1   General information

1.1   Application title

ODISSEI: Better Infrastructure, Better Science, Better Society

1.2   Scientific summary

Society is facing complex challenges, such as ageing, climate change, migration, populism, labour insecurity, and increasing social inequalities. Understanding these challenges and finding suitable solutions is at the core of social science. In today's rapidly changing world, the role of social scientists is more important than ever before. With their interest in explaining how society works, social scientists play a vital role in informing local, national and international authorities, non-governmental organisations, practitioners, and the public. The complex societal changes, trickling down to people's daily lives, cannot be described and explained from single types of data or one single discipline.

Currently, social scientists are too often constrained in the questions they can answer, not by the limitations of their thinking but by the fragmented data landscape that clusters disciplinary data in distinct silos, the lack of sustained and coordinated investment in longitudinal studies, and the under use of research opportunities offered by the digital age. For societies to gain resilience in the face of the enormous challenges they face, social scientists need linked data that bridge disciplinary fields, levels of analysis and historical time, and the tools, resources and expertise necessary to analyse them. To create a better society, there is a need for better social science, and this requires a better data infrastructure:

Better Infrastructure - Better Science - Better Society

In the last two years, the entire social science community in the Netherlands has united in an unprecedented level of financial, technical, and political support for a shared research infrastructure: ODISSEI, Open Data Infrastructure for Social Science and Economic Innovations. This momentum and initiative must now be supported, exploited, and extended, so that ODISSEI can facilitate a new generation of social science and avoid missing the research potential unleashed by the digital age. The Dutch social science community is committed to carrying ODISSEI into the future. This proposal seeks to transition ODISSEI from a fledgling collaboration into a truly national infrastructure that facilitates ground-breaking social science research at scale. To better address today's complex societal challenges and to find suitable solutions, social scientists in the Netherlands must pursue four key goals:

1. Accessing, linking, and analysing sensitive data—that bridge different disciplines, analytical levels and time periods—in a safe, secure, and ethical manner (Work stream: **Data Facility**)
2. Sustaining and optimising valuable, long-standing data collections (Work stream: **Observatory**)
3. Opening up new avenues of inquiry by exploiting innovative digital technologies (Work stream: **Laboratory**)
4. Acquiring skills and approaches for more complex and comprehensive modelling of social phenomena (Work stream: **Hub**)

The **Data Facility** will be a cluster of systems to strengthen the findability, accessibility, interoperability, and reusability (**FAIR**) of data from longitudinal panel studies in the social sciences, administrative data from Statistics Netherlands, and new forms of data such as large-scale digitally available, textual data, and social media data. It also includes a secure high-performance computing environment, incorporating linkable datasets and advanced analytics tools at the national supercomputer of SURFsara. Acknowledging that all data are potentially identifiable and potentially sensitive, ODISSEI will have policies and technical procedures in place to protect data privacy. High-speed analyses of timely, complex and diverse data will help better address social science's key intellectual and policy puzzles.

The **Observatory** will optimise and provide continuity in data collection, enabling the study of long-term trends and supporting the integration of these long-standing data collections with the wider ODISSEI infrastructure. Continuity is a prerequisite for the study of crucial social questions such as how people's lives are affected by the times in which they live and how their lives develop over time.

The **Laboratory** will work on innovating data collection and survey infrastructures, novel data sources and analysis techniques, and developing new methodological standards. It provides researchers with controlled environments in which to run social experiments and develop and test new measurement methods. It also includes tools for engaging in citizen science, a platform for running web experiments, and a suite for the analysis

of large-scale digitally available, textual data. The expansion of social scientists' research repertoire will provide enormous opportunities for the exploration of contradictions and complementarities of 'traditional' and 'novel' data sets in the pursuit of understanding the pressing social questions of our time.

The **Hub** will coordinate ODISSEI's educational and training programme so that students and researchers can familiarise themselves with advanced analytical techniques and tools for computational social science. The acquisition of computational skills will enable them to make full use of the new resources for social science research, available through the infrastructure.

By capitalising on the research opportunities that the digital age offers, ODISSEI indicates a major shift in the way social science research is conducted, enabling Dutch researchers to continue to be world leading in social science. ODISSEI opens up new avenues of enquiry in the social sciences and beyond that will attract the best talents from abroad.

1.3   Lay summary

ODISSEI is a shared research infrastructure that will bring together existing data from the social sciences and enrich them with data from government registries and online sources. The infrastructure ensures that Dutch social scientists are well equipped to exploit the opportunities offered by the digital age. By developing innovative analytics and presentation tools, and offering world class, secure and ethical data management facilities, ODISSEI will enhance the position of Dutch social scientists as global leaders and enable them to better address the pressing social questions of our time.

1.4   Key words

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

1.5   Scientific domain

Humanities and Social Sciences

1.6   Main field of research

Code	Field of research
21.10.00	Bioinformatics/biostatistics, biomathematics, biomechanics
23.44.00	Health economics
23.46.00	Medical sociology
23.80.00	Epidemiology
28.90.00	Gender studies
38.10.00	Microeconomics
38.30.00	Econometrics
39.90.00	Business administration
40.10.00	Clinical Psychology
40.20.00	Biological and Medical Psychology
40.30.00	Developmental Psychology
40.40.00	Psychonomics and Cognitive Psychology
40.50.00	Social and Organisational Psychology
40.60.00	Psychometrics
41.90.00	Educational Sciences
42.00.00	Pedagogics
43.40.00	Criminal law and Criminology
44.10.00	Public administration
44.20.00	Political science
45.90.00	Sociology
47.90.00	Communication science
48.90.00	Demography
49.11.00	Geography: Planning

2   Requested Infrastructure

2.1   Science and excellence case

2.1.1   Scientific quality and urgency of ODISSEI

2.1.1.1   Complex societal challenges

Society is facing complex challenges, such as ageing, climate change, migration, populism, labour insecurity, and increasing social inequalities. Understanding these challenges and finding suitable solutions is at the core of social science, which covers a wide spectrum of disciplines including economics, political science, human geography, sociology, psychology, communication science, anthropology, demography, criminology, and law. In today's rapidly changing world, the role of social scientists is more important than ever before. With their interest in explaining how society works, social scientists play a vital role in informing local, national and international authorities, non-governmental organisations, and the public.

Processes like modernisation, globalisation, and individualisation are making people's lives less predictable, more complex and interconnected (Castells, 2011). In these 'fluid times', social processes become individuals' challenges, and some people face more challenges than others (Bauman, 2013). For instance, the labour market is changing dramatically. Low skilled workers must adapt to more and longer periods of income and job insecurity via temporary or atypical contracts whilst skilled workers face the challenges of restructuring, automation, and international outsourcing (Arntz et al., 2016). The complex societal changes, trickling down to people's daily lives, cannot be described and explained from single types of data or by one single discipline. Inequalities echo and ripple through various spheres of life, paying no heed to the largely compartmentalised disciplines of the social sciences. Inequalities in work are reflected at home (Van der Lippe et al., 2010), disparities in education are echoed in health (Conti, Heckman & Urzua, 2010), and inherited genes resonate in day-to-day behaviours (Tropf et al., 2015). Neither can the influences of societal conditions on people's lives be described and explained with data from a single point in time. Changes such as the economic crisis and welfare state restructuring create varying opportunities and restrictions in the lives of individuals and their families (Hagestad & Dykstra, 2016). For societies to gain resilience in the face of the enormous challenges they face, social scientists **need linked data that bridge** disciplinary fields, levels of analysis and historical time, **and the tools**, resources and expertise necessary to analyse them.

2.1.1.2   Digital age

The transition from the analogue age to the digital age is **creating new opportunities** for social scientists (Mann, 2016). Digitisation has facilitated the use of existing data, such as censuses and population registers (Thorvaldsen & Østrem, 2018), and bibliometric databases (Robinson-Garcia et al., 2019). Novel digital sources can serve as the basis for 'nowcasting', the provision of real-time information on social trends. For example, online platforms used for posting job vacancies are a promising source for real-time data on employment (Reimsbach-Kounatze, 2015). Digital technologies have led to a re-energisation of research in self-administered instruments, focusing on styles of formatting and visual presentation (Groves, 2011), including gamification as a means to reduce survey fatigue (Turner et al., 2014). Smartphones prove to be particularly advantageous for research among low-resource, highly mobile, and hard-to-reach groups (Sugie, 2018). Online labour markets and social-network sites are a fast and easy way to recruit participants for surveys, but their samples tend not to be demographically representative (Boas et al., 2018). Running experiments by using digital devices in the physical world (e.g. Allcott & Rogers, 2014) or by doing them in a fully digital environment (e.g. Bail et al., 2018) is another new research opportunity, but randomised online experiments are not without controversy. In a project that has come to be known as 'Emotional Contagion', Facebook users felt the network was manipulating them without their consent (Verma, 2014). The public outcry underscores the need for social scientists to pay close attention to ethical research design. Finally, the digital age enables researchers to create mass collaborations to tackle scientific problems that cannot be solved individually. For example, Benoit et al. (2016) used an online microtask labour market to 'crowd code' over 4,000 political manifestos, a task that was impossible to carry out with their earlier expert-run coding system. ODISSEI will capitalise on the research opportunities that the digital age offers to continue to be world leading in social science.

More crucially, the digital age **challenges** social scientists to question their **style of research** (Salganik, 2018). Massive amounts of digital data on social behaviours are gathered through social media platforms like Facebook and Twitter, search engines, smartphones, and sensors interconnected through the Internet of Things for purposes other than research (Reimsbach-Kounatze, 2015). These 'ready-made' data are a big change from the carefully designed, sampled and collected data with which social scientists have typically worked. There is a **need**

for critical interrogation of these new data sources, to understand their possibilities and limits (Halford & Savage, 2017). While ready-made digital data provide exciting opportunities for new research, they are unlikely to make 'traditional' social science data obsolete. Traditional population data, for example, are crucial for identifying systematic bias in online sources and calibrating estimates made from these data (Zagheni et al., 2017). ODISSEI will combine the empirical power of multiple and diverse datasets, explore their contradictions and complementarities, in the pursuit of understanding major social questions.

The new digital landscape raises some well-recognised **ethical concerns** (Zook et al., 2017). The most common of these concerns focuses on the privacy of research subjects and informed consent. The risk of disclosure needs to be minimised. Acknowledging that all data are potentially identifiable and that all data are potentially sensitive (Salganik, 2018), ODISSEI will have policies and technical procedures in place to protect data privacy (for details, see section 2.3.3.5 and 2.3.3.6).

Many social scientists will need to acquire new and largely unfamiliar **skills and approaches** in order to work with new sources of large volumes of digital data. They are increasingly within reach thanks to the evolution of computational social sciences. Now is an opportune time for collaboration across the social and computational sciences towards a more mutually productive future (Wallach, 2016), which is precisely what ODISSEI envisions.

2.1.1.3 Description of the requested research infrastructure

To better address today's complex societal challenges and to find suitable solutions, social scientists in the Netherlands must pursue four key goals (Figure 1):
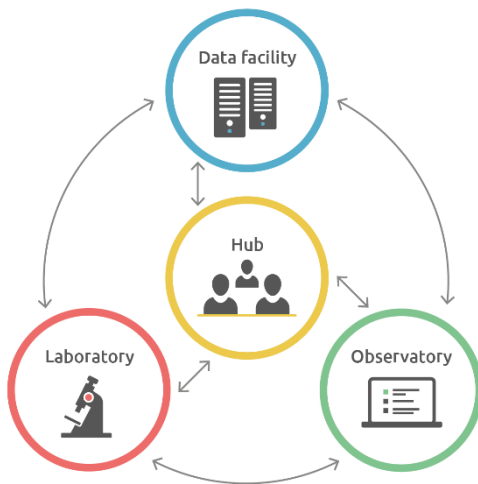
Figure 1 – Work streams

1. Accessing, linking, and analysing sensitive data—that bridge different disciplines, analytical levels and time periods—in a safe, secure, and ethical manner (Work stream: **Data Facility**)
2. Sustaining and optimising valuable, long-standing data collections (Work stream: **Observatory**)
3. Opening up new avenues of inquiry by exploiting innovative digital technologies (Work stream: **Laboratory**)
4. Acquiring skills and approaches for more complex and comprehensive modelling of social phenomena (Work stream: **Hub**)

ODISSEI will organise its activities in four Work streams. The **Data Facility** will help researchers to find the data they need, access administrative whole population data held by Statistics Netherlands, and make use of a secure high-performance computing facility. The **Observatory** will help maintain existing data collections within the Netherlands, put innovations into production, and link data that are collected with the administrative data accessed via the Data Facility. The **Laboratory** will allow researchers to develop and test new data collection methods and novel instruments. The **Hub** will coordinate training programmes for researchers and help them navigate the infrastructure. Further details are given below, and a technical description of the Work streams is provided in section 2.4.

Current ODISSEI Activities		By the end of the Project
ODISSEI Data Facility		
Secure supercomputer pilot for three projects	⇒	Administrative data access enhanced and extended Full production for secure supercomputer ODISSEI Portal developed
ODISSEI Observatory		
Limited support for data collections	⇒	Full participation in all Dutch Social Science European Research Infrastructure Consortiums (ERICs) Sustainability experiments Creation of a production pathway for survey innovations

ODISSEI Laboratory		
Longitudinal Internet studies for the Social Sciences (LISS) study	⇒	Innovations in data linkage and access piloted Experimental platforms piloted
ODISSEI Hub		
Basic coordination & communications operations	⇒	Educational programme Community managers & events Analytical support services

Table 1 – Summary of the Work stream activities

2.1.1.4   ODISSEI and scientific and technological innovation

The Data Facility

The ODISSEI Data Facility is a cluster of systems to strengthen the findability, accessibility, interoperability, and reusability (**FAIR**) (Wilkinson et al., 2016) of research data in the Netherlands. The first part of the data facility is the secure remote access **microdata services facility** offered by Statistics Netherlands, which enables analyses of not only administrative data, but also of administrative data linked with existing social scientific datasets (under the condition that respondents have provided consent). Apart from being more cost-effective to use existing administrative data to answer a research question than to undertake a de novo data collection (Currie, 2013), having factual data (e.g., about income, health care use) rather than self-reports greatly expands the options to generate accurate, timely, and effective knowledge to address the most pressing societal questions.

The second and most revolutionary part of the data facility, is a closed high-performance computing environment for social scientists carrying out complex modelling tasks at high speed: the **ODISSEI Secure Supercomputer**. It provides a secure remote-access environment incorporating linkable datasets and advanced analytics tools at the national supercomputer of SURFsara. With Roadmap funding, this facility will be scaled up from its recent successful pilot phase (see section 2.1.1.7) to service all ODISSEI member organisations. The ODISSEI Secure Supercomputer will offer an unmatched environment to enable ground-breaking social science and interdisciplinary studies by Dutch researchers and their collaborators.

