data will be implemented through workflows within SALADE (Task 1.1).

Access through space (i.e., maps) and time to vocabularies and reference datasets c.q. thesauri will be facilitated through application-customised deployment of an OpenStreetMap infrastructure[14] and building on previous experience through the HisGIS project. Providing access to finely grained reference data through such a platform will enable the linkage of disparate granularities of data (i.e., house, town, municipality, province, country, etc.) in the SSH through a shared dimension of geographical space throughout time. Researchers are then enabled to investigate patterns of environmental change and social inequality over the *longue durée*. In doing so, spatiotemporal, and domain-specific vocabularies will be made accessible and linkable which will support the linkage and data quality

[10] Participants need to be recruited and incentivized, the experiments have to be designed, programmed, hosted, tested, and supervised.
[11] Semantic entities such as people, places, events, etc.
[12] See the FAIR principles that describe how FAIR datasets necessitate FAIR vocabularies.
[13] SEMAF is a framework for FAIR semantic mappings and crosswalks across scientific domains.
[14] Inspired by the OpenHistoricalMap initiative, as a spatiotemporal extension of OpenStreetMap.

assessment work of Task 4.3. This also relates to the data workstream, given that datasets can only be considered FAIR, if the used (meta)data vocabularies are FAIR themselves.

> *⇒ The work in this task will be completed through 2.2 FTE for 5 years to be based at VU & HuC, and a material budget of €250,000*

Task 4.2. Integrating microservices to increase data FAIRness [Femmy Admiraal – DANS]

A Linked Open Data Model and an SSH wide vocabulary registry (Task 4.1) will allow for the enrichment of metadata with scientific concepts, provenance, versioning, accessibility, licensing, and multi-linguality. The development into meaningful applications and microservices is based around three subtasks.

In the first subtask the various registries from Task 4.1 will be connected and an overlay registry that provides information about the tools that are compatible with a certain dataset will be developed such that data types are linked to tool types. The Research Data Alliance (RDA) has a data type registry prototype that will be operationalised by DANS, thereby implementing the consensus of the global Research Data Management community. This controlled vocabulary of different data types will then be encoded in metadata on deposit or harvest. This will then allow SSHOC-NL to create new supplementary services for tools, analyses, VREs, visualisations, etc. that can be used when a specific type of dataset or sets are found. Finally, a register of crosswalks and content negotiation services will be implemented that supports processes where a datatype may not be suitable for a tool or analysis but can be converted to a type that is.

The second subtask will look to facilitate data harmonisation in an open and FAIR way. Datasets can generally be described in terms of one or more variables, and these variables are presented based on a combination of temporal, spatial, and semantic dimensions with varying granularity (resolution), precision, and accuracy. This subtask will develop a Virtual CUBE that provides a link between datasets that are poorly composed (schema and semantics), a community ontology, and APIs that provide datasets in a semantically and schematically standardised way.

The final subtask aims at reusability of datasets derived from research with complex workflows including annotated collections of cultural and heritage data or data points extracted from dense source material such as large text corpora. Holders of the underlying source data are not in the position to include such derived files in their holdings so they should be stored in a repository for long-term preservation, such as DANS. DANS will provide an API, based on Sword 2.0, that allows integration of data deposits including the corresponding metadata, from external systems into the DANS SSH Data Station, which will greatly enhance FAIR and Open Science in fields handling such complex workflows including heritage science and linguistics.

> *⇒ The work in this task will be completed through 1.5 FTE for 5 years to be based at DANS, and a material budget of €282,090*

Task 4.3. Reconciliation and Linkage Quality [Peter Lugtig – UU]

Provenance and data quality are essential for supporting the reuse of data, especially when various data sources are combined in complex and iterative ways. This task develops a rigorous and methodologically grounded vocabulary to improve the assessment of linked data so that it can be effectively used in research. It includes documenting metadata with data quality indicators, so that users can assess the suitability of linkages before analysis is conducted. Data across the SSH domain can be linked to the gold standard reference data such as administrative data held at CBS or historical population records. This linkage will allow for quality assessment indicators such as representivity, consistency and validity of the data. One of the most urgent use cases for linking and validation are citizen science data. Throughout the humanities, social, and natural sciences, citizens help to drive data collection projects on topics as diverse as documenting dialects (Sprekend Nederland project) or documenting waste in the urban environment (Plasticspotter). ODISSEI has developed a prototype package for linkages to data from spatial datasets or data from CBS in a FAIR manner. However, linkages are now conducted manually. This task will develop tools for semi-automated linkage and data annotation using common software packages (R and Python), that produce quality indicators that are incorporated into metadata. This will help inform researchers not only of a data source's provenance but also of any potential data quality issues.

> *⇒ The work in this task will be completed through 1.25 FTE for 4 years and to be based at UU, and a material budget of €100,000*

Workstream 5. Expertise [Erik-Jan van Kesteren (UU) and Suzan Verberne (UL)]

LiTask 5.1. Support [Kasia Karpinska – EUR & Susan Aasman – RUG]

The Support task bridges the gap between scientists and infrastructure by providing active support in the use of the infrastructures via three activities. (1) The Social Data Science (SoDa) team will serve as an entry point for researchers seeking support in execution of their research projects in e.g., text mining techniques, audio and image processing, and network analyses. The team will offer support for all resources available in ODISSEI and CLARIAH: data, tools, workflows, standards and training materials, building on, and including existing support structures like CLARIAH's Ineo. The projects executed with the support of the SoDa team may result not only in 'traditional' publications, but also in datasets and open-source reusable code, or enhanced publications, creating an active and collaborative open science research community. Emphasis will be placed on the complementarity of skills between SoDa team members, ensuring that all relevant types of expertise are well represented to support researchers across the SSH. In addition, (2) the task will support grants applications that stimulate the use of the infrastructure. Grants are a means to (a) support research and teaching ambassadorships to develop flagship projects across the social sciences and humanities, and (b) stimulate social impact under supervision of the eScience Center. The grants will be operated under NWO guidelines, independently peer reviewed, and administered by a grants officer. The grants will be aligned with those providing access to CBS administrative data (Task 1.3) and the LISS panel (Task 2.1). Finally, (3) the task will establish and operate the Ethical, Legal, and Social Implications desk (ELSI helpdesk), modelled on the implementation in Health-RI. This will be operated by a dedicated specialist who will be responsible for helping SSH scholars in addressing questions about ethical, legal and social dimensions of their research. Given the breadth of the expertise that exists across the consortium, the ELSI helpdesk will be able to draw on a vast breadth of existing knowledge and experience, making sure they are at the disposal of the broader community.

> *⇒ The work in this task will be completed through 2.2 FTE for 5 years and to be based at RUG & UU, and a material budget of €515,000*

Task 5.2. Training [Christian Olesen – UvA & Suze Zijlstra – EUR]

Organising training is essential to ensure actual usage of SSHOC-NL. This task develops training materials – from beginners to advanced level – that bring together SSH researchers in a joint effort to implement digital tools and data in university teaching and research. Over the course of five years the task will deliver 100 tutorials, focusing on facilitating interdisciplinary topics and methods, aligned with international training activities such as the SSH training community operated by SSHOC. The task will organise interdisciplinary hands-on workshops focusing on multiple data types (audio-visual, text, images, social media, geo and spatiotemporal, network, tabular) and qualitative and quantitative methods. Training activities will be designed, implemented, and overseen by the Training Manager who will be based at UvA (0.5 fte) and manage the budget for training materials. Partners who already give workshops on digital and computational methods will be supported in adapting and developing workshops to suit the needs of SSH scholars. The Training Manager will assess the existing training needs in the community and find scholars to support in developing workshops, webinars, and summer schools, in close collaboration with the national research schools in the SSH domain and the Digital Competence Centres (DCCs) at Dutch universities. Emphasis will be placed on a training programme that facilitates interdisciplinary training and the cross-pollination of skills between disciplines, resulting in illustrative data stories such as those generated by the CLARIAH media suite. Training concerns research as well as data management skills. Concerning the latter, the task will collaborate with the TDCC for SSH to ensure that all training materials are sustainably implemented in the local DCCs for widest possible take-up by the current and future community of SSH researchers, students and support staff.