To ensure researchers can find the data they need, the Data Facility includes the ODISSEI Portal, which will act as a necessary bridge between a wide range of data that social scientists use for research. The **ODISSEI Portal** will make administrative data, social science data collections, and new forms of data searchable through a single interface. The Netherlands has high-quality registers, and persistent identifiers are widely and consistently used across government institutions. Drawing on the scientific information NARCIS portal[1], ODISSEI will develop a FAIR metadata catalogue of the ODISSEI data collections (e.g. Longitudinal Internet studies for the Social Sciences (LISS), European Social Survey (ESS), Survey of Health, Ageing and Retirement in Europe (SHARE), European Values Study (EVS), Historical Sample of the Netherlands (HSN), and Netherlands Twin Register (NTR)) and connect it with the Statistics Netherlands catalogue. In addition to this, the ODISSEI Portal will help find datasets by providing advanced semantic search, facilitate data access requests, and, after acceptance by the data owner, transfer the data to the designated analysis environment (e.g. the ODISSEI Secure Supercomputer, described below). The Portal will significantly reduce the amount of time that researchers spend searching for suitable data (European Commission, 2018).

The Observatory

The ODISSEI Observatory provides continuity in data collection, enabling the study of long-term trends and supporting the integration of these long-standing data collections with the wider ODISSEI infrastructure. The Netherlands has a rich history in data collection within the social sciences and ODISSEI will seek to strengthen this. If continuity is not provided, the ability to examine how people's lives are affected by the times in which they live, and the ability to examine developments in people's lives, are lost. Currently, the fractured landscape of data collections is under pressure from the rising costs of data collection and the perpetual advances and innovations in data collection (Groves, 2011). Both trends threaten the longitudinal continuity of long-standing studies. Within the term of this project, ODISSEI will establish a **Council of Data Collections** to facilitate collaboration and transition the long-standing data collections into a coherent and sustainable future in which they are well placed within a more diverse and dynamic data landscape. To do this, ODISSEI will ensure the Dutch participation in cross-national data collections and will also build key strategic links with other disciplines.

---

[1] www.narcis.nl/?Language=en

It will also establish an independent review process for the evaluation and assessment of the Dutch data landscape. The Observatory will work to identify critical gaps in the landscape and allocate resources to sustain longitudinal data collections. The Observatory will support their transition to sustainable operational models, the integration of new data collection methodologies from the lab and linkages with secure, administrative data via the Data Facility.

ODISSEI will coordinate the development of a **push to web strategy** and infrastructure for studies which are moving face-to-face surveys into an online environment without damaging their long-term data comparability. This will be done through shared best practices, investment in adaptation strategies, consultation with fieldwork agencies, and coordination of international projects. These will then be made available to other countries fielding these surveys, demonstrating a clear production pathway from the Laboratory to the Observatory and beyond, strengthening the Netherlands as a hub of social science methodology and innovation. Finally, the Observatory will seek to **integrate long-standing data collections** within the ODISSEI infrastructure that will provide greater disciplinary diversity to ODISSEI. These include the National Twin Registry (NTR) and the Historical Sample of the Netherlands (HSN), both of which will encourage interdisciplinary research amongst ODISSEI users.

The Laboratory

The Laboratory is ODISSEI's centre for innovations in data collection, novel data sources and analysis techniques, and developing new methodological standards. It provides researchers with controlled environments in which to run social experiments and develop and test new measurement methods.

The Laboratory will prepare and test **operational innovations** for the wider ODISSEI infrastructure. First, there will be work on automated data linkage techniques to enhance the functionality of the ODISSEI Portal and support linkage across the infrastructure. These techniques are rapidly developing in the field of computer science but have yet to be fully applied to social science data. Such automated linkage would make data easier to find and enhance the power of the ODISSEI Portal. Second, ODISSEI will test the application of distributed computation on social science datasets. Distributed computation allows for analyses across datasets that are never centralised, by utilizing distributed learning algorithms.

Alongside these innovations for the wider infrastructure, ODISSEI will pilot **new platforms** to support research in burgeoning areas of social science. These include tools for engaging in citizen science, a platform for running web experiments and a suite for the analysis of large-scale digitally available, textual data. These platforms will provide tools for gathering and analysing data in new and innovative ways. The Media Content Analysis Lab will provide a suite of computational tools for the collection and textual analysis of large digital media corpora. Currently, these tools are largely developed in isolation by separate labs and are not easily adapted across corpora or tasks. The Media Content Analysis Lab will develop flexible versions of these tools for common use. The results can for example be used to gain insights into longitudinal trends in public debate. ODISSEI will collaborate with CLARIAH[2], the Dutch Humanities infrastructure with regards to wider uptake of these tools. The Mass Experiments Online Laboratory will enable researchers to deploy experiments by utilizing a series of tools which enable the creation and monitoring of social networks at scale. There will also be the Citizen Science Platform which helps scientists conduct citizen science projects incorporate tools developed using the best practices of social science methodologists. This will add greater rigour to citizen science approaches whilst broadening the types of data that ODISSEI users can link to.

Another key element of the Laboratory is the **Longitudinal Internet studies for the Social Sciences** (LISS) panel, administered by an ODISSEI member organisation, CentERdata. The LISS panel, based on a true probability sample of households drawn from the population register by Statistics Netherlands, consists of 5,000 households, comprising 8,000 individuals. The LISS panel was a pioneer in running experiments on the use of wearables and smart devices in survey research. One example is the use of accelerometers in an international study on comparing physical activity and health (Kapteyn et al., 2018) which has now been implemented in ten countries participating in the Survey of Health, Ageing and Retirement in Europe (SHARE). Another example is the use of advanced bathroom scales in a study on body mass measurement and health behaviour (Kooreman and Scherpenzeel, 2014).

All these activities will produce substantive, methodological, and technical innovations that can then be taken forward, scaled up, and implemented in other components of ODISSEI such as the Data Facility or Observatory.

---

[2] www.clariah.nl/en/

The Hub

The Hub will coordinate ODISSEI's **educational and training program** in collaboration with Statistics Netherlands, the Netherlands eScience Center (NLeSC), the national supercomputing institute SURFsara, and the national research data archive DANS. This will include training for: Supercomputer (SURFsara); Administrative microdata (Statistics Netherlands); Data Management (DANS), and Analytics (NLeSC).

ODISSEI will build and maintain a **large community** through regular contact with graduate program directors at ODISSEI member organisations. This will facilitate knowledge exchange, new ideas, and collaboration between the social and computational sciences, resulting in innovation and alignment. ODISSEI will reinforce this by organising conferences, expert sessions, workshops, and summer schools. Community Managers will oversee these educational and training programs to ensure that the next generation of social scientists are well placed to make use of the computational infrastructure.

The educational programme will not be aimed at any researcher with generic needs. In an advanced infrastructure such as ODISSEI, many of the training and support needs are specific and unique and bespoke solutions are required. The Hub will therefore coordinate a **Social Analytics team** that will work closely with researchers from member organisations to implement advanced analytical techniques and tools for computational social science that make use of the diverse infrastructure available via ODISSEI.

2.1.1.5   ODISSEI and scientific breakthroughs

It is broadly recognised that the current explosion in the amount and diversity of data available will advance our understanding of human behaviour and social phenomenon in a way that has never been possible before (Connelly et al., 2016). Conventional experiments, in actual laboratories, have very strong research designs, but such experiments are not situated in real life and lack external validity. Because ODISSEI allows for the integration of administrative data with data from social surveys and other sources such as social media, it will create the **first population-based social pseudo-laboratory in the world**. It will contain an ever-growing amount and range of data on the individual, household, family, ethnic group, firm, region, or any other aggregated level, to improve scientific testing of any social mechanism. Below are just a few examples of cutting-edge research that require ODISSEI to answer them:

(1) How do complex social policies interact with each other?

Social policy systems are exceptionally complex in that adjustments in one specific policy can have unintended effects elsewhere in the social policy system and wider economy. For example, Borghans, Gielen and Luttmer (2014) used data from Statistics Netherlands to show that a reduction in disability insurance led individuals to increase claims of social assistance and increase their labour market activity. The data from Statistics Netherlands not only include information from different policy spheres but also provide insights into different areas of individuals' lives such as their labour market activity or family situation. Designing good social policies requires a detailed understanding of what their broader impact will be. Doing this requires rich and complex data but analysing the complex butterfly effects of policies also requires a secure, high-performance computing facility that links sensitive data from diverse sources and enables them to be analysed.

"ODISSEI provides exceptional opportunities to consider the impact of social policies on adults and their offspring. This data facility plays a key role in showing that the actual impact of social policies often extends far beyond the direct effects the program aimed at. *I need ODISSEI* to help me examine these impacts with greater accuracy and clarity."

Anne Gielen, Associate Professor, Erasmus University Rotterdam

(2) How do social networks affect our opinions and behaviours?

We live in networked societies and understanding how information is shared across these networks is vital for understanding the modern world. Van de Rijt (2019) showed how social network recommendations and viral popularity improved people's attitudes towards objectively inferior products but that the positivity effect was short lived. This finding suggests that whilst networks can appear to popularise poor options in the short run, in the long run good options win out. To illustrate this, Van de Rijt ran small scale online network experiments, where information was introduced and then its spread and popularity in the network was monitored. Running these experiments is complex on a technical level and requires significant set up costs. Nevertheless, the resulting scale of the networks in these experiments is generally small and therefore limited in the kinds of insights it can provide. The platforms developed in the ODISSEI Laboratory will help researchers conduct online experiments at

scale and with high quality infrastructure that can support larger networks, providing further insights into our networked lives and their increasing impact on wider society.

 "ODISSEI will enable experiments with groups containing several hundred participants. We know that for many settings, small-group behaviour is fundamentally different from large-group behaviour. The scale of experiments that would be possible with ODISSEI will allow me to experimentally study the spread of fake news through synthetic social networks. **I need ODISSEI** to rapidly advance this research agenda."
Arnout van de Rijt, Professor, Utrecht University

(3) Is it better to have an educational degree that is linked to a specific career trajectory?

A recurring question is whether occupation specific skills enhance labour market outcomes. To answer this question, Bol et al. (2019) analysed data on educational and labour market trajectories to show that higher earnings associated with having a degree that closely matches subsequent employment depends on the existing clarity and strength of the pathway between educational credentials and the labour market. The earnings premium associated with a good occupational match is larger in contexts where qualifications have a stronger link to the labour market, but the penalty for a mismatch is simultaneously greater. Given increased enrolment in higher education and a changing labour market, these questions are vital in debates about the future of higher education systems. To enhance this analysis further requires combining data on education and labour market outcomes over the course of people's lives. The unique data at Statistics Netherlands provides exceptionally rich insights that would allow researchers to extend this analysis, but currently they are constrained by the computing capacity.

"The Dutch register data contain unique and detailed information on education and employment trajectories and would be perfectly suited to develop my research agenda. However, the current infrastructure is not up to task of analysing these amounts of data, thereby hindering potential innovation. **I need ODISSEI** to provide the computational capacity to handle the complexity inherent within educational and life course trajectories and reach a new level of understanding."
Thijs Bol, Associate Professor, University of Amsterdam

See section 2.1.1.7 for **more examples** of research questions that can be answered only through ODISSEI.

2.1.1.6   Relevant results from recent evaluations

Several recent reports have highlighted the excellent quality and reputation of Dutch social science (Koens et al., 2018; Van Dijck & Saarloos, 2017). International comparisons show that the scientific impact, as measured by the number of citations, is very high, higher than many other research areas in the Netherlands. The high citation impact is all the more remarkable because the research capacity of Dutch social science is of small to average size. Research in political science, sociology, anthropology, information science and communication science are cited particularly frequently, demonstrating that it is appreciably acknowledged by peers. No country has received more European Research Council (ERC) grants in Social Science & Humanities on a per capita basis in the last ten years than the Netherlands. According to the QS World University Rankings and the Academic Ranking of World Universities (also known as the Shanghai-ranking), Dutch social science taken together is **among the best in the world**. Individual departments hold remarkable positions in the league table. For example, Business Administration at Tilburg University is fifth worldwide in the Shanghai-ranking. Public Administration at the Erasmus University leads the Shanghai-ranking, and media and communication science at the University of Amsterdam is second in the QS-ranking.

2.1.1.7   Quality of results achieved so far

ODISSEI was launched in October 2016 with the financial support of NWO, 19 Dutch faculties, four research institutes, Statistics Netherlands (CBS) and DANS, the Dutch national research archive for scientific research. ODISSEI currently has grown to 37 member organisations (for details see section 2.1.2.1). It represents a milestone in Dutch social science as for the first time such a comprehensive consortium of stakeholders has been brought together to develop an open data infrastructure for the social sciences.

ODISSEI has already made significant strides in facilitating world class research. First, ODISSEI has **safeguarded the continuity of several high-quality and strategically important data collections within the Netherlands**. ODISSEI has secured the fieldwork for wave 8 of the Survey of Health, Ageing and Retirement in

Europe (SHARE)[3] which will start in October 2019. SHARE is a world leading survey infrastructure and part of G2Aging - an initiative to harmonize population survey data on aging around the world. In the Netherlands alone, SHARE has 964 users and the project publishes more than one peer-reviewed publication every week. The Netherlands plays a prominent role within SHARE as the technical infrastructure for the project is provided by CentERdata. ODISSEI also secured the Dutch participation in the European Values Study (EVS) in 2018. This survey has been conducted every nine years since 1981 and helps monitor long term changes in attitudes and values (e.g. Bréchon & Gonthier, 2017). The international project is based at Tilburg University. At the national level, ODISSEI has also supported the fielding of the National Election Study (NKO) in 2017. The nature of Dutch politics means that election cycles are hard to anticipate and predict and raising financing for election studies is exceptionally difficult. To mitigate this, ODISSEI provides continuity funding at short notice. ODISSEI has also taken on the responsibility of Dutch membership in international social science projects such as the Luxembourg Income Study, which has 50 active Dutch users annually.

Beyond supporting these data collections, ODISSEI has extended the 'Longitudinal Internet studies for the Social Sciences' (LISS) panel and provided nine **new research projects free panel time to collect data** in it. The projects were evaluated by a panel of leading scientists. 31 projects were submitted for consideration, illustrating a high demand for space in the panel and the quality of applications was high. The researchers who lead the successful projects are provided with around fifteen minutes of time with respondents in the panel. Amongst the nine projects there is research on privacy and data concerns, political discontent, digital health services, parental well-being and mental health, greed and self-interest, and collective nostalgia and group dynamics[4]. The researchers can target specific respondents or split their time over multiple waves of LISS to provide longitudinal data. The panel allows researchers access to a high-quality, representative sample on which to test their ideas within a relatively short period of time. PhD candidates and postdocs can gain access to the panel and test their own theories without having to construct their own data collection infrastructure or acquire large grants.