> *⇒ The work in this task will be completed through 1 FTE for 5 years and to be based at UvA & HuC, and a material budget of €250,000*

Task 5.3. Community and Communications [Suze Zijlstra – EUR & Sebastiaan Fluitsma – HuC]

Project management and communications are conducted within this task. The project manager is responsible for the implementation of SSHOC-NL and reporting, both to the Steering Board and NWO. They are also responsible for coordinating the sustainability strategy of the domain and ensuring that the necessary financial, governance and legal instruments are in place to sustain the project's services by the end of the project. The project manager is supported in his/her work by a communications manager who is responsible for implementing a wide range of outreach activities aimed at bringing together the various elements of the project, facilitating cross-pollination and representing the project to the wider research community and general public. The communications manager organises (1) joint engagement events aimed at SSH scholars about shared themes, methods, and techniques, and (2) monthly Task Insights meetings stimulating knowledge exchange and collaboration between tasks. The communications manager will also (3) facilitate community building and interaction, and (4) disseminate information

and project outputs via a newsletter, social media channels, a project website, and a shared events calendar. Finally, (5) they will develop a community platform accessible via the project website, which will provide users with easy access to the different support and training options available.

⇒ The work in this task will be completed through 1.8 FTE for 5 years and to be based at EUR, and a material budget of €300,000

Summary of Development

The workstreams and tasks laid out above are aimed at elevating existing services with regards to the technological and stakeholder readiness levels (TRL and SRL; see section 2.3c). This is the effective impact of SSHOC-NL and is summarised in Table 7. The TRL and SRL of each task is coded between 1 (low) and 9 (high), for both the start of the project and for the expected impact via SSHOC-NL. A coding of 1 for TRL means that the basic principle is observed; for SRL it means that there is no knowledge of the principle in the desired community. In contrast, a coding of 9 for TRL means that the system is proven in an operational environment; for SRL it means that there is widespread and extensive knowledge and use in the community. The codes are displayed via a colour gradient, where bright blue indicates a low score and bright green a high score.

Table 7: Anticipated impact of SSHOC-NL

Low  High

Stream	Task	Activity	TRL Start	SRL Start	Impact on TRL	Impact on SRL
Compute	1.1	sVRE for computationally intensive analysis				
Compute	1.1	sVRE for annotation and reflection				
Compute	1.2	Integration of SURF AAI within PORT				
Compute	1.2	Interoperability of PORT software within SANE				
Compute	1.2	Integration of PORT within LISS panel				
Compute	1.3	Implementation of AAI within the CBS Remote Access Environment				
Compute	1.3	Integration and translation of CBS metadata in SSH Data Station				
Compute	1.3	Legal & Ethical support for administrative data usage				
Compute	1.3	Training for the use of SERVE at CBS				
Data	2.1	Audio Transcription Modules in LISS for NLP research				
Data	2.1	Deployment of Mass Online Experiments in LISS				
Data	2.1	Refreshment of the LISS panel				
Data	2.2	Digitization of Historical Population Cards				
Data	2.2	Enrichment of CBS Network File with Historical Population Records				
Data	2.2	Linkage of Surveys to Historical Population Data				
Data	2.3	Integration of EHRI and NDF metadata in SSH data station				
Data	2.3	Standardised workflows and containers for GLAM collections				
Data	2.3	Cataloguing of annotated GLAM datasets in SSH data station				
Data	2.3	Replicable workflows and data stories				
Tools	3.1	Validation and documentation of existing NLP tools				
Tools	3.1	Use cases for tools that are deployable in SVRE's				
Tools	3.1	Validation of manual annotation tools for ML models				
Tools	3.1	Training for use of manual annotation tools				
Tools	3.2	Validation of disclosure risk algorithm on CBS data				
Tools	3.2	Visualisation tool for complex, privacy sensitive data				

Tools	3.3	Participant management platform for Mass Online Experiments
Tools	3.3	Code Library for standardised Mass Online Experiments
Tools	3.3	Scalable servers for Mass Online Experiments
Tools	3.3	Training materials for Mass Online Experiments
FAIR	4.1	Linked open data model for the SSH Domain
FAIR	4.1	Framework for publishing standards and vocabularies
FAIR	4.1	Deployment of an OpenStreetMap Infrastructure
FAIR	4.1	Publication of Spatio-temporal vocabularies
FAIR	4.2	Overlay registry for data and tools
FAIR	4.2	Virtual CUBE for data harmonisation
FAIR	4.2	Sword 2.0 API for ingest of complex workflows and derivative data
FAIR	4.3	Publication of data quality vocabulary
FAIR	4.3	Enriched ODISSEI Portal Metadata
FAIR	4.3	Python package for semi-automated, quality data linkage

5.2 IT infrastructure

SSHOC-NL aims to enable digital scholarship and the use of computational methods by developing an advanced scholarly IT infrastructure for distributed, data intensive, collaborative research. It must do this whilst finding an optimal balance between local developments (specialised technologies & domain services) that have a quick turnaround on the one hand, and global developments (generic/provisioning services) that require long term sustainability on the other. When categorising the digital methods of SSH scholars on a computational scale it is crucial to consider the diversity of their disciplines. SSH includes diverse forms of analysis from very computationally and data intensive approaches through to more explorative, diachronic, or reflective analysis. This diversity is also evident in the data itself which are very heterogeneous in terms of their medium (multimedia, enabling cross/trans media research) but also in terms of accessibility (open, sensitive, copyright). The "computational" facilities of a shared SSH infrastructure must therefore be flexible and multifaceted to accommodate these broad requirements.

The SSHOC-NL infrastructure spans a wide range of heterogeneous functional components that do justice to the wide range of requirements in social sciences and humanities. It means providing solid computational solutions (Compute workstream), ensuring high-quality, well documented data that is findable and accessible (Data workstream), developing specialised software and tooling solutions for analysing and enriching data (Tools workstream), enabling new knowledge generation by combining information sources within the infrastructure (FAIR workstream) and, last but not least, supporting researchers in making optimal use of the facilities that the infrastructure provides (Expertise workstream).

The activities that are taken up in SSHOC-NL will seek the balance between technology and usage, in terms of technology readiness levels (TRL) –off-the-shelf versus experimental– and stakeholder readiness levels (SRL) –seamless versus disruptive: on the one hand developing promising but innovative, experimental facilities for a “niche” user group, and on the other hand establishing root facilities that can incorporate innovations and in general, benefit large communities in both the social sciences and the humanities (see section 2.3c). The experimental facilities are to be escalated towards high TRL and high SRL.

a. Software management

The SSHOC-NL IT infrastructure consists of three layers and components to which connections will be made (Figure 8).

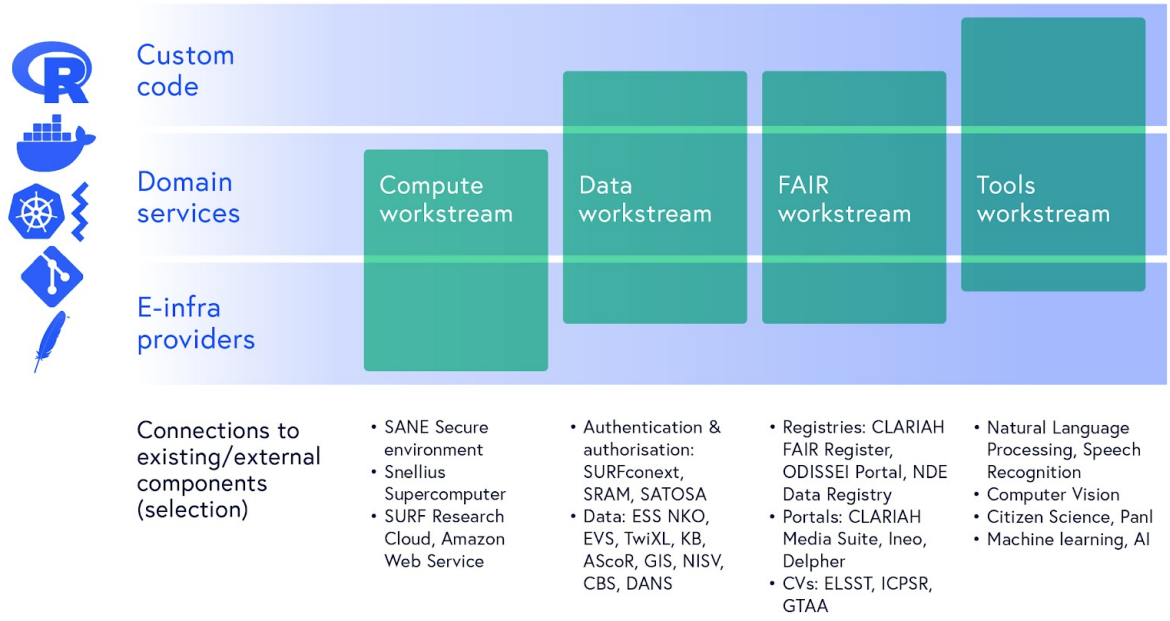


Figure 8 – The three layers of the SSHOC-NL IT infrastructure (in light-blue), the technical workstreams that will be developed (in green), and a selection of connections to existing or external existing components that will be made throughout the project (bottom of the figure). These connections can be made across each infrastructure layer.

At the bottom layer, national E-infrastructures such as DANS, SURF, NLeSC, and HuC-DI provide to SSHOC-NL a relatively small number of existing *provisioning services which will be selected and integrated into the SSHOC-NL infrastructure* (section 'Integrate & Reuse'). These services facilitate core functions such as *supercomputing* (e.g., Snellius supercomputer), *authentication and authorisation* (e.g. SURFconext, SATOSA, SURF Research Access Management – SRAM), and *registries and portals* (e.g., the ODISSEI Portal and Ineo that guide users to the data and services they need). SSHOC-NL manages these components and services by developing an infrastructure architecture that identifies for each component or service the *most diligent* party or parties. SURF is nationally tasked with developing and maintaining a wide array of building blocks that can be used for infrastructures, including SURF Research Cloud, SURF Research Drive, SRAM, and SURFconext. SURF is also heavily involved in the universities' reference architecture, HOSA. Some of the SURF tools are already integrated in the ODISSEI and CLARIAH infrastructures, but others are not yet even considered. Similarly, heritage and research institutes such as HuC-DI, NISV, and KB commit themselves to the provisioning of core infrastructure services such as access to catalogues, storage of research data or the hosting of specialised local cloud infrastructures.

Next to these key provisioning services, infrastructure components, which we call *domain services,* will be *developed and/or maintained* in the workstreams. They serve more specific needs in the respective SSH research domains. These domain services cannot be developed by the national E-infrastructures as, per the E-infrastructures' mandate, they should focus on delivering generic provisioning services. Therefore, domain services must be developed within SSHOC-NL that *use or are built upon* these provisioning services. Such domain services include compute environments (Compute workstream), data connections to datasets hosted by external parties (Data workstream), analysis tools (Tools workstream), or FAIR services (FAIR workstream), hosted and maintained at various places in the infrastructure. Bringing these distributed resources on a functional and organisational level under the umbrella of SSHOC-NL is key to the development of an infrastructure. On an organisational level SSHOC-NL will (1) focus on identifying guidelines and best practises for developing and maintaining (open-source) software developed in the project, (2) publish standardised software documentation (codemeta) via the user portals, and (3) steer towards agreements with local service maintainers on service/maintenance levels.

Finally *custom code* will be developed, typically in proximity of researchers, that is specialised towards specific research domains or questions. The code is written in analysis software packages such as R or using Jupyter Notebooks. SSHOC-NL will aim towards an open architecture that allows the embedding of these experimental services in the infrastructure, while stressing the importance of following standards (interoperability, documentation) as agreed upon in the workstreams and the Technical Working Group (see below). Facilities developed in the Compute workstream operate with national E-infrastructure components but hardly facilitate custom researchers' code. In contrast, because of its applied nature, the Tools workstream deals most with custom code and relatively little with components from national E-infrastructure providers. The Data and FAIR workstream operate in the middle and have a fairly equal part in those two infrastructure layers.

Integrate & Reuse

The existing ODISSEI and CLARIAH infrastructures have been built independently, without having particular joint use cases in mind. This resulted in two infrastructures that share functional components that are sometimes implemented differently or have a slightly different perspective. CLARIAH and ODISSEI will *integrate* their infrastructures on these key functional components, thereby facilitating joint SSH use, increasing the flexibility and efficiency of the infrastructures, and establishing domain wide best practises so that stakeholders such as NICAS and EHRI can be better integrated.

CLARIAH and ODISSEI have already taken the first steps towards the alignment of their infrastructures in exploratory collaborations in pilot projects (e.g. SANE, HSN). This resulted in a better understanding of the requirements of the stakeholders in the ecosystem: different types of scholars working on various levels with datasets and tools, but also data providers, and multiple (local/national/European) infrastructure facilitators. Now it is time to integrate these efforts to facilitate a *common ecosystem* in an infrastructure-agnostic, scalable and sustainable way.

The project will make sure to *reuse existing provisioning services* and build upon the knowledge and infrastructure components developed in predecessors of the projects. Specifically, SSHOC-NL will align and reuse services across the four technical workstreams (see also section 5.2g):

(1) Compute: *secure data workflows and protocols* that are being developed (the ODISSEI/CLARIAH pilot project SANE) to solve IPR/GDPR related (computational) access restrictions at various types of data owners in the infrastructure (CBS focusing on GDPR, KB on IPR, third-parties on competition aspects). In addition, to perform analysis on a large scale (large data sets, computationally intensive processes) SSHOC-NL builds upon *components and workflows for High Performance Computing* (HPC) such as currently facilitated in several ways (Amazon Cloud, Snellius supercomputer or ODISSEI Secure Supercomputer). The alignment with other research infrastructures with similar (or often even higher) computational requirements is evident.

(2) Data: in close collaboration with SURF, SSHOC-NL uses case studies to align the *authorization and authentication* (AAI) for access to datasets being developed and used (SATOSA, SURFconext, SRAM) with the broader set of requirements of the shared CLARIAH and ODISSEI ecosystem, e.g., including higher levels of security or more granularity in authorization levels, both on user level (student versus researcher) and data level (restricting parts of data sets). AAI is applicable across all four technical workstreams.

(3) Tools: exploiting and enhancing (improving TRL/SRL levels) the rich set of tools developed in predecessors of SSHOC-NL made available via *tool registries* (CLARIAH Tool Discovery service via Ineo) and deployed via containers (see containerization below). A tremendous amount of work has been invested previously in the development of toolboxes that suit the different needs of scholars, ranging from tools for data processing (Natural Language Processing (NLP), Audio-video analysis) to data analysis (annotation, workspaces) and workflows (large scale analysis, linking). Some of the core tools, such as services for NLP, speech recognition, computer vision, and annotation are (being) made available as containers that can be shipped flexibly within SSHOC-NL and other infrastructures.

(4) FAIR: Stimulating findability, accessibility, interoperability, and reusability of datasets by continuing investments in preparing, modelling, and enriching data sets and connecting these to various infrastructure levels, including on a European level. Components in this area are (a) *registries* that aim to improve the findability of datasets (e.g., the FAIR Dataset & Vocabulary Registries in CLARIAH, the ODISSEI Portal, the LISS Data Archive, and the NDE Data Registry), (b) *domain portals* that provide access (searching, viewing) to data sets (e.g., Media Suite, Delpher, NederLab, institutional portals), (c) controlled vocabularies (e.g., European Language Social Science Thesaurus (ELSST), Inter-university Consortium for Political and Social Research (ICPSR) Thesaurus, and Common Thesaurus for Audiovisual Archives (GTAA) Tand (d) *end-points* at collection owners that provide low-level access (SPARQL, search API) to data sets for computationally oriented use cases.

Containerisation

To foster interoperability, reuse and sharing of infrastructure developments within and outside the SSHOC-NL context, all functional components will be developed and implemented as so-called *containers* – building blocks with a distinct functionality, such as user authentication. Containers can easily connect with other containers, enabling the creation of scholarly workflows agnostic of any technological differences. SSHOC-NL will use the Docker and Kubernetes industry standards and make the containers *loosely-coupled*, meaning that they can easily be replaced with different containers with the same functionality. The container infrastructure is described with a Common Workflow Language to make it portable and scalable across a variety of software and hardware environments. This setup provides the necessary flexibility with the wide array of available components. Containers will be registered in *container registries* to foster sharing and reuse.