Due to the success of the LISS access grants, a similar design was applied to microdata access at Statistics Netherlands. These data provide an accurate insight into the lives of the entire population in the Netherlands and are therefore extremely valuable to identify social cleavages at the micro-level. Often, however, there are substantial barriers to access the data. For that reason, ODISSEI has launched the two-pronged Microdata Access Programme. First, the Microdata Access Grant provides five selected projects with **free access to Statistics Netherlands microdata**. The call was launched at a free workshop at Statistics Netherlands which introduced researchers to the data that was available. At the time of writing, the 24 proposals submitted are being reviewed by a panel of experts. Second, in the Microdata Access Discount programme, ODISSEI has also supported a further 111 projects with access to the microdata at reduced costs.

Both the LISS call and the Microdata Access Grant **have been successful in enabling early career researchers to pursue their pressing scientific questions and test them using the highest quality data available**. Looking forward, ODISSEI will seek to integrate these calls to make it possible for researchers to access the LISS Panel, Statistics Netherlands data and the ODISSEI Secure Supercomputer simultaneously within a single project. This opens up the opportunity for highly innovative and impactful studies to use an integrated data infrastructure to its full capacity. The LISS calls and Microdata Access Grant allow ODISSEI to identify demand and the needs of the research community.

The ODISSEI community includes many advanced and heavy users of the data held by Statistics Netherlands, but they are currently limited by the available computing capacity at Statistics Netherlands. Identifying this demand has led to the biggest advance within ODISSEI over the last two years which has been the creation of the ODISSEI Secure Supercomputer. A fully functioning pilot of the ODISSEI Secure Supercomputer has been delivered by SURFsara in March 2019, on which three high-capacity projects have simultaneously run (described below). The **ODISSEI Secure Supercomputer can be used to analyse highly sensitive data in a secure environment**, including administrative microdata on persons, households, companies etc. from Statistics Netherlands. This is due to the system's architecture: SURFsara acts as Trusted Third Party between the data provider and the researcher, and analysis environment is strictly controlled and shielded. The ODISSEI Secure Supercomputer facilitates analysis of any sensitive data, and research communities ranging from humanities to health sciences have shown interest.

---

[3] www.share-project.org

[4] For an overview of the projects, see odissei-data.nl/nl/2019/05/liss-projects-2018/

The ODISSEI Secure Supercomputer provides the fastest computing capacity in the Netherlands via the national supercomputer[5] and obeys the highest standards for data protection imposed by both Statistics Netherlands Law and Europe's General Data Protection Regulation (GDPR), both legally and technically. The ODISSEI Secure Supercomputer: (1) Is powerful, offering unprecedented computing capacity to social scientists that had previously been unattainable within the Statistics Netherlands microdata services; (2) Provides access to tools and software as well as data ensuring that scientists are fully equipped to answer the most complex of social problems; and (3) Is interoperable with other research infrastructures through a shared architecture provided by SURFsara.

The three pilot projects that have been successfully completed thus far, show the extraordinary possibilities. These initiatives provide only a glimpse of the potential of ODISSEI but already demonstrate a seismic shift in social science analysis as part of the digital data revolution.

Pilot Project 1: New Dimensions in Geospatial Analysis

The interest in neighbourhood effects goes back decades, but there is a lot of discussion on their importance and magnitude. There are two problems in the literature. First, neighbourhoods are often defined by using standard administrative neighbourhoods, such as census tracts, postal codes or other administrative units. The size of these units can vary from several hundreds of people to tens of thousands. Second, the neighbourhood context is often measured at one point in time, overlooking that the spatial context affects people over the course of their lives. This ODISSEI pilot project from TU Delft delivers both a conceptual and methodological contribution to the literature by developing a multi-scale and longitudinal approach to measuring **bespoke spatial contexts** of people. A multi-scale approach allows researchers to assess which spatial scale is more important for which people, and at which point in their lives, and whether there are differences between cities and urban structures in these spatial context effects. Analyses of heterogeneity in intergenerational mobility across spatial contexts or in very specific subpopulations become possible on a scale that is unprecedented in social research.

Figure 2 - The proportion of the population that are non-western and the degree of entropy in Amsterdam. For more information visit: odissei-data.nl/nl/2018/10/tudelft-odf-pilot/

The researchers used the individual-level characteristics of every person in the Netherlands. These grid cells were then used to construct spatial context characteristics (for example on poverty) for 100 different spatial scales surrounding every individual. Such analyses have been carried out before, but often using highly selective or problematic data sources, and rarely covering a whole country. Statistics Netherlands microdata allow much more precise and comprehensive analyses. Starting from very small 100 by 100 metre grid cells, a distance profile of the residential context was calculated for each individual up to 10 by 10 kilometres. For a single indicator, this process would take months to calculate in the standard Statistics Netherlands' secure microdata facility. There are nearly 400,000 inhabited grids cells in the Netherlands and the calculation of longitudinal entropy values of distance profiles generates more than 700 million new data points. The parallelisation that is made possible in **the ODISSEI Secure Supercomputer reduces this process from months to hours**, allowing much more experimentation with different operationalisations, and additional insights to be published in less time.

Pilot Project 2: Insights from Combining Socio-genomics and Administrative Data

Reaching out to disciplines such as behavioural genetics, socio-genomics, and genetic epidemiology is of increasing importance for the social sciences. However, linking genome-wide genetic data, or transcriptomic, epigenetic or other 'omics' data to social science data poses a set of challenges: the sheer size of the genetic
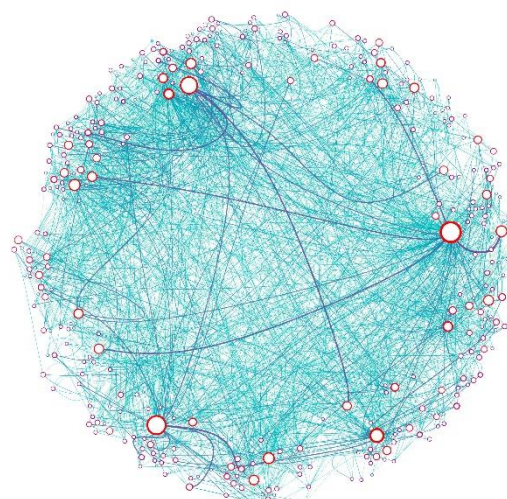
---

data, computing requirements, privacy concerns and the requirement that data available at Statistics Netherlands could thus far only be analysed within its own secure remote access facility. In this ODISSEI pilot project, the Netherlands Twin Register (NTR), Statistics Netherlands and SURFsara developed algorithms and solutions for these challenges. NTR collects genetic data that currently contains over 40 million data points ('Single Nucleotide Polymorphisms', SNPs) per individual. These data had to be linked with data collected at Statistics Netherlands.

Linkage of phenotype and register data was done using the standard pseudo-anonymisation of person identifiers at Statistics Netherlands (see section 2.4.9.3) for participants in the NTR who had provided their consent. This procedure was first tested by the NTR researchers in a proof-of-concept study for NTR data on genetic liability to schizophrenia and Statistics Netherlands data on population density (Colodro-Conde et al., 2018), resolving a longstanding question in psychiatry about the aetiology of the relation between schizophrenia and urbanicity. Next, a study with full genome-wide data was done. An additional privacy protection challenge for working with these data is that they could be not organised through classical linkage keys. These data need to be analysed with a defined order of subjects in the data file, thereby potentially creating a means to identify persons. To solve this problem, a **pseudorandom record shuffling procedure** was developed. This was successfully executed by SURFsara as a Trusted Third Party in such a way that neither the NTR researchers nor those involved at Statistics Netherlands could relate the identities of the subjects to the data.

The NTR pilot study was tailored to the current possibilities of the ODISSEI Secure Supercomputer and entailed a Genome-Wide Association study (GWAS) including millions of association tests between the genetic data (SNPs) and health care expenditures held at Statistics Netherlands. The sample with genotype data that could be linked to Statistics Netherlands data consisted of ~15,000 individuals and the GWAS was an important step in successfully demonstrating that linking genetic data to administrative data is feasible. This opens up an unprecedented range of new research areas as many longitudinal panels in the social sciences have now begun to include genotype information (as well as other big data such as Magnetic Resonance Imaging (MRI) and recordings from wearables). In this proposal, the ODISSEI Secure Supercomputer will accommodate these data and make it possible to link multiple longitudinal panels simultaneously to integral information at Statistics Netherlands. Running new and innovative genetic analyses on multiple traits, outcomes over time and gene-environment interactions will provide the opportunity to obtain insights into the role of genetics in social phenomena within Dutch society. For example, it will allow researchers to track genetic effects related to mental health on social inequalities in educational and health outcomes throughout the lifespan and allow the investigation previously established associations that might reflect **gene-environment interplay** rather than causal effects.

Pilot Project 3: Whole Population Network Analysis

One of the potentially powerful applications of administrative microdata is to examine how people are linked together through their work, education, neighbourhoods and families and how these networks structure resources and inequalities. This has normally been done on a limited scale, for example looking at how parents and children are correlated on specific variables or assessing the difference in outcomes between siblings. The integration of Statistics Netherlands' whole population data into the ODISSEI Secure Supercomputer has removed this limitation. Researchers from Statistics Netherlands are now able to analyse individuals' whole network of connections. The researchers have created a dataset comprising all 17 million residents of the Netherlands and have linked every person to their neighbours, their colleagues, their relatives and people whose children attended the same school. This is made possible by the high quality, longitudinally consistent identifiers that capture not only individuals but also the institutions to which they are connected. The research team created a dataset of over

Figure 3 - The intensity of first order family connections between municipalities in the Netherlands. For more information visit: odissei-data.nl/nl/2018/10/cbs-odf-pilot/

800 million links which describe the **latent structure of social network opportunities** within the Netherlands.

The created network opens up numerous research possibilities: e.g. on segregation/integration of networks of different social groups, inequalities, but also on vulnerabilities in network opportunities, on 'contagion' of social behaviours (such as fertility, divorce or crime), and so on. As a first example of the possibilities, the researchers

investigated exposure to others with a different migration background. They applied a **random walk procedure** over the network for every individual in the country. In effect, this procedure would select a connection within someone's network at random and then see how many connections would have to be 'travelled through' until someone from a different migration background is encountered. This random walk procedure can summarise how 'exposed' an individual person is to persons of a variety of different backgrounds. Random walk analyses are not unusual in network analysis but were never performed on this scale: the team was able to investigate for the entire population of the Netherlands how this exposure to various groups varied between age groups, genders, or even specific towns and cities.

The analyses so far give only a glimpse of the possibilities which ODISSEI is eager to further explore and facilitate. In a next step, these analyses can be extended to any number of indicators, for instance income, education or other indicators of social inequality. Social network analysis with this depth, detail and scale is unprecedented and unique. It is made possible only through the combination of the supercomputer at SURFsara and the high-quality administrative data at Statistics Netherlands. If this work is continued, Dutch social sciences can become front runners in the application of network analysis for social research. Such data can be linked to any respondent participating in an ODISSEI data collection or Mass Experiments Online Laboratory, providing insights on individuals' lives which are currently not possible through traditional research methods, revealing new dimensions of inequalities and extended over time to reveal the changing nature of the society around us.

2.1.1.8   Researchers and research groups involved

To ensure that the infrastructure remains focused on servicing the social science research community, a leading team of scientists has been appointed to oversee the development of ODISSEI and direct its implementation (Table 2). These scientists are drawn from across the social sciences, irrespective of their affiliation, and supplemented with expertise from methodology, data science, information science, and computer science. The team also includes a group of young researchers who will receive opportunities via ODISSEI to consolidate their academic careers. The specific role of each scientist is elaborated in section 2.4.9 of the proposal.

Name and title	Specialisation	Institution	DAI or ORCID
Management Board			
Pearl Dykstra	Sociology	Erasmus University Rotterdam	75229773
Marcel Das	Methodology	CentERdata/Tilburg University	137115008
Tom Emery	Sociology	NIDI	0000-0001-6137-9577
Jacco van Ossenbruggen	Computer Science	VU Amsterdam	215695461
Tanja van der Lippe	Sociology	Utrecht University	073416320
Dorret Boomsma	Genetic Epidemiology	VU Amsterdam	100730663
Rob van Nieuwpoort	Computer Science	Netherlands eScience Center	255170823
Other key researchers			
Frank van Harmelen	Computer Science	VU Amsterdam	298805235
Herbert van de Sompel	Information Science	DANS	0000-0002-0715-6126
Rens Vliegenthart	Communication Science	University of Amsterdam	298806991
Daniel Oberski	Methodology	Utrecht University	331779420
Marike Knoef	Economics	Leiden University	305472631
Michel Dumontier	Data Science	Maastricht University	0000-0003-4727-9435
Kees Mandemakers	History	IISG	069702624
Arnout van de Rijt	Sociology	Utrecht University	0000-0002-4208-3452

Table 2 – Involved key scientists

2.1.1.9   Attracting top talent

ODISSEI represents an opportunity for the Netherlands to strengthen its position in social science research by providing researchers with unprecedented access to data and the computing capacity, skills, and tools needed to analyse them. There are many institutes and facilities across the world who provide the facilities to analyse large and complex datasets, but ODISSEI will provide data access and computing capacity in combination, by bringing together government data with research data and big data within the national ICT infrastructure at SURFsara.

To attract top talent, ODISSEI will prioritise the recruitment of the best early career researchers and provide opportunities for career progression. Specifically, it will provide free and easy access to the ODISSEI Data Facility,

small grants for innovative projects exploring the use of big data in social research, the ability to run experiments through the LISS panel, grants for microdata projects at Statistics Netherlands, and actively support interdisciplinary collaborations with leaders in related fields. ODISSEI will also seek additional international recruitment and retention through initiatives such as the Marie Curie programme and Horizon 2020 calls, and in collaboration with international EU funded projects such as SSHOC (Social Sciences & Humanities Open Cloud) and international consortia such as CESSDA (Consortium of European Social Science Data Archives - see section 2.1.2.2). The capabilities of ODISSEI will be actively promoted at major international conferences across the social sciences in Europe and beyond and these will be supplemented with short term research stays for international researchers wanting to come and utilise ODISSEI.

2.1.2   Embedment of the investment

2.1.2.1   The consolidation of research groups within the Netherlands

ODISSEI is **supported by all deans** of social sciences and all deans of economics and business administration in the Netherlands. It is unprecedented in the level of support and commitment received. ODISSEI currently has 39 member organisations (Table 3), including faculties, governmental research agencies, Royal Netherlands Academy of Arts and Sciences (KNAW) and NWO institutes, and Statistics Netherlands. The number of member organisations is constantly growing. All member organisations show their **commitment** by paying an annual cash contribution determined by the size of their research staff. For the period 2020-2029, the cash and in-kind commitments of member organisations combined already totals € 12.18 million.