Open infrastructure

The SSHOC-NL infrastructure will be open, so that other developers can easily connect with and extend its services. This will be done through Application Programming Interfaces (APIs). Where possible, open source software will be reused. All our developed software will be published in a FAIR way, as set out by the NLeSC in www.fair-software.nl (section 3.2e). This includes publishing under an open source Apache2 licence. All software is assessed using the Open Source Security Foundation (OpenSSF) checklist. Source code will be managed via an open Git repository which allows for versioning control. In the first year of the project, SSHOC-NL will establish a federated test environment to deploy agile iterations of deliverables.

b. Technical Working Group

The Technical Working Group (TWG), consisting of a ODISSEI's and CLARIAH's Chief Technical Officers (CTOs) and a representation of existing national E-infrastructures (SURF, DANS, NLeSC, HuC-DI), oversees the technical implementation of products and services across SSHOC-NL (also see section 4.1). The TWG optimally aligns the technical elements of the workstreams and tasks and makes sure that existing software building blocks are reused as much as possible.

At the outset of the project, the TWG will finetune the technical implementation plans laid out in the proposal to ensure that integration is established early and reduce implementation risks. In addition, the TWG will continuously document and improve the IT architecture for the domain, closely aligning with relevant reference architectures such as HOSA for Dutch higher education institutions. The technical implementation plans and IT architecture provide guidance on the software engineering process aligned with EOSC Technology Readiness Levels, supported by appropriate standards and exemplars/code implementations. The TWG is specifically tasked with ensuring continuity between developments in the SSHOC-NL work program and those between partners in other SSH infrastructural projects. It is therefore designed to be the central Technical Working Group for SSH where alignment with other initiatives is undertaken. Each SSHOC-NL task includes dedicated technical staff and is responsible for the technical implementation of their domain services. For integration and implementation of the SSHOC-NL provisioning services and providing the necessary connections to relevant domain services, a team of developers is available at the national E-infrastructure providers within the SALAD task (Task 1.1).

c. Data management

SSHOC-NL encourages the reuse of existing data and facilitates the collection of new data. Both the access to existing data as well as the collection of new data are managed by the respective SSHOC-NL partners and *not* by SSHOC-NL directly. Much of the data is *strictly protected* because it contains privacy-sensitive information or because of intellectual property right restrictions and the data partners have the appropriate infrastructure in place to securely process the data. SSHOC-NL not processing these data prevents duplication of facilities and reduces SSHOC-NL risks. At the outset of the project, SSHOC-NL's host organisation Erasmus University Rotterdam will conduct a Data Protection Impact Assessment with all project partners.

While the data management responsibilities for existing and new data thus lie with the SSHOC-NL partners, SSHOC-NL mandates all partners to make their data FAIR and available *'as open as possible, as secure as necessary'* and provides guidelines in the mandatory user agreement. The user agreement will be incorporated into the consortium agreement at the outset of the project (section 4.1). Generated data must be made accessible to researchers via a trusted repository within six months (at most) after a research project is completed. The privacy of respondents must be respected, and de-identification measures must be taken. Where possible, data generated via SSHOC-NL should be available to researchers at no cost. All metadata must be made open,machine-readable and searchable.

SSHOC-NL supports researchers in the re-use of existing data (via the FAIR and Expertise workstreams), including data from the Microdata Catalogue of CBS, the Eye film institute, World Press Photos, NIOD Holocaust survivor testimonies, the broadcast magazine collection of NISV and KB, the LISS panel, the Survey of Health, Ageing and Retirement in Europe (SHARE), the Generations and Gender Programme (GGP), and the Historical Sample of the Netherlands (HSN), and many others.

Both existing and new data that are part of the SSHOC-NL ecosystem will be documented using the appropriate community standards (section 3.2a). Survey data will be documented according to the standards already in use by the respective surveys. This will most likely be either DDI Lifecycle or DDI Codebook depending on which is most applicable to the design of the survey. The HSN is documented using the Intermediate Data Structure (IDS) and CBS microdata is documented using the metadata standard for statistical agencies SDMX.

SSHOC-NL will expose annotations as W3C WebAnnotations, which act as a bridge between the data and structured information in the knowledge graph represented as RDF/LOD. Annotations will be stored in annotation stores at multiple centres.

d. FAIR

The FAIR data principles are regarded as an explicit 'driving force' of the SSHOC-NL activities: exactly the ability to provide FAIR data and tools will bring a shared infrastructure that enables the envisioned innovative research that crosses domains and data boundaries. This is reflected by having a FAIR workstream which is led by one of the authors of the original FAIR paper (Wilkinson et al. 2016), dr. Tobias Kuhn, and via the close connections to European initiatives such as Go-FAIR, FAIRsFAIR and EOSC. SSHOC-NL permeates the FAIR principles by putting them at the core of the Data workstream, by emphasising interoperability and reuse for tool development in the Compute workstream, and by supporting researchers to follow these principles in their projects via the Expertise workstream. The user agreement (see section 3.2e) will require all SSHOC-NL projects (e.g., research pilots or access grants) to include a data management plan that lays out the documentation and archiving strategy of the project and ensures FAIR compliance. Data stewards within the Expertise workstream will help researchers in developing appropriate data management plans. The metadata will contain the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, scientific concepts, provenance, versioning, accessibility, licensing and multi-linguality. Within the Technical Working Group for software management, the SSH Data Station Managers (Femmy Admiraal & Ricarda Braukmann) monitor and steer compliance with the FAIR principles throughout the project.

e. Ethical data use

All legal and ethical considerations will be overseen by the SSHOC-NL Data Officer. The Data Officer will be the current Data Protection Officer at EUR as host institution. They report directly to the Project Board. The Data Officer will evaluate the annual program of work for ethical risks and ensure that sufficient oversight is in place. The Data Officer is responsible for approving the user agreement (section 3.2). The Data Officer will also check that ethical review boards of SSHOC-NL partners watch over activities conducted by partners and advise these boards where the specific nature of SSHOC-NL data or facilities makes ethical reviews complex or atypical. The Data Officer also evaluates proposals in response to calls for ethical considerations and ensures the highest ethical standards are applied. The Data Officer is supported in their work by the ELSI helpdesk team (Task 5.1).

Across SSHOC-NL, data collection within tasks must be completed with explicit prior consent of the data subject that adheres to the GDPR. This means that consent must be informed, explicit, and retractable. Informed consent must contain the nature, scope, context, and purposes of processing and the severity of the risks to the data subjects' fundamental rights. If the data processing entails potential risks to the data subjects' rights and freedoms, they will be made aware of these risks during the informed consent procedure. Any research including children as data subjects must include informed consent of the parent or guardian, and to the extent that it is possible, the informed consent of the child. Research data that are collected as part of SSHOC-NL will be minimised to only that which is necessary and proportionate to the research question at hand. All research data containing personal information will be pseudonymised, securely stored, and subject to the principle of privacy-by-design. This states that where the possibility to enhance the level of data protection is afforded to data subjects, SSHOC-NL will apply such measures by default rather than just considering them or making them available as an optional extra.

f. Training and learning

SSHOC-NL will facilitate (technical) knowledge about the infrastructure components such as how they work and interoperate, and what opportunities or limitations are (e.g., tool criticism). Especially the exchange of this knowledge via *documentation* and *training and learning modules* are key for the successful integration of SSHOC-NL features in the daily routines of researchers. As this exchange requires a close collaboration between technical experts and user representatives, we will build upon experience from the ODISSEI and CLARIAH projects with knowledge exchange via various methods, ranging from the organisation of workshops to the development of training videos, in collaboration with the Expertise workstream.

g. IT costs

Classified IT costs are those that are spent on IT, but not on personnel. Budget for SURF is reserved for assisting SSHOC-NL in designing and developing SALAD and for sitting in the Technical Working Group (total €300k). CBS allocated €150k for SURF and €200k for CBS for supporting the improvement of their technical server infrastructure. Task 4.3 asked SURF to provide an engineer for FAIR data linking (€90k). Task 1.2, 2.3, 3.1, 4.1, 4.3 and 4.2 budgeted €600k for hiring software developers and for licensing fees. The total IT costs account to €1,350k.

h. Use of existing national E-infrastructures (SURF, eScience, DANS)

SSHOC-NL aligns with, extends, and integrates existing national E-infrastructures with the aim to establish a healthy and vibrant research ecosystem that fosters collaboration and sharing of expertise. The aim is to ensure a balanced allocation of resources that fits in with the natural responsibilities of participants in the E-infrastructures ecosystem. For example, the ODISSEI Supercomputer facility (OSSC) is built upon the existing national supercomputer Snellius of SURF. The SSH Linked Open Data Model that SSHOC-NL develops will be distributed amongst a large number of services such as the data architecture of the DANS SSH datastation, the CESSDA data catalogue, and the EOSC marketplace, CBS and many other repositories. The SoDa team is trained by NLeSC, experienced software developers of NLeSC can be insourced for tackling complex problems. A successful harmonisation of SSHOC-NL and the national E-infrastructures is accomplished by having SURF, DANS and NLeSC represented in the Technical Working Group.

5.3 Risk analysis and mitigation strategy

Like any other enterprise, SSHOC-NL will be confronted with some risks, which can be particularly significant at the interface between implementation and operational phase. A simple risk management system will be implemented, checked quarterly, and kept up to date by the Project Board. Each identified risk has a risk owner, who is responsible for monitoring the development of the risk, estimating the probability of the risk and the severity of the effects in case the risk materialises. The risk owner updates the risk assessment system whenever needed and the Project Board will perform bimonthly audits of the risk register and flag problems to the Project Manager.

Table 8 – Risk Analysis

Risk factors	Risk owner[15]	Probability	Impact	Mitigating actions
Financial				
Financial commitments limited to term of project	PB	Moderate	Moderate	Stakeholder engagement & sustainability strategy developed from year 1
Lack of budgetary compliance	PCT	Low	Moderate	Pre-proposal Financial, Legal, Accounting and Tax (FLAT) check at EUR
Organisational				
Dysfunctional governance	PB	Moderate	High	Annual evaluation by PB & SB
Dysfunctional coordination team	PB	Low	High	Quarterly Evaluation by PB
Insufficient resources for tasks	PCT	Low	Moderate	Bi-Monthly updates on budgets and progress by task leaders
Insufficient commitment from task leads	PCT	Moderate	High	Consortium Agreement signed
Poor integration within international initiatives	PB	Low	Low	Responsible PB member assigned
Technical				
Delay in delivery of platforms or tools	PCT	High	Low	Each task to develop contingency plans at project outset
Limited connections within SALAD	PCT	Moderate	Moderate	Utilisation of standard codings and classifications where possible; shared Technical Working Group; iterative approach
Low demand for analytical support	PCT	Low	Low	Community consultation and awareness raising in first 6 months
Legal & Ethical				
Inadvertent personal data disclosure	PB	Very Low	High	Transparent and robust disclosure checks by Data Controllers
Security breach	PB	Very Low	High	Pen testing, continuous monitoring of security (logging) and training of researchers in the context of (privacy) security must reduce those risks.
Scientific malpractice using SSHOC-NL	PB	Low	High	Code of Ethics embedded in the user agreement signed by all users

[15] SB = Steering Board, PB = Project Board, PCT = Project Coordination Team

6. Literature and other relevant information

6.1 Glossary

Acronym	Long
AAI	Authentication and Authorisation Infrastructure
API	Application Programming Interfaces
ASCoR	Amsterdam School of Communication Research
CBS	Statistics Netherlands
CESSDA	Consortium for European Social Science Data Archives
CID	Consortium Individual Development
CLaaS	CLARIAH as a Service
CLARIAH	Common Lab Research Infrastructure for the Arts and Humanities
CLARIN	Common Language Resources and Technology Infrastructure for Europe
CPB	Bureau for Economic Policy Analysis
CWI	Centrum Wiskunde & Informatica
DANS	KNAW Data Archiving and Networked Services
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DCC	Digital Competence Centre
DDP	Data Download Package
DESCA	Development of a Simplified Consortium Agreement
E-RIHS	European Research Infrastructure Heritage Science
EHRI	European Holocaust Research Infrastructure
ELSI	Ethical Legal Social Implications
ELSST	European Language Social Science Thesaurus
EOSC	European Open Science Cloud
ERC	European Research Council
ERIC	European Research Infrastructure Consortium
ESFRI	European Strategy Forum on Research Infrastructures
ESS	European Social Survey
EUR	Erasmus University Rotterdam
EVS	European Values Study
EYE	Film Institute Netherlands
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
GGP	Generations and Gender Programme
GTAA	Common Thesaurus for Audiovisual Archives
HPC	High Performance Computing
HSN	Historical Sample of the Netherlands
HuC(-DI)	KNAW Humanities Cluster (-Digital Infrastructure Department)
Huygens ING	Huygens Institute for the History of the Netherlands
ICPSR	Inter-university Consortium for Political and Social Research
IISG	International Institute of Social History
INT	Institute for the Dutch Language
IPR	Intellectual Property Rights
KB	National Library of the Netherlands
KNAW	Royal Netherlands Academy of Arts and Sciences
KPI	Key Performance Indicator
LDA	LISS data archive
LISS	Longitudinal Internet studies for the Social Sciences
Meertens	KNAW Meertens Institute
NCSR	Netherlands Institute for the Study of Crime and Law Enforcement

Acronym	Long
NDE	Network Digital Heritage
NICAS	Netherlands Institute for Conservation+Art+Science+
NIDI	Netherlands Interdisciplinary Demographic Institute
NIOD	National Institute for War, Holocaust and Genocide Studie
NISV	Netherlands Institute for Sound and Vision
Nivel	Netherlands Institute for Health Services Research
NKO	Dutch National Election Study
NLesC	Netherlands eScience Centre
NLP	Natural Language Processing
NPOS	National Platform for Open Science
NWA	National Science Agenda
ODISSEI	Open Data Infrastructure for the Social Science and Economic Innovations
OKiN	Parents and children in the Netherlands
OSSC	ODISSEI Secure Supercomputer
PB	Project Board
PBL	Netherlands Environmental Assessment Agency
PCT	Project Coordination Team
RDA	Research Data Alliance
RIVM	National Institute for Public Health and the Environment
RU	Radboud University (Nijmegen)
RUG	University of Groningen
RvIG	National Office for Identity Data
SALAD	Secure, Advanced, Large-Scale Analysis of Data
SANE	Secure Analysis Environments
SB	Steering Board
SCOOP	Sustainable Cooperation: Roadmaps to a Resilient Society
SCP	Netherlands Institute for Social Research
SHARE	Survey of Health, Ageing and Retirement in Europe
SoDa	Social Data Analytics Team
SRAM	SURF Research Access Management
SRL	Stakeholder Readiness Levels
SSH	Social Sciences and Humanities
SSHOC	Social Science and Humanities Open Cloud
SSHOC-NL	Social Science and Humanities Open Cloud for the Netherlands
sVREs	Secure Virtual Research Environments
TDCC	Thematic Digital Competence Centre
TRL	Technology Readiness Levels
TWG	Technical Working Group
UL	Leiden University
UM	University of Maastricht
UMCU	University Medical Center Utrecht
UNL	Universities of the Netherlands
UU	Utrecht University
UvA	University of Amsterdam
VU	Free University Amsterdam
YOUth	Youth of Utrecht

6.2 Literature references

Adger, W. Neil, Jon Barnett, Katrina Brown, Nadine Marshall, and Karen O'Brien. 2013. 'Cultural Dimensions of Climate Change Impacts and Adaptation'. *Nature Climate Change* 3 (2): 112–17. doi.org/10.1038/nclimate1666.

Bennett, W Lance, and Steven Livingston. 2018. 'The Disinformation Order: Disruptive Communication and the Decline of Democratic Institutions'. *European Journal of Communication* 33 (2): 122–39. doi.org/10.1177/0267323118760317.