Member organisation		Member organisation	
CentERdata	€ 5.000	Radboud University Nijmegen (RU) - Faculty of Social Sciences	€ 30.000
Data Archiving and Networking Services (DANS)	€ 5.000	Statistics Netherlands (CBS)	€ 15.000
Erasmus School of Economics (ESE)	€ 20.000	SURFsara	€ 5.000
Erasmus School of Health Policy Management (ESHPM)	€ 20.000	Tilburg Law School (TLS)	€ 20.000
Erasmus School of Social and Behavioural Sciences (ESSB)	€ 30.000	Tilburg School of Catholic Theology (TSCT)	€ 10.000
International Institute of Social History (IISG)	€ 5.000	Tilburg School of Economics and Management (TiSEM)	€ 30.000
Leiden University (UL) - Faculty of Law	€ 5.000	Tilburg School of Humanities and Digital Sciences (TSHD)	€ 10.000
Leiden University (UL) - Faculty of Social Sciences	€ 30.000	Tilburg School of Social and Behavioural Sciences (TSB)	€ 30.000
Maastricht University (UM) - Faculty of Science and Engineering	€ 5.000	TU Delft (TUD) - Faculty Architecture and the Built Environment	€ 15.000
Maastricht University (UM) - School of Business and Economics	€ 30.000	Twente University (UT) - Faculty of Behavioural, Management and Social Sciences	€ 20.000
National Institute for Public Health and the Environment (RIVM)	€ 15.000	University of Amsterdam (UvA) - Faculty of Economics and Business	€ 30.000
Netherlands Bureau for Economic Policy Analysis (CPB)	€ 15.000	University of Amsterdam (UvA) - Faculty of Social and Behavioural Sciences	€ 30.000
Netherlands Environmental Assessment Agency (PBL)	€ 10.000	University of Groningen (RUG) - Faculty of Behavioural and Social Sciences	€ 30.000
Netherlands eScience Center (NLeSC)	€ 5.000	Utrecht University (UU) - Faculty of Geosciences	€ 15.000
Netherlands institute for health services research (Nivel)	€ 10.000	Utrecht University (UU) - Faculty of Social Sciences	€ 30.000
Netherlands Institute for Social Research (SCP)	€ 10.000	VU Amsterdam (VU) - Faculty of Behavioural and Movement Sciences	€ 15.000
Netherlands Institute for the Study of Crime and Law Enforcement (NSCR)	€ 5.000	VU Amsterdam (VU) - Faculty of Science	€ 5.000
Netherlands Interdisciplinary Demographic Institute (NIDI)	€ 5.000	VU Amsterdam (VU) - Faculty of Social Sciences	€ 15.000
Open University (OU)	€ 10.000	VU Amsterdam (VU) - School of Business and Economics	€ 30.000
Radboud University Nijmegen (RU) - Faculty of Management Sciences	€ 20.000		

Table 3 – Annual Contributions of ODISSEI member organisations

In the Netherlands, there is a longstanding tradition of collecting high-quality survey data. The **large data collections** (listed below) are owned and operated independently by ODISSEI member organisations, with ODISSEI coordinating activities on shared risks, challenges and opportunities. These data collections cover a diverse range of topics including household income, personal finances, labour market activity, political behaviour,

attitudes and values, family dynamics, child development, health, ageing, and education. Each data collection is designed to address specific scientific questions and thus it would be inaccurate to suggest that there is duplication within the community. The data collections have been funded by a diverse array of funding instruments at the European and national levels, totalling more than **€ 200 million** in existing investments. However, inefficiencies in the way these data collections operate can be identified. By implementing joint standards, utilising joint tenders, pooling methodological and technical expertise, and intensified knowledge exchange, the community of data collections will reduce costs, accelerate the adoption of new technologies, and instigate interdisciplinary cross-pollination. Before ODISSEI, there existed no formal mechanism for cooperation. In cooperating via the Observatory, the data collection community benefits from economies of scale, data complementarities and knowledge exchange.

- Children of Immigrants Longitudinal Survey (CILSNL)
- DNB Household Survey (DHS)
- European Social Survey (ESS-NL)
- European Values Study (EVS)
- Generations & Gender Programme (GGP)
- Health Behaviour in School-Aged Children (HBSC)
- International Social Survey Programme in the Netherlands (ISSP-NL)
- Longitudinal Ageing Study Amsterdam (LASA)
- Nationaal Cohortonderzoek Onderwijs (NCO)
- Nationaal Kiezersonderzoek (NKO)
- Netherlands Twin Register (NTR)
- Netherlands Longitudinal Lifecourse Study (NELLS)
- Research on Adolescent Development and Relationships (RADAR)
- Safety Monitor Survey (VM)
- Socio-Cultural Developments (SOCON)
- Survey of Health, Ageing and Retirement in Europe (SHARE-NL)
- Tracking Adolescents' Individual Lives Survey (TRAILS)
- YOUth cohort

2.1.2.2 ODISSEI in Europe

Various ODISSEI member organisations have a central role in social science infrastructure at the European level. CentERdata, Tilburg University, the University of Amsterdam, NIDI, and DANS are partners in the Horizon 2020 project SSHOC (Social Sciences & Humanities Open Cloud)[6], illustrating the central role of Dutch partners in the integration of Social Science Infrastructure within the European Open Science Cloud. A third of the SSHOC budget and a third of all partners **are based in the Netherlands** and, alongside the Dutch Humanities Infrastructure CLARIAH, ODISSEI has helped strengthen the role of Dutch Social Science Infrastructure within the European landscape.

The Dutch social science community played a crucial role in establishing the Survey for Health, Ageing and Retirement in Europe (SHARE) governance structure at the European level, acting as the first legal and administrative seat of SHARE-ERIC[7]. CentERdata developed the SHARE survey instrument using Blaise[8], a world leading survey software produced by Statistics Netherlands. 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[9]. 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[10]. The Netherlands plays a pivotal role in the European Values Study (EVS) since its inception[11]. 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 Study[12]. The **leading position** of the

---

[6] www.shopencloud.eu/
[7] www.share-project.org/organisation/share-eric.html
[8] www.cbs.nl/en-gb/our-services/blaise-software
[9] seriss.eu/about-seriss/work-packages/wp4-interactive-tools-for-cross-national-surveys/
[10] www.ggp-i.org/about/background
[11] europeanvaluesstudy.eu/category/organization/board-of-evs-foundation/
[12] www.worldvaluessurvey.org/wvs.jsp

Netherlands in the European Social Survey[13] (ESS) project is underlined by its 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-ERIC)[14] 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. In 2016, CESSDA, ESS and SHARE were recognised as landmarks of the European Strategy Forum on Research Infrastructure (ESFRI) in the field of social and cultural innovation.

In ODISSEI, these infrastructures will be coordinated for their mutual benefit and strengthening the position of Dutch Social Science within the **European research area**. Through the partners in ODISSEI, further collaboration with international research groups will be actively pursued, especially in the context of the European Open Science Cloud. As the European Open Science Cloud develops further, tools developed by ODISSEI partners will be made available to the wider European research community and provide a product development pathway that begins in the ODISSEI Laboratory and leads into full production within the major international surveys. This vision of integrating ODISSEI within the European Open Science Cloud framework opens up future possibilities for the infrastructure's development and strengthening of the Dutch position in European social science innovation.

In addition, ODISSEI will seek to make several other services and tools available via the **European Open Science Cloud**. The protocols that govern the Secure Supercomputer at SURFsara can be utilised by universities, research institutes and statistical offices outside the Netherlands. This would allow the services developed in ODISSEI to become more sustainable but would also open up new lines of research for European partners and the tantalising possibility of analysing secure data in a comparative framework for the first time. This would be a revolutionary leap forward for the social sciences and our understanding of societies. SURFsara will explore the feasibility of this in this project. In addition to the ODISSEI Data Facility, there are several other services and tools developed in ODISSEI that could be brought into the European Open Science Cloud such as the tools in the Media Content Analysis Laboratory and the Mass Experiments Online Laboratory. ODISSEI is therefore emerging at a crucial juncture in the development of social sciences and their utilisation of e-infrastructures within the European Open Science Cloud.

2.2 Strategic case and innovation

2.2.1 The importance for Dutch science and international positioning and appeal

2.2.1.1 Long term planning of the social sciences and economics

In 2018, Mark Bovens (Public Administration) was appointed by the VSNU (the Association of Dutch Universities)[15] to consult with NWO, Royal Netherlands Academy of Arts and Sciences and the Young Academy with regards to the needs of the social science and humanities domain over the next six years. The subsequent report (called the **SSH Sector plan**) highlighted the excellent quality and reputation of Dutch social sciences and the vital role that ODISSEI plays in linking data and facilitating research. The report[16] (in Dutch) also highlighted the urgent need for investment in ODISSEI so that it can be continued and strengthened.

Similarly, the Royal Netherlands Academy of Arts and Sciences has developed the **Academy Agenda** which sets out researcher's 'dream' facilities. These facilities are described with a view to begin construction in 2025 or thereafter. The **M3 infrastructure**[17] is part of this agenda. M3 aims to bridge the studies of populations across health sciences, social sciences and humanities by linking data on molecules, people, and society. ODISSEI is a prerequisite for such an ambitious program, providing secure data linkage facilities, shared standards, and the basis for data linkage across disciplines. M3 provides a long-term strategic focus for ODISSEI and represents key and essential investments for the realisation of the Academy Agenda.

Both the SSH Sector plan and the Academy Agenda envisage the integration of existing infrastructures. The SSH Sector plan is supporting the development of an infrastructure for legal studies that will sit alongside ODISSEI and CLARIAH. ODISSEI will also seek alignment and collaboration with BBMRI in the health sciences, given the potential synergies that are outlined in M3. Cooperation and long term integration across these infrastructures is highly desirable and this proposal sets out **several bridging activities** which lay the foundations of such work including the integration of National Twin Registry and the Historical Sample of the Netherlands as ODISSEI data

---

[13] www.europeansocialsurvey.org/
[14] www.cessda.eu/About/Consortium/CESSDA-Countries/CESSDA-Members/Netherlands
[15] www.vsnu.nl/en_GB
[16] www.sectorplan-ssh.nl/binaries/content/documents/sectorplan-ssh-nl/algemeen/documenten/samen-sterker.-beeld-van-het-ssh-domein/samen-sterker.-beeld-van-het-ssh-domein/subsites%3Adocument
[17] www.knaw.nl/shared/resources/adviezen/bestanden/KNAWAgendaM3.pdf

collections, interdisciplinary use of the ODISSEI Secure Supercomputer and the alignment of closely related tools such as the Media Content Analysis Lab (quantitative analysis) with the CLARIAH Media Suite (qualitative analysis).

2.2.1.2   The position of the Netherlands in international social science

In ODISSEI, the Netherlands will have the opportunity to strengthen its position within international social science. ODISSEI is a **unique project**, not only in the Netherlands but globally. No other national data infrastructure includes the national statistical office (Statistics Netherlands), high-performance computing (SURFsara), the national data archive (DANS) as well as such a broad diversity of independent large-scale data collections and big data. No other national data infrastructure has such a strong institutional base in the social science research community. There are several initiatives in other countries that share some of the ambitions of ODISSEI but none have been able to bring so many datasets together in a single community and engage with so many distinct data collections.

In France, Progedo[18] brings together the major national social science infrastructures that are on the European Strategic Forum on Research Infrastructures (ESFRI) roadmap including Consortium of European Social Science Data Archives (CESSDA), European Social Survey (ESS), Generations and Gender Programme (GGP), and Survey for Health, Aging and Retirement in Europe (SHARE). Like ODISSEI, Progedo encourages knowledge exchange and cooperation between the infrastructures, distributing funds to the national nodes of these infrastructures to bolster their contributions to the international projects. In Germany, GESIS[19], an internationally renowned institute, is developing a national survey data infrastructure called the Integrated Data and Survey Infrastructure (IDSI), which aims to provide a single data facility for innovations in survey research and methodology. This facility aims to address many of the same challenges as all social science infrastructures, including low response rates and a changing technological landscape. Neither Progedo nor IDSI have yet been able to provide administrative data access that links to these data collections within a secure high-performance computing environment.

One of the foremost centres in Europe regarding data collection activities is in the United Kingdom and this reputation is highly warranted. The United Kingdom has invested heavily in survey research through its world-renowned cohort studies, household panels, educational studies, and a wide range of other sustained and high-quality longitudinal data collections. The British Household Panel has been transformed into the ground-breaking and highly innovative Understanding Society survey, with around 40,000 households. The United Kingdom also has several excellent initiatives that look to improve the United Kingdom's data infrastructure including the United Kingdom 'Biobank[20] and the Urban Big Data Centre[21]. These are world-leading research centres. However, the United Kingdom is organisationally complex and this has inhibited linkage with administrative data sources underlined by the problems encountered by the United Kingdom Administrative Data Service (ADRS) which has been discontinued despite significant investment. This illustrates the need, not only for investment, but also for the buy-in and collaboration of data owning parties such as the national statistical office and key data owners.

In the United States there are several excellent and rapidly developing facilities for the secure analysis of administrative data such as the Massive Data Institute at Georgetown University[22] and the Coleridge Initiative[23]. These facilities share many of the same challenges as ODISSEI and collaboration will be sought on shared aims. However, the challenges facing initiatives in the United States are different from those in the Netherlands given the highly federated and diverse nature of the government and its data infrastructure. ODISSEI is aided by the Netherlands relatively centralised and standardised data structures which represent a scientific comparative advantage.

The Netherlands must invest in its data infrastructure for social science if it is to stay at the forefront of these developments. Not only this, but given the rapid developments in infrastructure and technology, now is also exactly the right time to be doing so.

---

[18] www.progedo.fr/en/
[19] www.gesis.org/en/institute/
[20] www.ukbiobank.ac.uk/
[21] www.ubdc.ac.uk/
[22] mccourt.georgetown.edu/massive-data-institute
[23] www.coleridgeinitiative.org/

2.2.2 The importance for society and industry and the connection with societal developments

ODISSEI is highly relevant for issues included in the central research agendas of the Netherlands: the Dutch Top Sectors Research Agenda[24], the Digital Society Agenda of the Association of Universities in the Netherlands (VSNU)[25], the Dutch National Research Agenda[26] and the SSH Sector plan (see section 2.2.1.1). Furthermore, ODISSEI can make important contributions to the Societal Challenges defined by the European Commission[27].

2.2.2.1 Dutch Top Sector Policy

The Dutch Top Sectors do not operate in isolation. They are embedded in and influenced by social structures, such as the educational system, the labour market, social security, governance and the public interest. Organising these structures appropriately creates a fertile ground, enabling the individual Top Sectors to perform well. ODISSEI will provide the kinds of multi-dimensional and multi-level data that are crucial for gaining insight into the major societal challenges where social and natural systems interact such as health, food, water, energy, materials, climate, transport, urbanisation, and security, as identified in the so-called Social Infrastructure Agenda[28]. **Examples of questions were ODISSEI can help**, are: What conditions ensure a highly trained, technically skilled workforce? How can the Netherlands remain an appealing business location? What conditions contribute to political, economic, and social stability in the Netherlands? What are the conditions for a healthy, safe, self-sufficient, happy society? The integrated approach in ODISSEI will facilitate answering these questions by using insights from different large-scale studies in combination with administrative and novel forms of data. ODISSEI will also facilitate the diffusion of methodological innovations in the collection, processing, and analysis of data that is increasingly vital in a knowledge economy.

2.2.2.2 The Dutch National Research Agenda

The Dutch National Research Agenda (NWA) was developed by the Dutch knowledge coalition and represents the focus of Dutch science over the coming years (De Graaf & Rinnooy Kan, 2017). It provides a large number of research questions and ODISSEI is the central infrastructure to give better and sustainable answers to many of these questions. ODISSEI directly contributes to at least 11 of the identified 25 areas. These areas are given below, with examples of exactly how ODISSEI will contribute to advancing this agenda provided below.

Creating value through responsible access to and use of big data: Responsible access to secure sensitive data is a key aim of ODISSEI. Researchers will be able to choose from a diverse range of computational resources to suit their needs and will be supported in the technical, legal, and ethical use of these facilities (task 1.1).

Logistics and transport in an energetic, innovative and sustainable society: ODISSEI data collections contain a diverse range of measures on mobility, transport, use and energy use. These data collections will be made more sustainable through ODISSEI and innovations that facilitate flexible and adaptive data collection will be supported (e.g. via smartphone and GPS) (task 2.1).

Circular economy and resource efficiency: sustainable circular impact: Monitoring and analysing resource use requires access to large, diverse, and complex forms of data. ODISSEI Data Scouts will identify and integrate data from across the economy and make them linkable in a secure environment (task 2.2).

Health care research - sickness prevention and treatment: The ODISSEI Data Facility will be used to link data from various biomedical studies. Pooling such datasets will support the study of rare conditions and small sub-populations (task 2.2).

Living history: The Historical Sample of the Netherlands holds data on a representative sample of Dutch residents dating back to the early 19th century. In ODISSEI, this sample will be linked with the population of today to enable researchers to explore how current behaviours, outcomes and patterns are rooted in the country's past (task 2.2).

Quality of the living environment: The ODISSEI data collections contain a wealth of information on attitudes to the environment and environmental behaviours. In the Observatory, these data collections will be extended and it will be made possible to link them with detailed neighbourhood level data on peoples immediate social environments (task 2.3).

---

[24] www.government.nl/topics/enterprise-and-innovation/contents/encouraging-innovation
[25] www.vsnu.nl/files/documents/Publications/VSNU_The_Digital_Society.pdf
[26] www.wetenschapsagenda.nl/wp-content/uploads/2016/12/nwa_deel_eng_digitaal.pdf
[27] ec.europa.eu/programmes/horizon2020/en/h2020-section/societal-challenges
[28] www.nwo.nl/en/research-and-results/programmes/magw/the-social-infrastructure-agenda/index.html

Smart liveable cities: The borders of smart cities are defined by how citizens move and interact, not by the boundaries of municipalities and data owners. To understand smart cities and make them liveable, it is necessary to link social data from multiple sources without violating the privacy and security of the data subject. ODISSEI will explore the use of distributed computational methods which could help breakdown the boundaries and barriers to data flows within smart cities (task 3.1).

Towards resilient societies: We live in a networked society and the behaviours exhibited by networks can often be dramatic and unpredictable, undermining their own resilience. The Mass Experiments Online Laboratory will allow for cutting edge research on network behaviours and their consequences (task 3.2).

Measuring and detecting - anything, anytime, anywhere: With smart devices in proliferation and data flowing from all corners of our lives, it is possible for anyone to become a citizen scientist and contribute measurements and data to scientific research. The Citizen Science Platform will develop methodological tools to improve data quality from these projects and support this trend (task 3.2).

Child and adolescent upbringing and education: ODISSEI data collections provide longitudinal evidence on the development of children and adolescents that will be enhanced by the application of new innovative measures developed in the Laboratory (task 3.3).

Sport & exercise: New biomarkers will be tested and deployed in the LISS panel which capture people's activity levels and provide insights into individuals every day health and exercise regimes (task 3.3).

2.2.2.3 The Digital Agenda

As a response to the National Research Agenda, the VSNU Digital Society agenda focuses on effective connections between digital technology, people, and their society. It notes that to establish the Netherlands' as a 'digital main port' requires new data technologies, research, and applications. This means investing in high-end facilities, services, and technologies. ODISSEI will also have a vital role to play in this research agenda by:

1. Collecting, processing, and linking data across diverse areas to facilitate social and economic innovations;
2. Contributing to methodological and technological breakthroughs in the use of new technologies in social science;
3. Creating an open national data infrastructure for the social sciences;
4. Training the next generation of social and data scientists.

In recent years, innovative data collection and retrieval have become one of the most promising and interesting developments in the social sciences. Research based on small, one-off surveys still play a vital role, but is now often enriched with complementary data collection in big data endeavours like social media, mobile phones, and wearable devices such as activity meters, sphygmomanometers, and GPS trackers. Without knowing exactly where these current trends will lead, it is safe to say that we are experiencing revolutionary developments in data infrastructure. ODISSEI will invest in innovations in data collections through its Laboratory and actively stimulates their uptake and application through its Observatory. Below are a few examples of existing and future innovations within LISS and ODISSEI data collections:

- **Avatars and voice recognition**: Interviewer-respondent interactions affect the quality and content of interview data collected. By using avatars, one can tailor the 'interviewer' to a respondent's preferences. Respondents who do not like reading or typing, or who have difficulty doing so can be interviewed online or using a home assistant using an avatar, while their responses are recorded by voice recognition software (Conrad & Schober, 2007; Pickard et al., 2016).

- **Biomarkers**: This is a very broad class and includes various objective health measures. ODISSEI studies such as LASA, TRAILS, RADAR, and NTR have been collecting extensive biomarkers including blood pressure meters, glucose measures, step counters, devices using galvanic skin response, cognitive tests, MRI, and many more (Hoogendijk et al., 2016; Willemsen et al., 2013; Oldehinkel, 2015). The measurement of physical activity using accelerometers also falls in this class, as does the use of advanced electronic bathroom scales to objectively measure weight and fat percentage (e.g. Kapteyn et al., 2018; Kooreman & Scherpenzeel, 2014). One of the biggest challenges in the medical field is monitoring compliance with drug regimens. One can use chips to detect when a drug is taken or when a note is entered into a diary (e.g. Stone et al., 2002).

- **Webcams for eye-tracking**: Eye-tracking is an important tool to monitor what a respondent looks at when answering surveys. This indicates what part of the screen receives most attention or whether some parts are missed altogether (Galesic et al., 2008; Lenzner et al., 2011). Until recently eye-tracking was

mainly done in a laboratory setting with fairly complicated devices. Eye-tracking software is reaching the stage where it can be done with a simple webcam in respondents' homes, so that in principle this technology can now be used in population surveys.

- **High frequency data collection**: In traditional surveys, either by phone, mail, or by a personal visit by an interviewer, it is infeasible to have frequent interviews. With internet surveys, one can go back frequently and for instance ask a daily time-diary, or even prompt respondents several times during the day on their cell phone. The frequency of the use of sensor technology in experiments, e.g. on health, crime prevention, education, or consumer behaviour, is determined by how often measurement takes place.
- **Interventions**: Technology makes interventions easier to implement and scalable. An example would be an intervention aiming to induce a healthier lifestyle through regular reminder text messages, in combination with feedback through the electronic devices that objectively measure health indicators (such as weight, fat percentage, and physical activity).

These are only examples, which moreover are already technically possible today. No doubt many more opportunities for cost effective and accurate data collection will emerge in the years to come. ODISSEI is well positioned to take advantage of those developments due to the broad scope of the data collections involved, its dedicated innovation panel at LISS, its intensive integration of big data and its incorporation of expertise from SURFsara and NLeSC.

2.2.2.4 Dutch & European societal challenges

Prominent (policy) research institutes in the Netherlands, such as the Netherlands Bureau for Economic Policy Analysis (CPB), the Netherlands Institute for Social Research (SCP), the Dutch Central Bank (DNB), the Netherlands Institute for Transport Policy Analysis (KiM), the Netherlands Environmental Assessment Agency (PBL), the National Institute for Public Health and the Environment (RIVM), and the Research and Documentation Centre of the Ministry of Justice and Security (WODC) will benefit from ODISSEI. The use of new technology and data science, and the development of new innovative instruments for policy making and privacy safeguards will further improve and enable **evidence-based policy making**. The policy research organisations consider the ODISSEI infrastructure of great importance in terms of valorisation and use for policy and societal relevant themes including the realisation of the digital agenda. Several are already ODISSEI member organisations and during the project, the others will be invited to join.

ODISSEI will also contribute in important ways to tackling the societal challenges identified by the European Commission including:

- **Health, demographic change, and wellbeing**: ODISSEI member organisations hold survey data from important comparative and national level studies that look at ageing (SHARE), family change (GGP), and genetics (NTR). ODISSEI combines these with administrative data and innovative forms of data collection such as the use of smart scales and other biomarkers to create powerful insights into health over the course of life.
- **Smart, green, and integrated transport**: GPS tracking is already used by ODISSEI partners and provides insights into how individuals move around. However, using expertise from NLeSC, ODISSEI will glean more from this data by using machine learning technologies to identify distinct patterns of travel behaviour that can be matched to specific transport modes. This technology negates the need for complex travel diaries and provides a clear and accurate measure of transport usage that is integrated within wider social science infrastructures.
- **Climate action, environment, resource efficiency, and raw materials**: ODISSEI member organisations cover a wealth of information on individuals' consumption habits, including energy consumption. Combining household surveys on consumption such as in the LISS panel and European Union Statistics on Income and Living Conditions (EU-SILC) with innovative data collection methodologies such as smart thermostats and small area measures of green infrastructure enables the creation of a more holistic picture of energy consumption behaviour in the Netherlands.
- **Europe in a changing world - inclusive, innovative, and reflective societies**: The capacity to address this challenge is vast given that ODISSEI member organisations hold a wealth of data on attitudes, values, communications, and behaviours. For example, ODISSEI includes the CILSNL study which tracks the children of immigrants, their social integration, and outcomes. Studies on inclusion and exclusion will be greatly supported by ODISSEI as data linkages enable the local area context to be captured and specific subpopulations to be targeted for analysis.

- **Secure societies - protecting freedom and security of Europe and its citizens**: ODISSEI includes the Safety Monitor Survey (VM) which is an annual population survey on safety, liveability, and victimisation. The VM is carried out on behalf of the Ministry of Justice and Security, Statistics Netherlands, municipalities, and police. The VM has a sample size of 100,000. This large sample is perfect to be combined with local area geo data and population registries to examine individual and neighbourhood causes and consequences of crime and feelings of unease. The computing power and data infrastructure that was previously lacking for such analysis will become possible through ODISSEI.

2.3   Management case
2.3.1   Organisation and governance

ODISSEI is a consortium which includes faculties, Statistics Netherlands, 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 and entered a development phase starting in early 2017. ODISSEI now consists of 39 member organisations which contribute to an operational budget for the infrastructure (see section 2.1.2.1 for an overview of the member organisations). To pursue ODISSEI's strategic aims, to elaborate on the work plan, and to implement the infrastructure operations, ODISSEI has developed a clear governance and organisational structure based on the recommendations in the OECD Global Science Forum report on the sustainability of Research Infrastructures[29] and the InRoad project recommendations of the European Strategic Forum on Research Infrastructures[30].

ODISSEI governance consists of the Supervisory Board, the Management Board, and the Advisory Board (Figure 4). ODISSEI's host member organisation is Erasmus University Rotterdam (EUR) who acts as the host of the Coordination Team and secretariat to all governance bodies. The ODISSEI governance structure was unanimously approved by all ODISSEI member organisations and written down in By-Laws.
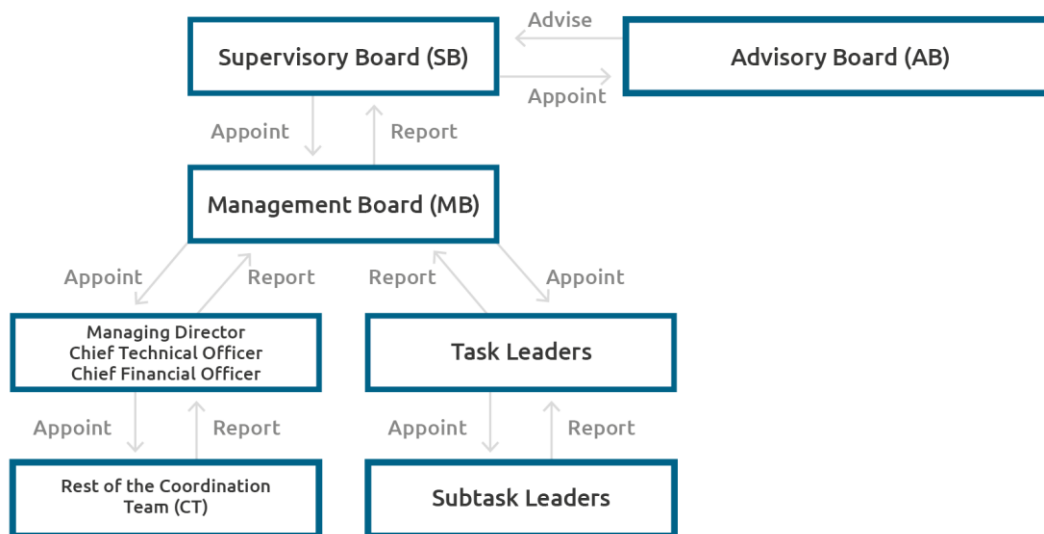

Figure 4 – Governance structure

2.3.1.1   Supervisory Board

The Supervisory Board is the highest body and consists of seven members who are elected by the member organisations. The group is broad enough to represent the diverse member organisations of ODISSEI, yet succinct enough to enable effective oversight and consultation. It is composed of leading scientists and senior managers. The Supervisory Board meets at least once a year to approve the scientific agenda, strategic direction, the annual plan of work, and the budget. Supervisory Board members are responsible for collating the opinions and preferences of member organisations ahead of this annual meeting and communicating decisions afterwards. The Supervisory Board appoints the Management Board (see section 2.3.1.3). Other tasks of the ODISSEI Supervisory Board include:

- Approval of strategic direction of ODISSEI;

---

[29] www.innovationpolicyplatform.org/strengthening-sustainability-and-effectiveness-international-research-infrastructures-oecd-project-0
[30] www.inroad.eu/description/

- Supervision of legal, ethical and scientific integrity of the work conducted by ODISSEI;
- Definition of risk tolerance, including financial setbacks, delivery delays, quality of partner participation;
- Approval of the budget, including calls for funding and projects;
- Approval of membership fee structures;
- Requested and unrequested advice and consultancy for the Management Board.