Bodenhamer, David J., John Corrigan, and Trevor M. Harris. 2015. *Deep Maps and Spatial Narratives*. Indiana University Press. doi.org/10.2307/j.ctt1zxxzr2.

Boeschoten, Laura, Adriënne Mendrik, Emiel van der Veen, Jeroen Vloothuis, Haili Hu, Roos Voorvaart, and Daniel L. Oberski. 2022. 'Privacy-Preserving Local Analysis of Digital Trace Data: A Proof-of-Concept'. *Patterns* 0 (0). doi.org/10.1016/j.patter.2022.100444.

Bressers, Johannes T. A., and Kristiaan R. D. Lulofs. 2010. *Governance and Complexity in Water Management; Creating Cooperation through Boundary Spanning*. Edward Elgar / IWA Publishing. www.e-elgar.com/shop/gbp/governance-and-complexity-in-water-management-9781848449558.html.

Brown, Étienne. 2018. 'Propaganda, Misinformation, and the Epistemic Value of Democracy'. *Critical Review* 30 (3–4): 194–218. doi.org/10.1080/08913811.2018.1575007.

Bucher, Bénédicte, Carola Hein, Dorit Raines, and Valérie Gouet Brunet. 2021. 'Towards Culture-Aware Smart and Sustainable Cities: Integrating Historical Sources in Spatial Information Infrastructures'. *ISPRS International Journal of Geo-Information* 10 (9): 588. doi.org/10.3390/ijgi10090588.

Graham, Shawn, Ian Milligan, and Scott Weingart. 2015. *Exploring Big Historical Data: The Historian's Macroscope*. Imperial College Press. doi.org/10.1142/p981.

Hein, Carola, ed. 2020. *Adaptive Strategies for Water Heritage: Past, Present and Future*. Springer Nature. doi.org/10.1007/978-3-030-00268-8.

Jennings, Julia A., and Clark L. Gray. 2017. 'Climate and Marriage in the Netherlands, 1871–1937'. *Population and Environment* 38 (3): 242–60. doi.org/10.1007/s11111-016-0266-7.

Kaptijn, Ralf, Fleur Thomese, Aart C. Liefbroer, Frans Van Poppel, David Van Bodegom, and Rudi G. J. Westendorp. 2015. 'The Trade-Off between Female Fertility and Longevity during the Epidemiological Transition in the Netherlands'. *PLOS ONE* 10 (12): e0144353. doi.org/10.1371/journal.pone.0144353.

Laan, Jan van der, Marjolijn Das, Saskia te Riele, Edwin de Jonge, and Tom Emery. 2021. 'Using a Network of the Whole Population of the Netherlands to Measure Exposure to Differing Educational Backgrounds'. Working paper. doi.org/10.31235/osf.io/7jtb2.

Melgar Estrada, Liliana, Eva Hielscher, Marijn Koolen, Christian Gosvig Olesen, Julia Noordegraaf, and Jaap Blom. 2017. 'Film Analysis as Annotation: Exploring Current Tools'. *The Moving Image: The Journal of the Association of Moving Image Archivists* 17 (2): 40–70. doi.org/10.5749/movingimage.17.2.0040.

Noordegraaf, Julia, Marieke van Erp, Richard L. Zijdeman, Mark Raat, Thunnis van Oort, Ivo Zandhuis, Thomas Vermaut, et al. 2021. 'Semantic Deep Mapping in the Amsterdam Time Machine: Viewing Late 19th- and Early 20th-Century Theatre and Cinema Culture Through the Lens of Language Use and Socio-Economic Status'. In *Research and Education in Urban History in the Age of Digital Libraries*, edited by F. Niebling, S. Münster, and H. Messemer, 1501:191–212. Communications in Computer and Information Science. Springer.

Nunner, Hendrik, Vincent Buskens, Rense Corten, and Mirjam Kretzschmar. 2021. 'Effects of Risk Perception on Epidemics in a Small-World Network Game'. Working paper. Utrecht University.

Ordelman, Roeland, Carlos Martínez Ortíz, Liliana Melgar Estrada, Marijn Koolen, Jaap Blom, Willem Melder, Jasmijn van Gorp, et al. 2018. 'Challenges in Enabling Mixed Media Scholarly Research with Multi-Media Data in a Sustainable Infrastructure'. In *Digital Humanities 2018 Conference*. research.utwente.nl/en/publications/challenges-in-enabling-mixed-media-scholarly-research-with-multi-.

Paiva, Diogo, Francisco Anguita, and Kees Mandemakers. 2020. 'Linking the Historical Sample of the Netherlands with the USA Censuses, 1850–1940'. *Historical Life Course Studies* 9 (September): 1–23. doi.org/10.51964/hlcs9312.

Petrović, Ana. 2020. 'Multiscale Spatial Contexts and Neighbourhood Effects'. Delft University of Technology, Architecture and the Built Environment: 1–192. journals.open.tudelft.nl/abe/article/view/5194.

Petrović, Ana, Maarten van Ham, and David Manley. 2022. 'Where Do Neighborhood Effects End? Moving to Multiscale Spatial Contextual Effects'. *Annals of the American Association of Geographers* 112 (2): 581–601. doi.org/10.1080/24694452.2021.1923455.

Roozenbeek, Jon, Claudia R. Schneider, Sarah Dryhurst, John Kerr, Alexandra L. J. Freeman, Gabriel Recchia, Anne Marthe van der Bles, and Sander van der Linden. 2020. 'Susceptibility to Misinformation about COVID-19 around the World'. *Royal Society Open Science* 7 (10): 201199. doi.org/10.1098/rsos.201199.

Schellekens, Jona, and Frans van Poppel. 2012. 'Marital Fertility Decline in the Netherlands: Child Mortality, Real Wages, and Unemployment, 1860–1939'. *Demography* 49 (3): 965–88. doi.org/10.1007/s13524-012-0112-1.

Schulz, Wiebke, and Ineke Maas. 2012. 'Studying Career Success—the Role of Resources and Norms for Occupational Status Attainment in The Netherlands, 1865–1940'. *European Sociological Review* 28 (2): 220–40. doi.org/10.1093/esr/jcq060.

Van Noord, Nanne, Christian Olesen, Roeland Ordelman, and Julia Noordegraaf. 2021. 'Automatic Annotations and Enrichments for Audiovisual Archives': In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence:* 633–40. Vienna, Austria: SCITEPRESS – Science and Technology Publications. doi.org/10.5220/0010387706330640.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. *Scientific Data* 3 (1): 160018. doi.org/10.1038/sdata.2016.18.

Zeeuw, Eveline L. de, Kees-Jan Kan, Catharina E. M. van Beijsterveldt, Hamdi Mbarek, Jouke-Jan Hottenga, Gareth E. Davies, Michael C. Neale, Conor V. Dolan, and Dorret I. Boomsma. 2019. 'The Moderating Role of SES on Genetic Differences in Educational Achievement in the Netherlands'. *Npj Science of Learning* 4 (13): 1–8. doi.org/10.1038/s41539-019-0052-2.

Zeeuw, Eveline L. de, Lykle Voort, Ruurd Schoonhoven, Michel G. Nivard, Thomas Emery, Jouke-Jan Hottenga, Gonneke A. H. M. Willemsen, et al. 2021. 'Safe Linkage of Cohort and Population-Based Register Data in a Genomewide Association Study on Health Care Expenditure'. *Twin Research and Human Genetics* 24 (2): 103–9. doi.org/10.1017/thg.2021.18.

6.3 Relevant key publications

Dykstra, Pearl A. 2018. 'Cross-National Differences in Intergenerational Family Relations: The Influence of Public Policy Arrangements'. *Innovation in Aging* 2 (1): igx032. doi.org/10.1093/geroni/igx032.

Dykstra, Pearl A., Christoph Bühler, Tineke Fokkema, Gregor Petrič, Rok Platinovšek, Tina Kogovšek, and Valentina Hlebec. 2016. 'Social Network Indices in the Generations and Gender Survey: An Appraisal'. *Demographic Research* 34 (35): 995–1036. doi.org/10.4054/DemRes.2016.34.35.

Karpinska, Kasia, T. Fokkema, Nina Conkova, and **Pearl A. Dykstra**. 2016. 'Codebook of the Families of Poles in the Netherlands (FPN) Survey, Wave 1, Version 1'. Erasmus University Rotterdam.