The Supervisory Board currently consists of Pieter Hooimeijer (Utrecht University), Henk van der Kolk (University of Twente), Carlo Schuengel (VU Amsterdam), Arthur van Soest (Tilburg University), Astrid Boeijen (Statistics Netherlands), Peter van der Laan (Netherlands Institute for the Study of Crime and Law Enforcement), and Sander Steijn (Netherlands Institute for Social Research). They were elected in 2019 in accordance with the ODISSEI By-Laws. The Supervisory Board has formally approved this Roadmap proposal and was consulted extensively during its development.

2.3.1.2 Advisory Board

The Advisory Board currently consists of five members and meets at least once a year. It can be extended to up to eight members. The Advisory Board advises the Supervisory Board on content-related issues and identifies potential synergies and alignment between the different Work streams of ODISSEI. It can do so on its own initiative or upon request by the Management Board or Supervisory Board. Members of the Advisory Board are appointed by the Supervisory Board based on their expertise, notwithstanding their affiliation, residency or nationality. The Advisory Board is responsible for advising on strategic development of ODISSEI and looking for opportunities to facilitate innovation and increase cooperation with other initiatives, especially in an international context. The current Advisory Board consists of Melinda Mills (University of Oxford), Amy O'Hara (Georgetown University), Ron Dekker (CESSDA Director and EOSC Board Member), Julia Lane (New York University) and Sally Wyatt (Maastricht University).

2.3.1.3 Management Board

The Management Board is responsible for the operation of ODISSEI and meets ten times a year. The Management Board:

- Holds responsibility for the implementation of ODISSEI;
- Drafts the annual work plan, annual budget, and annual report;
- Approves the design reports of products and services and allocates budget for further development (see section 2.4.6);
- Approves production versions of products and services before launch and allocates budget for operation and further development (see section 2.4.6);
- Can appoint partners who complete tasks on behalf of ODISSEI;
- Manages risk in line with tolerances identified by the Supervisory Board (see section 2.4.7);
- Determines expenditures from ODISSEI's operating budget and is responsible for procurement;
- Implements the funding decisions, mandated by the Supervisory Board;
- Can reallocate budget and tasks in accordance with funding guidelines;
- Identifies initiatives for future funding of the ODISSEI framework in the short and long term.

When there is a budget issue or time overruns, the Management Board will be responsible for adjusting the original project plan and this requires approval by the Supervisory Board. If disputes arise between partners, the Management Board will be tasked with reconciling differences and in the event of persistent conflicts, the matter will be deferred to the ODISSEI Supervisory Board. Procurement in excess of € 100,000 per annum that falls outside the annual budget requires explicit authorisation by the Supervisory Board. The Scientific Director chairs the Management Board. The Management Board has nine members who represent the key partners and activities within the project: Pearl Dykstra (EUR - Scientific Director), Tanja van der Lippe (UU - Scientist), Dorret Boomsma (VU - Scientist), Ruurd Schoonhoven (Statistics Netherlands - Microdata Access), Marcel Das (CentERdata - LISS Panel), Irene Nooren (SURFsara - Secure Computing), Jacco van Ossenbruggen (VU - Data Science), Rob van Nieuwpoort (NLeSC - Research Software), and Tom Emery (NIDI - Data Collections).

In addition, the Management Board meetings are attended by the Managing Director, Chief Technical Officer and Chief Financial Officer from the Coordination Team (see next section) who will report on progress across the Work streams. The Managing Director, Chief Technical Officer, and Chief Financial Officer report on 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 Management Board or Supervisory Board identifies a breach by an executing organisation, the host member organisation will

give formal notice to such Party. If a breach is substantial and is not remedied within 30 days, the Management Board as well as the Supervisory Board may propose to declare a Party to be a defaulting party and the consequences thereof, which may include termination of its participation in the execution of this proposal.

2.3.1.4   Coordination Team

The Coordination Team consists of the people employed under ODISSEI at the host member organisation. ODISSEI's host member organisation is the Erasmus School of Social and Behavioural Sciences (ESSB) at the Erasmus University Rotterdam (EUR). The Coordination Team is responsible for supporting, facilitating, and monitoring the work conducted by the (sub)task leaders (see section 2.3.1.5). This includes coordination, finance, and the secretariat. Its work is included within the ODISSEI Hub.

The Coordination Team is run by the **Managing Director** who coordinates the implementation of the ODISSEI work programme. The Managing Director is appointed by the Scientific Director in coordination with the host member organisation. The initial Managing Director will be Lucas van der Meer, the current project manager. In the first year, the Coordination Team will consist of at least five positions; by the fifth year the team will have expanded to up to eight positions. The Managing Director is supported by three senior staff members who address specific areas of operations. They will be recruited at the outset of the project and will report directly to the Managing Director.

The **Chief Technical Officer** deals with implementation across ODISSEI. Each Work stream will include dedicated technical staff and will be responsible for the technical implementation of their work. The Chief Technical Officer and supporting staff at EUR will be responsible for monitoring the development of products and services within tasks and identifying interdependencies and developing contingency plans. All stages of product and service development must be approved by the Chief Technical Officer (see section 2.4.6). The Chief Technical Officer will also be responsible for identifying synergies and facilitating cross-pollination and collaboration across Work streams and where necessary, liaising with technical teams in partner organisations.

The **Chief Financial Officer** provides financial and administrative support for the operation of ODISSEI and monitors spending and budgets across Work streams. All budgets set out and submitted by partners must be approved by the Chief Financial Officer. The Chief Financial Officer is responsible for liaising with NWO and the various governing bodies within ODISSEI and ensuring that the project reporting is complete and timely. The Chief Financial Officer will also serve as the interface with the host member organisation and ensure that ODISSEI is well embedded within it. The Chief Financial Officer will ensure the budget is implemented and that it is compliant with the applicable legal and statutory requirements.

The **External Relations Officer** coordinates the external relations of ODISSEI. Many ODISSEI member organisations have existing work programmes which overlap with the aims and goals of ODISSEI such as the use of new methodologies and support for computational social science. The External Relations Officer will be responsible for liaising with communication and educational officers across all member organisations and organising events to raise awareness of ODISSEI. This includes communicating the user agreement (see section 2.3.2) developed by DANS and facilitating communication between various parts of the infrastructure. The External Relations Officer will also be responsible for developing a communications strategy in the first year of the project.

In addition to these three senior members of staff and the Managing Director, the coordination team will consist of support staff to help conduct day-to-day operations. The Secretariat will support all ODISSEI staff and governing bodies with the execution of their work.

2.3.1.5   Exploitation & Work stream management

All four Work streams are made up of three tasks, each consisting of three broadly interrelated subtasks. Each task has a designated member organisation as leader who is responsible for coordinating work between the three subtasks. Given the diverse nature of the tasks, each task leader is able to manage this coordination in the manner they see fit. Each subtask has a designated member organisation who is responsible for execution of the subtask. Some organisations are leads for multiple subtasks, resulting in 13 subtask leaders across the 36 subtasks. The subtask leaders report to the task leaders, who report to the Management Board. On a quarterly basis, the Coordination Team will meet with each task leader and the corresponding subtask leaders to discuss the progress of work and ensure the smooth running of the project. Task leaders are appointed by the Management Board. The task leaders are supported in their daily work by the Coordination Team, who are the central coordinating point for all cross-task-related issues (see section 2.4.6). The task and subtask leaders are identified in section 2.4.9.

All member organisations executing work within this proposal (called 'partner organisations') signed a multilateral consortium agreement, based on the widely used DESCA standard[31]. In it, the partner organisations 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 intellectual property (the ODISSEI partners remain owner of their work, but all of that work should be freely available to the ODISSEI member organisations).

2.3.1.6 Key Performance Indicators

Key performance indicators have been outlined to help the Management Board and the Supervisory Board track the progress and development of the infrastructure across several different areas of the project (Table 4). They were developed based on the recommendations of the ESFRI framework[32]. They relate to the specific implementation of the project but also to the broader strategic aims and development of ODISSEI. Some of the indicators have been ascribed targets whilst others are for monitoring purposes and therefore have no target. Targets will be added, updated, revised, and refined as the project develops. The list of key indicators will also be extended, and further specifications added.

Indicator	Target
Development	
Participation of economics faculties	All social science and economics faculties to be full ODISSEI members by 2025
Participation in European projects	Active participation in at least two European Level projects by 2025
International collaborations	Formal collaboration with similar initiatives in UK, US, FR, and DE
Data Access	
Percentage of researchers at ODISSEI member organisations submitting a proposal[33]	Receive project proposals from 10% of researchers at ODISSEI member organisations
Percentage of researchers at ODISSEI member organisations being granted a project	Receive project proposals from 5% of researchers at ODISSEI member organisations by 2025
Number of unique visitors to the ODISSEI Portal	No current target set
Number of projects spanning multiple Work streams	By 2025, at least five projects should have used infrastructure from multiple Work streams simultaneously
Financial	No current target set
Project money from further projects	No current target set
Total amount committed by member organisations	By 2025, to receive continuation of existing commitments to 2029 as a minimum
Training & Education	
Number of attendees at training events and workshops	No current target set
Number of faculties cooperating in educational program	All ODISSEI member organisations that are university faculty to be integrated into educational program
Number of researchers supported by Social Analytics team	No current target set
Science & Technology	
Number of peer-reviewed papers describing the infrastructure	3 per year over 5 years
Number of peer-reviewed publications derived from ODISSEI research projects	No current target set
Number of citations received for publications derived from ODISSEI research projects	No current target set
Impact for Industry	
Number of datasets contributed by industry partners	Five datasets from industry partners contributed by 2025
Number of sectors from which industry partners are derived	Partners from more than one sector by 2025
Impact for Society	
Number of government agencies participating as member organisations	No current target set
Number of website visits	No current target set
Press mentions in the Netherlands	No current target set

---

[31] www.desca-2020.eu/
[32] www.ceric-eric.eu/2018/08/30/key-performance-indicators-of-research-infrastructures/
[33] See section 2.3.2 for a list of grants available to the ODISSEI community

Communication	
Number of social media subscribers	Total of 5,000 by 2025
Number of newsletter subscribers	Total of 8,000 by 2025
Attracting New Talent	
Number of international researchers recruited by ODISSEI partners	No current target set
Number of international projects (such as MSCA[34]) using ODISSEI	No current target set

Table 4 – Key Performance Indicators

2.3.2   Accessibility

The ODISSEI Portal, the Media Content Analysis Lab and all non-sensitive data generated within ODISSEI will be freely available to researchers irrespective of their affiliation, seniority or location. For services with limited resources and capacity, access will be allocated through calls open to researchers at ODISSEI member organisations (see section 2.1.2.1). Calls are overseen by the Management Board and administered by the Coordination Team through a standardised call process. This involves a clear and transparent process with all calls disseminated across the whole ODISSEI community. All applications are then evaluated for technical feasibility by subtask leaders before being evaluated for ethical risks by the ODISSEI data officer. They are then assessed for scientific excellence by a panel of at least three domain experts from across ODISSEI member organisations. These domain experts are appointed by the ODISSEI Management Board.

This call process will be managed in alignment with NWO best practices as is currently the case. Those projects that are ranked highest for scientific excellence will be granted use of the infrastructure. Awards only provide access to the infrastructure and no cash will be transferred to successful applicants. Calls which span multiple parts of the infrastructure (e.g. LISS and Statistics Netherlands data in combination) can be put forward by task leaders and are then subject to the same process. The research staff across all of ODISSEI member organisations totals around 5,000. As a researcher at a member organisation, users are able to apply for grants associated with:

- Microdata Access at Statistics Netherlands (task 1.1);
- ODISSEI Secure Supercomputer (task 1.2);
- ODISSEI Mass Experiments Online Lab (task 3.2);
- ODISSEI Citizen Science Platform (task 3.2);
- The LISS Panel (task 3.3);
- ODISSEI eScience calls (task 4.3);

Over the lifetime of the project, across the infrastructure, approximately 150 separate projects will be granted. Usage of ODISSEI facilities or services is dependent on the acceptance of the user agreement which will be affirmed by the ODISSEI Management Board and the Supervisory Board in the first three months of the project. This **user agreement** will cover the data management, research ethics, and usage policy of the entire ODISSEI infrastructure and will be developed by DANS. All users of the infrastructure will be required to accept this user agreement. Usage of ODISSEI infrastructure by external, non-commercial parties is currently restricted to a discretionary basis in consultation with the Management Board. As the infrastructure develops further, the Management Board may seek to develop a service catalogue and pricing structure for external users that supports the sustainability of the infrastructure.

2.3.3   IT infrastructure

2.3.3.1   Required IT infrastructure

IT infrastructure is a critical part of ODISSEI and is used to advance data analysis via the ODISSEI Secure Supercomputer (see task 1.2). The ODISSEI Secure Supercomputer is a virtual part of the existing national supercomputer Cartesius of SURFsara. Cartesius consists of a large number of batch nodes and a small number of special purpose nodes. For the batch nodes ODISSEI differentiates between so-called thin nodes, fat nodes, GPU nodes and Xeon Phi nodes. Thin nodes constitute the majority of the available batch nodes. GPU nodes use GPGPUs to accelerate the computations. The Xeon Phi nodes have Intel Many-Core CPUs. Fat nodes have more memory (256 GB) than the thin nodes and more physical cores (32) than most thin nodes.

Given that the national supercomputer Cartesius is being used by other researchers at all times, virtualisation based on PCOCC software (developed at CEA in France) has been deployed. This creates a virtual computing

---

[34] Marie Sklodowska Curie Actions

environment on one or more supercomputer nodes. It is a separate environment that is shielded from the outside world. The virtual nodes retain all the properties of a physical Cartesius node, such as connections to other nodes via InfiniBand and access to (Luster) storage and data on the central ODISSEI Secure Supercomputer environment. SURFsara has configured the virtual environment in such a way that data exchange can only take place between the node and the virtual machine on which the analysis is done.

The IT infrastructure of other prominent partners including DANS, Statistics Netherlands, University of Amsterdam and CentERdata is utilised and specified, where necessary, in the task descriptions of section 2.4.

2.3.3.2   Harmonisation with existing IT infrastructures

ODISSEI will align with and use existing, high-quality IT infrastructures that are available in the Netherlands. ODISSEI's Secure Supercomputer is built upon the existing national supercomputer Cartesius of SURFsara. The ODISSEI Portal integrates existing data catalogues of DANS, Statistics Netherlands and many other repositories. The Social Analytics team is trained by NLeSC. The pathfinder grants for computational social science are run by NLeSC. A successful harmonisation is accomplished by having SURFsara, Statistics Netherlands, and NLeSC represented in the Management Board. ODISSEI represents the integration of social sciences within existing IT infrastructure and not the duplication or replication of any such facilities.

ODISSEI continuously aligns with infrastructures in other disciplines such as CLARIAH (humanities) and BBMRI-NL (health sciences) via SURFsara and sharing in solutions for joint challenges such as identification, authentication and authorisation (IAA) systems and federated data access.

2.3.3.3   IT team

ODISSEI has significant IT expertise at strategic positions to manage the IT infrastructure. Jacco van Ossenbruggen (Associate Professor of linked data at VU) and Rob van Nieuwpoort (Professor of eScience at UvA) are represented in the Management Board. The Managing Director (Lucas van der Meer) has a master's degree in computer science and several task leaders (Narges Zarrabi and Herbert van de Sompel) have doctorates and multiple years of professional experience in computer science for large scale research infrastructure. In addition, prominent scholars in computer science, such as Frank van Harmelen and Michel Dumontier will be frequently consulted over the course of the project. The Chief Technical Officer who will be hired for the project will have a master's degree in computer science as a minimum requirement.

2.3.3.4   FAIR data policy

Stimulating FAIR data within its community is one of ODISSEI's fundamental aims. All ODISSEI project proposals must include a data management plan that lays out the documentation and archiving strategy of the project. Data stewards within the ODISSEI Hub will help researchers in developing appropriate data management plans. Data that are subject to ODISSEI's user agreement terms include Statistics Netherlands microdata, survey data, data collected via online experiments, sensors, apps, and biomedical data.

The **user agreement** terms will be structured around the **FAIR** principles and require that data and metadata are properly documented, archived, and accessible to the wider scientific community. For instance, one requirement is that all data collected within ODISSEI are documented using the CESSDA Metadata Model[35], based on the Data Documentation Initiative (DDI)[36] or pre-approved, industry-specific equivalent standard. The data must be archived in a sustainable and trusted repository and be searchable in the ODISSEI Portal. The documentation must include a clear description of the access protocols and whether the data are generally considered non-sensitive or sensitive.

Generated data must be made accessible to researchers at ODISSEI member organisations 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. All metadata must be made open and searchable. Where possible, ODISSEI data should be available to ODISSEI users at no cost. By establishing and enforcing the user agreement, ODISSEI will establish new standards of data access and interoperability upon which the social science community in the Netherlands can build. A team of Data Stewards in collaboration with experts from DANS will help researchers **make their data FAIR**.

---

35 www.zenodo.org/record/1118374#.XO7TcIgzY2w
36 www.ddialliance.org/

2.3.3.5   Safe Data

Making data FAIR ensures that researchers can find data and readily discover the access procedures and links to other data sources and scientific outputs. However, given that ODISSEI data tend to involve individuals, FAIR data must be balanced with safe data. Data within ODISSEI are grouped into two categories: non-sensitive and sensitive data. Non-sensitive data have been depersonalised and can be made accessible and distributed to the research community. Sensitive data are data that include an appreciable risk of re-identification and can therefore only be analysed within a secure and controlled environment. ODISSEI is not a legal entity and will not process or hold data[37] and it is therefore the responsibility of the relevant data controller to determine whether their data are sensitive data. ODISSEI employs the **five safes principle**[38] across the infrastructure to handle data that are deemed sensitive:

- **Safe Projects**: All projects that have been granted use of the ODISSEI Secure Supercomputer at SURFsara will be evaluated for security and ethical risks by the relevant data controllers[39] and the ODISSEI Data Officer (see section 2.3.3.6). All must approve a project before analyses can proceed.
- **Safe People**: All researchers analysing sensitive data must be employees of an ODISSEI member organisation and adhere to the ODISSEI user agreement (see section 2.3.3.4) and abide by the VSNU code of conduct[40]. SURFsara has considerable expertise in implementing Identification, Authentication, and Authorisation (IAA) systems and will advise on an access procedure for the Secure Supercomputer that is effective and scalable.
- **Safe Data**: The data that are brought into the ODISSEI Secure Supercomputer undergo data minimisation by the respective data controller who removes variables that are not of interest and pseudo-anonymises the data in order to meet only the immediate research needs of the scientists. SURFsara is a Trusted Third Party that links sensitive data using persistent identifiers (see section 2.1.1.7).
- **Safe Settings**: SURFsara meets strict international standards for information security and holds an ISO 27001 certification. The ODISSEI Secure Supercomputer was designed to obey all legal and technical safety requirements imposed by the strict Statistics Netherlands law, and passed a thorough penetration test by Secura in December 2018, demonstrating the system's security. Data in this facility does not leave the Netherlands. Taken together, the ODISSEI Secure Supercomputer facilitates ODISSEI users in the strict adherence to Article 89(1) of the General Data Protection Regulation (GDPR).
- **Safe Outputs**: All outputs from the ODISSEI Secure Supercomputer are checked manually by the data controller for disclosure risks under the four-eyes principle[41].

2.3.3.6   Ethical data use

All legal and ethical considerations will be overseen by the **ODISSEI Data Officer**. The Data Officer will be Marlon Domingus, the current Data Protection Officer at EUR and a member of the ODISSEI Coordination Team. He reports directly to the Management Board. The Data Officer will evaluate the ODISSEI 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 developed by DANS. The Data Officer will also check that ethical review boards of ODISSEI member organisations watch over activities conducted by partners and advises these boards where the specific nature of ODISSEI data or facilities makes ethical reviews complex or atypical. The ODISSEI Data Officer also evaluates proposals in response to calls for ethical considerations and ensures the highest ethical standards are applied. The ODISSEI Data Officer is supported in his work by the Data Stewards from the Coordination Team.

Across ODISSEI, 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.

Data that are collected in ODISSEI will be minimised to only that which is necessary and proportionate to the research question at hand. All data collected in ODISSEI will be pseudonymised, securely stored, and subject to

---

[37] ODISSEI data collections listed on page 18 remain the legal owners of all their data with full control over access.
[38] www.fivesafes.org
[39] Hosted and financed by ODISSEI member organisations as hosts of an ODISSEI data collection
[40] www.vsnu.nl/files/documents/Netherlands%20Code%20of%20Conduct%20for%20Research%20Integrity%202018.pdf
[41] www.unido.org/overview/member-states/change-management/faq/what-four-eyes-principle

the principle of privacy-by-design. This states that where the possibility to enhance the level of data protection is afforded to data subjects, ODISSEI will apply such measures by default rather than just considering them or making them available as an optional extra.

2.4 Technical & business case

2.4.1 Technical feasibility

ODISSEI and the products and services offered through ODISSEI partner organisations will be further developed through a continuous and iterative process. The challenges set out below and the resulting solutions and work plan illustrate an infrastructure in development in line with best practices of ESFRI and the OECD. Services and facilities will not be launched in a single release but result from continuous development and improvement over the five-year period based on user feedback of earlier versions. Identifying the expertise needed and defining more precisely the role of ODISSEI partners in different steps and stages will also be an ongoing process. Consequently, descriptions of tasks and timing are indicative, and milestones in which the work plan and strategic direction are evaluated might potentially be refined. Associated with this is the scalability of the ODISSEI infrastructure. Over time, a substantial increase is anticipated in the number of member organisations, the number of available datasets, the variety of datasets, the ways they may be interlinked, the computational demands, etc. Flexibility of processes and facilities towards such expansions whilst maintaining the highest levels of security is a crucial requirement.

The technical implementation of ODISSEI is managed by four Work streams: Data Facility, Observatory, Laboratory, and the Hub. These Work streams consist of 3 tasks each, 12 tasks in total. These 12 tasks consist of 36 subtasks spread over the five years. The next section, following Figure 5, describes the 36 subtasks and how they will be addressed. The program of work is explained in section 2.4.6, followed by the risk analysis (section 2.4.7), financial feasibility (section 2.4.8), and roles and responsibilities (section 2.4.9).

Task	Leader	Subtask	Responsible	
Data Facility				
Widening Microdata Access	Stat. NL	1.1b	Integrate tools into the RA environment	Stat. NL
		1.1c	Provide data stewardship	Stat. NL
Secure Supercomputer Scaling	SURFsara	1.2a	Create a scalable OSSC	SURFsara
		1.2b	Diversify and increase OSSC accessibility	SURFsara
Developing the ODISSEI Portal	DANS	1.3a	Extend the dataset coverage	DANS
		1.3b	Extend and improve search functionality	VU
		1.3c	Manage data access policies	SURFsara
Observatory				
Optimising Social Data Collection	NIDI	2.1a	Develop a Cost Reduction Strategy	NIDI
		2.1b	Develop protocols for online surveys	EUR
		2.1c	Coordinate community for data collections	NIDI
Data Scouts	Uni Leiden	2.2a	Scout for data	NIDI
		2.2b	Integrate historical data	IISG
		2.2c	Incorporate biomedical studies	VU
International Survey Integration	NIDI	2.3a	Link SHARE to administrative data	EUR
		2.3b	Create a production pathway for innovations	Uni Leiden
		2.3c	Assess the data landscape	Uni Leiden
Laboratory				
Operational Innovations	VU	3.1a	Explore distributed analytics techniques	Uni Maastricht
		3.1b	Develop automated data linkage	VU
Software for New Data Paradigms	UvA	3.2a	Develop the Media Content Analysis Lab	UvA
		3.2b	Pilot the Mass Online Experiment Lab	Uni Utrecht
		3.2c	Pilot the Citizen Science Suite	Uni Utrecht

Innovative Survey Experiments	CentERdata	3.3b	Refresh the LISS Panel	CentERdata
		3.3c	Improve access to the LISS Panel	CentERdata
Hub				
Community Management	EUR	4.1a	Administer the project	EUR
		4.1b	Develop an ODISSEI community	EUR
		4.1c	Create FAIR data	DANS
Educational Programme	EUR	4.2a	Organise ODISSEI conference	EUR
		4.2b	Organise ODISSEI workshops	EUR
		4.2c	Coordinate teaching programmes	EUR
Stimulating Social Analytics	Uni Utrecht	4.3a	Manage the Social analytics team	Uni Utrecht
		4.3b	Develop social science benchmarks	NLeSC
		4.3c	Administer grants for eScience support	NLeSC

Figure 5 – The Work streams, tasks, subtasks, and who takes ownership

2.4.2 Work stream 1 - The ODISSEI Data Facility

The ODISSEI Data Facility is a cluster of systems to strengthen the findability, accessibility, interoperability, and reusability of research data in the Netherlands.

Task 1.1 Widening and upgrading the Microdata Access Program [Lead: Statistics Netherlands]

The use of the Statistics Netherlands secure Microdata Services Facility (as described in section 2.4.9.3) by researchers from the social sciences has been growing rapidly and is anticipated to do so over the coming years within the context of ODISSEI. Statistics Netherlands' Microdata Services Facility, started more than fifteen years ago as a small-scale service to selected researchers, has now become a standard facility for an extended research community. For that reason, the facility has been included in the Dutch national map of research infrastructures[42]. The Microdata Services Facility has grown such that it requires the direction and investment of the research community to fulfil its potential. For example, requirements for regular (i.e. non-supercomputer) computation and associated services are increasing to a level that can no longer be supplied independently by Statistics Netherlands. The research community requires services and infrastructure beyond the remit of Statistics Netherlands. To address this, ODISSEI will create a more comprehensive and sustainable basis for linkage of administrative microdata, at the levels of IT infrastructure, tooling, and data stewardship. Statistics Netherlands has been open, collaborative and highly supportive of increasing this capacity, as illustrated by their long-term commitments to ODISSEI.

First, ODISSEI will *integrate Statistics Netherlands tools, both existing and to be further developed, into the ODISSEI Portal* (subtask 1.1b). The correct, efficient, and fast linking of large microdata sets is not easy for every researcher. This phase of the research cycle would therefore benefit strongly from flexible, easy-to-handle tools. In Statistics Netherlands internally, the SSD (System of Social Statistical Datasets) includes sets of tools for the flexible selection and linkage of populations and variables by Statistics Netherlands employees. The proposed project will explore and test whether this facility can be used in the Microdata Services Facility for external users, as it is.

In addition to these tools and resources, ODISSEI will provide *additional data stewardship* (subtask 1.1c) via experts at Statistics Netherlands. The data stewardship will be provided to projects selected through open calls via the Microdata Access Program which will be administered by the Coordination Team at EUR, for both the Statistics Netherlands Microdata Services Facility and ODISSEI Secure Supercomputer. Due to the growing complexity of both research data and the various analytical tools, leaving it to researchers to find their way through the administrative data catalogue and range of available compute options is becoming increasingly inefficient. Support from substantive data experts, data analysts, and technical consultants is required. Data stewards will be involved in the support of individual research projects but will also be available as a help desk.

---

[42] www.onderzoeksfaciliteiten.nl/facility/microdata-services

Across this task, the team at Statistics Netherlands will be complemented with approximately 1,5 person-year of data stewards (€ 138,000). Material costs are dedicated to Microdata calls (€ 1,750,000). [Total € 1,888,000].

Task 1.2 ODISSEI Secure Supercomputer Scaling [Lead: SURFsara]

The current ODISSEI Secure Supercomputer has shown its massive potential but has so far been used as a Trusted Third Party (TTP) for Statistics Netherlands data only and uses a Statistics Netherlands secure access connection and secure analysis environment. It is not easily scalable to other providers of sensitive data. It is therefore imperative that ODISSEI builds on the success of it by extending it towards *a scalable and generic secure high-performance computer* (subtask 1.2a – see Figure 6).

As part of ODISSEI, SURFsara will create a scalable secure data transfer environment to transfer privacy sensitive and large data to a supercomputer storage cluster. The connection will enable the analysis of data held by ODISSEI member organisations and other data sources, whilst enabling the data controllers at these organisations to remain in full control of the data throughout. Transferring data to the cluster will be supported by DANS through the ODISSEI Data Node (see subtask 1.3c). SURFsara will act as a Trusted Third Party by combining multiple sensitive datasets in a secure manner (see pilot project 2 in section 2.1.1.7). The researcher can then perform the analysis on the ODISSEI Secure Supercomputer. This environment will also easily scale to multiple use cases and will be able to handle the transfer of large amounts of data in a timely manner. After any necessary disclosure check by the data controller, the output data are released to the researcher.
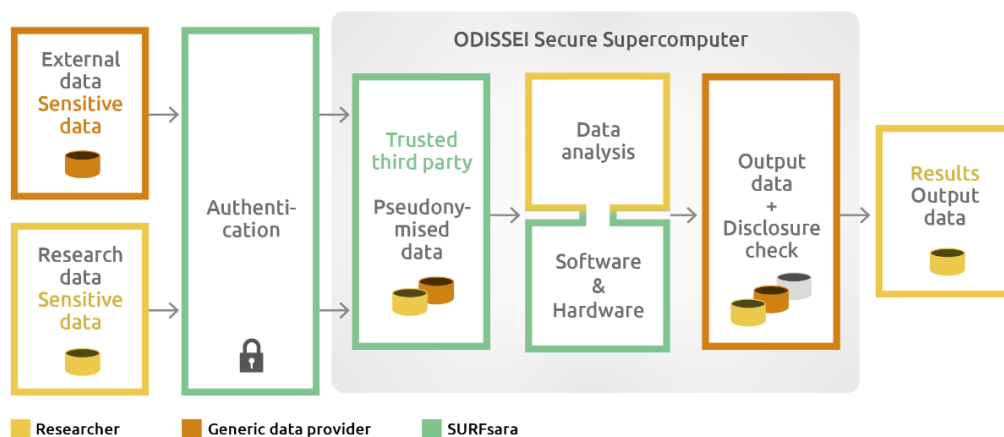

Figure 6 – Architecture of the ODISSEI Secure Supercomputer that has been improved for generic use

In the pilot phase, the emphasis was on providing typical high-performance workloads. Nevertheless, users tend to have more diverse needs. Some require a 'classic' supercomputer cluster for batch-like workloads, while others simply require a 'bigger' workstation for interactive work. SURFsara will undertake consultations within the ODISSEI community and *diversify and increase the accessibility of the compute facilities available via the ODISSEI Secure Supercomputer* (subtask 1.2b) in an iterative fashion through a gradually expanding set of open calls for new and increasingly diverse projects. SURFsara will offer access to different infrastructures, more cloud-based data-analytics tools such as RStudio and Jupyter Notebooks, and an intuitive interface to Python, R, and STATA for accessing and processing data. SURFsara will also ensure the further integration of storage systems, so that the storage on the supercomputer cluster and HPC Cloud VMs at SURFsara are unified from the user perspective.