Dykstra, Pearl A., and Tineke Fokkema. 2011. 'Relationships between Parents and Their Adult Children: A West European Typology of Late-Life Families'. *Ageing & Society* 31 (4): 545–69. doi.org/10.1017/S0144686X10001108.

Mandemakers, Jornt J., and **Pearl A. Dykstra.** 2008. 'Discrepancies in Parent's and Adult Child's Reports of Support and Contact'. *Journal of Marriage and Famil*y 70 (2): 495–506. doi.org/10.1111/j.1741-3737.2008.00496.x.

Waterschoot, Cedric, **Antal van den Bosch**, and Ernst van den Hemel. 2021. 'Calculating Argument Diversity in Online Threads'. In *3rd Conference on Language, Data and Knowledge (LDK 2021)*, edited by Dagmar Gromann, Gilles Sérasset, Thierry Declerck, John P. McCrae, Jorge Gracia, Julia Bosque-Gil, Fernando Bobillo, and Barbara Heinisch, 93:39:1-39:9. Open Access Series in Informatics (OASIcs). Dagstuhl, Germany: OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany. doi.org/10.4230/OASIcs.LDK.2021.39.

Karsdorp, Folgert, and **Antal van den Bosch**. 2016. 'The Structure and Evolution of Story Networks'. *Royal Society Open Science* 3 (6): 160071. doi.org/10.1098/rsos.160071.

Willems, Roel M., Stefan L. Frank, Annabel D. Nijhof, Peter Hagoort, and **Antal van den Bosch**. 2016. 'Prediction During Natural Language Comprehension'. *Cerebral Cortex* 26 (6): 2506–16. doi.org/10.1093/cercor/bhv075.

Bosch, Antal van den, and Walter Daelemans. 2013. 'Implicit Schemata and Categories in Memory-Based Language Processing'. *Language and Speech* 56 (Pt 3): 309–28. doi.org/10.1177/0023830913484902.

Bucur, Cristina-Iulia, Tobias Kuhn, Davide Ceolin, and **Jacco van Ossenbruggen**. 2021. 'Expressing High-Level Scientific Claims with Formal Semantics'. In *Proceedings of the 11th on Knowledge Capture Conference*, 233–40. K-CAP '21. New York, NY, USA: Association for Computing Machinery. doi.org/10.1145/3460210.3493561.

Bogaard, Tessel, Laura Hollink, Jan Wielemaker, **Jacco van Ossenbruggen**, and Lynda Hardman. 2018. 'Metadata Categorization for Identifying Search Patterns in a Digital Library'. *Journal of Documentation* 75 (2): 270–86. doi.org/10.1108/JD-06-2018-0087.

Boer, Victor de, Jan Wielemaker, Judith van Gent, Michiel Hildebrand, Antoine Isaac, **Jacco Van Ossenbruggen**, and Guus Schreiber. 2012. 'Supporting Linked Data Production for Cultural Heritage Institutes: The Amsterdam Museum Case Study'. In Extended Semantic Web Conference, 733–47. Springer. doi-org.eur.idm.oclc.org/10.1007/978-3-642-30284-8_5.

Geurts, Joost, Stefano Bocconi, **Jacco van Ossenbruggen**, and Lynda Hardman. 2003. 'Towards Ontology-Driven Discourse: From Semantic Graphs to Multimedia Presentations'. In *The Semantic Web – ISWC 2003*, edited by Dieter Fensel, Katia Sycara, and John Mylopoulos, 597–612. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. doi.org/10.1007/978-3-540-39718-2_38.

Van Noord, Nanne, Christian Olesen, Roeland Ordelman, and **Julia Noordegraaf**. 2021. 'Automatic Annotations and Enrichments for Audiovisual Archives': In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence*: 633–40. Vienna, Austria: SCITEPRESS – Science and Technology Publications. doi.org/10.5220/0010387706330640.

Ceolin, Davide, **Julia Noordegraaf**, Lora Aroyo, and Chantal van Son. 2016. 'Towards Web Documents Quality Assessment for Digital Humanities Scholars'. In *Proceedings of the 8th ACM Conference on Web Science*, 315–17. WebSci '16. New York, NY, USA: Association for Computing Machinery. doi.org/10.1145/2908131.2908198.

Olesen, Christian Gosvig, Eef Masson, Jasmijn Van Gorp, Giovanna Fossati, and **Julia Noordegraaf**. 2016. 'Data-Driven Research for Film History: Exploring the Jean Desmet Collection'. *The Moving Image: The Journal of the Association of Moving Image Archivists* 16 (1): 82–105. doi.org/10.5749/movingimage.16.1.0082.

Noordegraaf, Julia. 2010. 'Who Knows Television? Online Access and the Gatekeepers of Knowledge'. *Critical Studies in Television: An International Journal of Television Studies* 5 (November): 1–19. doi.org/10.7227/CST.5.2.3.

Oberski, Daniel L., and Frauke Kreuter. 2020. 'Differential Privacy and Social Science: An Urgent Puzzle'. *Harvard Data Science Review* 2 (1). doi.org/10.1162/99608f92.63a22079.

Van Erp, Sara, Joris Mulder, and **Daniel L. Oberski.** 2018. 'Prior Sensitivity Analysis in Default Bayesian Structural Equation Modeling.' *Psychological Methods* 23 (2): 363–88. doi-org.eur.idm.oclc.org/10.1037/met0000162.

Oberski, Daniel L., Antje Kirchner, Stephanie Eckman, and Frauke Kreuter. 2017. 'Evaluating the Quality of Survey and Administrative Data with Generalized Multitrait-Multimethod Models'. *Journal of the American Statistical Association* 112 (520): 1477–89. doi-org.eur.idm.oclc.org/10.1080/01621459.2017.1302338.

Mayor, Jordan R., Nathan J. Sanders, Aimée T. Classen, Richard D. Bardgett, Jean-Christophe Clement, Alex Fajardo, Sandra Lavorel, Maja K. Sundqvist, Michael Bahn, and Chelsea Chisholm, Ellen Cieraad, Ze'ev Gedalof, Karl Grigulis, Gaku Kudo, **Daniel L. Oberski** and David A. Wardle. 2017. 'Elevation Alters Ecosystem Properties across Temperate Treelines Globally'. *Nature* 542 (7639): 91–95. doi-org.eur.idm.oclc.org/10.1038/nature21027.

Strömbäck, Jesper, Yariv Tsfati, Hajo Boomgaarden, Alyt Damstra, Elina Lindgren, **Rens Vliegenthart**, and Torun Lindholm. 2020. 'News Media Trust and Its Impact on Media Use: Toward a Framework for Future Research'. *Annals of the International Communication Association* 44 (2): 139–56. doi.org/10.1080/23808985.2020.1755338.

Tsfati, Yariv, Hajo G. Boomgaarden, Jesper Strömbäck, **Rens Vliegenthart,** Alyt Damstra, and Elina Lindgren. 2020. 'Causes and Consequences of Mainstream Media Dissemination of Fake News: Literature Review and Synthesis'. *Annals of the International Communication Association* 44 (2): 157–73. doi-org.eur.idm.oclc.org/10.1080/23808985.2020.1759443.

Vliegenthart, Rens, and Liesbet van Zoonen. 2011. 'Power to the Frame: Bringing Sociology Back to Frame Analysis'. *European Journal of Communication* 26 (2): 101–15. doi.org/10.1177/0267323111404838.

Boomgaarden, Hajo G., and **Rens Vliegenthart.** 2007. 'Explaining the Rise of Anti-Immigrant Parties: The Role of News Media Content'. *Electoral Studies* 26 (2): 404–17. doi.org/10.1016/j.electstud.2006.10.018.

Kerssens, Niels, and **José van Dijck**. 2021. 'The Platformization of Primary Education in The Netherlands'. *Learning, Media and Technology* 46 (3): 250–63. doi.org/10.1080/17439884.2021.1876725.

Dijck, José van, Thomas Poell, and Martijn de Waal, eds. 2016. *De platformsamenleving: Strijd om publieke waarden in een online wereld*. Amsterdam University Press. doi.org/10.5117/9789462984615.

Dijck, José van, and Thomas Poell. 2016. 'Understanding the Promises and Premises of Online Health Platforms'. *Big Data & Society* 3 (1): 1–11. doi.org/10.1177/2053951716654173.

Dijck, José van. 2013. *The Culture of Connectivity: A Critical History of Social Media*. New York: Oxford University Press. doi.org/10.1093/acprof:oso/9780199970773.001.0001.

Van Rossum, Lisanne, Joris J. van Zundert, and **Karina van Dalen-Oskam**. 2020. 'I Catching: : Computationally Operationalising Narrative Perspective for Stylometric Analysis'. In *DHBenelux 2020*. pure.knaw.nl/portal/en/publications/f2d24089-354a-452f-ac96-61712e0d6e38.

Riddell, Allen, and **Karina van Dalen-Oskam**. 2018. 'Readers and Their Roles: Evidence from Readers of Contemporary Fiction in the Netherlands'. *PLOS ONE* 13 (7): e0201157. doi.org/10.1371/journal.pone.0201157.

Dalen-Oskam, Karina van. 2016. 'Corpus-Based Approaches to Names in Literature'. In *The Oxford Handbook of Names and Naming,* edited by Carole Hough, 344-54. doi.org/10.1093/oxfordhb/9780199656431.013.40.

Herrmann, J. Berenike, **Karina van Dalen-Oskam**, and Christof Schöch. 2015. 'Revisiting Style, a Key Concept in Literary Studies'. *Journal of Literary Theory* 9 (1): 25–52. doi.org/10.1515/jlt-2015-0003.

Rutten, Gijsbert, and Rik Vosters. 2021. 'Language Standardization "from Above"'. In *The Cambridge Handbook of Language Standardization*, edited by Wendy Ayres-Bennett and John Bellamy, 65–92. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press. doi.org/10.1017/9781108559249.003.

Rutten, Gijsbert, Andreas Krogull, and Bob Schoemaker. 2020. 'Implementation and Acceptance of National Language Policy: The Case of Dutch (1750–1850)'. *Language Policy* 19 (2): 259–79. doi.org/10.1007/s10993-019-09527-y.

Rutten, Gijsbert. 2019. *Language Planning as Nation Building: Ideology, Policy and Implementation in the Netherlands, 1750–1850*. Amsterdam: John Benjamins Publishing Company. doi.org/10.1075/ahs.9.

Rutten, Gijsbert, and Marijke van der Wal. 2014. *Letters as Loot. A sociolinguistic approach to seventeenth-and eighteenth-century Dutch.* Advances in Historical Sociolinguistics 2. Amsterdam: John Benjamins Publishing Company. benjamins.com/catalog/ahs.2.

Hessels, Roy S., Gijs A. Holleman, Alan Kingstone, Ignace T. C. Hooge, and **Chantal Kemner**. 2019. 'Gaze Allocation in Face-to-Face Communication Is Affected Primarily by Task Structure and Social Context, Not Stimulus-Driven Factors'. *Cognition* 184 (March): 28–43. doi.org/10.1016/j.cognition.2018.12.005.

Boomen, Carlijn van den, Victor A.F. Lamme, and **Chantal Kemner**. 2014. 'Parallel Development of ERP and Behavioural Measurements of Visual Segmentation'. *Developmental Science* 17 (1): 1–10. doi.org/10.1111/desc.12093.

Vlamings, Petra Hendrika Johanna Maria, Lisa Marthe Jonkman, Emma van Daalen, Rutger Jan van der Gaag, and **Chantal Kemner**. 2010. 'Basic Abnormalities in Visual Processing Affect Face Processing at an Early Age in Autism Spectrum Disorder'. *Biological Psychiatry* 68 (12): 1107–13. doi.org/10.1016/j.biopsych.2010.06.024.

Vandenbroucke, Myriam W. G., H. Steven Scholte, Herman van Engeland, Victor A. F. Lamme, and **Chantal Kemner.** 2008. 'A Neural Substrate for Atypical Low-Level Visual Processing in Autism Spectrum Disorder'. *Brain: A Journal of Neurology* 131 (Pt 4): 1013–24. doi.org/10.1093/brain/awm321.

Mulder, Clara, Isabel Palomares-Linares, and Sergi Vidal. 2022. 'Internal Migration, Living Close to Family, and Individual Labour Market Outcomes in Spain'. *Comparative Population Studies* 47 (January): 3–28. doi.org/10.12765/CPoS-2022-01.

Wiel, Roselinde van der, Niels Kooiman, and **Clara Mulder**. 2021. 'Family Complexity and Parents' Migration: The Role of Repartnering and Distance to Non-Resident Children'. *European Journal of Population* 37: 877–907. doi.org/10.1007/s10680-021-09594-0.

Mulder, Clara, Emma Lundholm, and Gunnar Malmberg. 2020. 'Young Adults' Return Migration from Large Cities in Sweden: The Role of Siblings and Parents'. *Population, Space and Place* 26 (7): e2354. doi.org/10.1002/psp.2354.

Thomas, Michael J., and **Clara Mulder**. 2016. 'Partnership Patterns and Homeownership: A Cross-Country Comparison of Germany, the Netherlands and the United Kingdom'. *Housing Studies* 31 (8): 935–63. doi.org/10.1080/02673037.2016.1164832.

6.4 Number of pages

The number of pages of Sections 1 to 6 of this grant application is: 50 pages *(max. 50 pages)*

7. Data management

This is a standard section that is part of all NWO grant applications and is not included in the page limit. In filling out this section, you may refer to other specific parts of the application for more detailed information compared to the information you provide here.

1. Will this project involve re-using existing research data?

   - [x] Yes: Are there any constraints on its re-use?
   - [ ] No: Have you considered re-using existing data but discarded the possibility? Why?

   If no, please briefly explain why; if yes, state any constraints on re-use of existing data if there are any.

   The re-use of existing research data within the project is only constrained in certain instances that the use be restricted and limited to research purposes only.

2. Will data be collected or generated that are suitable for reuse?

   - [x] Yes: Please answer questions 3 and 4.
   - [ ] No: Please explain why the research will not result in reusable data or in data that cannot be stored or data that for other reasons are not relevant for reuse.

3. After the project has been completed, how will the data be stored for the long-term and made available for the use by third parties? Are there possible restrictions to data sharing or embargo reasons? Please state these here.

   All data collected or generated within the project must be archived within a trusted repository and documented according to the FAIR principles. Data that is collected or generated within the project must be made available for research purposes to third parties. This will follow the principle of 'Open as possible, secure as necessary' such that where possible data will be made freely available, and where necessary it will be accessible via suitable sVRE's.

4. Will any costs (financial and time) related to data management and sharing/preservation be incurred?

   - [x] Yes: Then please be sure to specify the associated expenses in the budget table of this application.
   - [ ] No: All the necessary resources (financial and time) to store and prepare data for sharing/preservation are or will be available at no extra cost.

   The budgets for the Compute, Data and FAIR work streams are in large part dedicated to aspects of data sharing and preservation.

8. Declarations and signature

Funding elsewhere

Have you requested funding via a grant application for this research infrastructure elsewhere?

X No

☐ Yes

Ethical aspects

	Not applicable	Not yet applied for	Applied for	Received
Approval from a medical ethics review committee	**X**			
Approval from an animal experiments committee	**X**			
Permission for research with the population screening Act	**X**			
Approval from any other recognised ethics review committee(s)		**X**		

By submitting this form, I declare that:

X I have completed this form truthfully;

X I satisfy the nationally and internationally accepted standards for scientific conduct as stated in the Netherlands Code of Conduct for Research Integrity 2018;

X I have submitted non-referees (no more than 3 in total);

X I have entered all co-applicants in ISAAC;

☐ I endorse and follow the Code Openness Animal Experiments (if applicable);

☐ I endorse and follow the Code Biosecurity (if applicable);

X The consortium partners are aware of the NWO Grant Rules and obligatory establishment of an agreement containing IP&P arrangements and will adhere to this if the proposal is awarded.

Initial(s) and surname(s): P.A. Dykstra

Place: Aerdenhout

Date: 25 March 2022