Across this task, the team at SURFsara will consist approximately of a project coordinator and senior technicians (€ 1,176,000). In addition, a non-scientific employee at DANS will deliver FAIR support (€ 102,000) and Statistics Netherlands will provide a project manager and will work on the user interface (€ 256,000). Material costs involve an infrastructure upgrade (€ 225,000) and a penetration test (€ 20,000) [Total € 1,819,000].

Task 1.3 The ODISSEI Portal [Lead: VU]

The ODISSEI Portal combines metadata from a wide variety of research data repositories into a single interface, allows advanced semantic queries to support findability, and facilitates data access (Figure 7).

The ODISSEI Portal's metadata catalogue will *extend the coverage of available data* (subtask 1.3a) by including key research datasets that are currently not findable via NARCIS (the main national portal for information about researchers and their work). The project will extend the current catalogue with the metadata of (a) all datasets of Statistics Netherlands, including the metadata of the microdata catalogue, (b) all datasets developed in the ODISSEI Laboratory (LISS), and (c) all datasets developed in the ODISSEI Observatory (EVS, GGP, SHARE, ESS,

NTR, HSN). This dataset extension task is a joint effort of trained data stewards at DANS, the ODISSEI team of **Data Scouts** at the Observatory, and the aforementioned data repositories. In collaboration with these partners and experts at VU Amsterdam, this task will also develop a metadata ingestion pipeline to make sure that the Portal can be maintained and kept up to date. This pipeline will be used by the hosting party to add new datasets during and after the end of the project[43]. All new datasets that will become available via the ODISSEI Portal will also be added to the national NARCIS research dataset catalogue.

Currently, existing tools for findability in the social sciences are limited in that they only identify specific terms or synonyms (e.g. United Kingdom question bank or the Question Variable Database). The ODISSEI Portal will *extend and improve search functionality* (subtask 1.3b) by using semantic queries which will enable broader probabilistic matching and link functions over an enriched knowledge graph representation of the FAIR metadata catalogue. This incorporates the context of specific terms which are crucial in social research. By using rich and extensible data structures developed within the linked data community, ODISSEI will evolve the relatively flat metadata catalogues in use today into the highly interlinked and graph-based structures needed to conduct advanced semantic searches. The richness in the metadata catalogue does not solely rely on the high manual documentation and curation standards that already exist across ODISSEI associated data such as DDI. The social sciences are fortunate to have an advanced automated metadata capture system that documents the data collection process, principally through survey software[44]. Besides the manual documentation and curation standards, it takes advantage of automatic and semi-automatic metadata enrichment and entity linking to enrich the available curated information. Where relevant, there will be alignment with standards used in CESSDA (CESSDA Metadata Model) at the European level and the Open Science Framework[45].

The search functionality will be designed by information scientists at VU Amsterdam in close collaboration with Data Stewards at DANS and the ODISSEI Data Scouts. Lessons learned from the enriched data format and extended search and discovery functionality will naturally feed into the development of future versions of NARCIS to become also used outside the ODISSEI community.
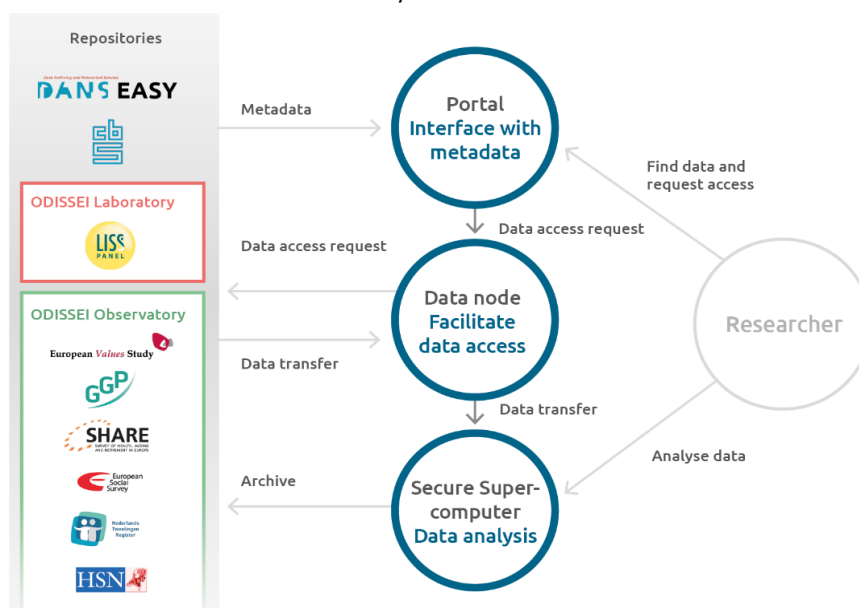

Figure 7 – The Portal, Data Node, Secure Supercomputer, together with the Microdata Facilities, form the ODISSEI Data Facility.

(3) The Portal also facilitates *automatic and semi-automatic data access policy management* (subtask 1.3c) between the producers and users of research datasets. Unclear data licensing or access policies are currently an obstacle in open science and the application of the FAIR principles, even for research datasets that are available as open data. ODISSEI will enrich its research data catalogue with explicit, and as detailed as possible information on licensing and access policies, preferably in a machine-readable format. The owners of each dataset will be able to provide the Portal with metadata describing what the policy for obtaining access entails. The access

---

[43] DANS has committed to the continued maintenance of this pipeline between 2025-2029
[44] For examples, see Colectica (www.colectica.com) or Blaise (www.blaise.com)
[45] www.osf.io

process varies between data providers: Statistics Netherlands requests that the user is affiliated with an authorised research institute and using their data involves formalities and costs, whereas other research data are often freely available for download to anyone around the globe. For datasets with machine-readable access policy metadata, the ODISSEI Data Node, an automated system that is closely connected to the Portal, will be able to facilitate the researcher, for example by sending data access request to the data owner, by initiating a federated authentication session, or by redirecting researchers to the landing page of the open dataset. In case a dataset does not yet have fully machine-readable access policy metadata, the ODISSEI Data Steward based at EUR will help the data owner and researcher with the access process.

Once the data owner reaches an agreement with the researchers, the owner allows the ODISSEI Data Node to transfer the data to the designated analysis environment, typically the ODISSEI Secure Supercomputer (in case of large, complex or sensitive data) or the computer of the researcher (in case of small and/or open data).

The Data Node will be designed and prototyped by SURFsara (secure authentication and link to the Secure Supercomputer), DANS (owner of NARCIS), and VU Amsterdam (linked data expertise). Statistics Netherlands will make its metadata available and provide expertise on the secure data transfer connection. Design and development will happen within the first two years of the project by information scientists at VU Amsterdam and data stewards at EUR, DANS and Statistics Netherlands. DANS will then operate the Portal/Data Node.

Researchers who have created or altered data, will be encouraged to properly store them according to the ODISSEI user agreement, with the help of the FAIR support team (see the Hub).

Across this task, the team at VU Amsterdam will consist of a full-time senior scientist and a PhD student in information sciences to build a proof of concept of the Portal (€ 409,000). They will be supported by a team of four non-scientific employees at DANS and SURFsara including software developers and data stewards (€ 700,000). There also are licensing costs (€ 16,000) [Total € 1,125,000].

2.4.3   Work stream 2 - The ODISSEI Observatory

The ODISSEI Observatory provides continuity for long-standing, excellent data collections in the social sciences and links them to the ODISSEI infrastructure.

Task 2.1 Managing Change in Social Science Data Collections [Lead: NIDI]

ODISSEI will address high fieldwork costs in the Netherlands by *developing a cost reduction strategy for fieldwork* (subtask 2.1a) to identify potential efficiencies throughout the data collection process in survey infrastructures. ODISSEI will commission a strategic assessment of the face-to-face fieldwork process and systematically identify the cost components within the current tender requirements. Specifically, workload management and caseload constraints that data collections apply within tender requirements[46] are preventing fieldwork agencies from deploying more cost-efficient face to-face-operations. Sometimes these requirements have high scientific merit, but some have limited scientific merit and significant consequences for cost. The results of the assessment will then be fed into a working group which consists of leading data collections. This working group will be tasked with developing guidelines for future data collections and facilitating the cost reduction strategies outlined in the assessment. Based on the report and its findings, ODISSEI will provide a single tendering framework with clear guidelines on how call for tenders are to be structured. This will be used for data collections from January 2022 onward.

Whilst face-to-face fieldwork costs are increasing rapidly, the feasibility of high-quality online surveys are improving. Online surveys have been constrained by the degree to which they reach the whole population and the degree to which they can be applied in international contexts. ODISSEI will address these issues by *developing protocols for adopting web only designs* (subtask 2.1b). These will specify mitigation strategies for the negative effects of online only data collection, specifically for hard to reach sub-populations and the internationally comparable nature of the data. The next stage in development will be to set out the conditions and timetable for this transition. These will then be used to inform and structure future data investments within ODISSEI.

As well as financing and monitoring data collection, ODISSEI will help *develop and strengthen the existing community of data collections in the Netherlands* (subtask 2.1c). These activities will be overseen by survey

---

[46] Data collections generally set specific guidelines such as a minimum number of contact attempts and place restrictions on agencies overusing specific interviewers or targeting the respondents who are easiest to reach.

methodologists based at NIDI and will also include resources to support sustainable fieldwork in ODISSEI data collections.

The work across this task will be coordinated by a senior scientist at NIDI (€ 145,000) and supported by a non-scientific community manager at EUR (€ 170,000). Material costs are for the European Social Survey (€ 580,000), the assessment (€ 20,000), and for the NKO (€ 350,000) [Total € 1,265,000].

Task 2.2 Linking to new forms of data [Lead: UL]

One of the aims of the ODISSEI Observatory is to broaden the types of data that are integrated into ODISSEI in order to maximize its scientific potential and scope. The Observatory will seek to integrate data from three new sources and open up new lines of research. First, NIDI will *employ a Data Scout* (subtask 2.2a) which will focus on integrating data into the ODISSEI Portal from non-research sources. These will include private companies, municipalities, NGO's, and civil society organisations that have data relevant for research purposes. These data providers will be brought together through a series of engagement events in which the current and future potential of ODISSEI will be illustrated. ODISSEI will allow for data donations from such organisations whose metadata and access protocols will be then transferred to the ODISSEI Portal. The data donors would not be allowed to access the Microdata Access Program or ODISSEI Secure Supercomputer themselves but would instead rely on researchers based at ODISSEI member organisations to conduct analysis. However, data donors would be able to stipulate that any output which includes their data, be shared with them and that access to their data be restricted in scope. This initiative will greatly diversify the data within ODISSEI and lay the groundwork for future industry interactions and a powerful model of societal impact. *This work will be overseen by a senior scientist at NIDI [Total € 145,000].*

Second, *data from the Historical Sample of the Netherlands (HSN) will be linked with the current Statistics Netherlands microdata and catalogued in the ODISSEI Portal* (subtask 2.2b). This will enable links between the historical research conducted using the HSN and contemporary society and the outcomes studied by social scientists, enabling stronger links with CLARIAH. The HSN sample is based on the birth certificates from the period 1812-1922. To achieve the link with Statistics Netherlands microdata, all HSN research persons from the birth cohorts 1900-1922 (n=20,000) will be extended with the life courses of their children. Most of these children are born between 1925 and 1960. In total, ODISSEI will link at least four generations of families that lived in the Netherlands between 1850 and 2023. The linking will be done on the basis of the combination of the birth date of the person and his/her parents. A test proved that 85 to 90% of the persons from the second generation could be immediately linked to the Statistics Netherlands microdata. However, more extensive linkage procedures are necessary to resolve ambiguities (twins, double links) and errors including more variables in the procedure (e.g. names and partners). The linking of this data will create a meaningful infrastructural bridge between ODISSEI and CLARIAH and the resulting links will be findable via the ODISSEI Portal. *This work will be overseen at HSN by a senior scientist for three years (€ 236,000).*

Third, data from ODISSEI data collections will be linked in the ODISSEI Secure Supercomputer (subtask 2.2c). Many data collections in the social sciences have begun to include genotype information or other measures such as MRI and recordings from wearables. These can be pooled to achieve larger sample sizes to study rare conditions or small and hard to reach sub-populations. To accommodate and link multiple sensitive data sources together, the ODISSEI Secure Supercomputer is required. The NTR will take the lead in linking the ODISSEI Secure Supercomputer project to a diverse range of ODISSEI data collections. Running new and innovative genetic analyses on multiple traits, longitudinal outcomes, and gene-environment interactions will allow insights into the role of genetics in social phenomenon within Dutch society. The aim will be to set up standard operating procedures (SOP) for running parallel analyses within multiple ODISSEI data collections with their data linked to other sensitive data sources such as Administrative Data from Statistics Netherlands. Because of privacy, informed consent, and linkage issues, combining data from different data collections must be done in the secure access facility that ODISSEI provides. Combining data across data collections will substantially enhance statistical power, which is essential in many fields of research. *A dedicated NTR postdoc will work closely with SURFsara for three years [Total € 236,000] to link the data.*

Task 2.3 International Survey Integration [Lead: UL]

The coordination of all ODISSEI data collections by a single data infrastructure opens the possibilities for greater substantive and methodological synergies, the coordinated application of new technologies, the implementation of shared standards, and the efficient distribution of resources. This can reduce costs, improve data quality and support multi-disciplinarily. This is particularly true with regards to international data infrastructures such as

SHARE, GGP, ESS, LIS, and EVS. ODISSEI will coordinate the Dutch representation with these infrastructures and seek to align the work within ODISSEI. This work is represented by three separate subtasks. Firstly, ODISSEI will *link the survey data collected from SHARE-NL waves in 2020 and 2022 with administrative data stored at Statistics Netherlands* (subtask 2.3a). The legal framework for this was recently established and preliminary linking of the earlier waves of SHARE work was completed in January 2019. The linkages made between the survey data and administrative records would then be analysed for methodological and substantive research and demonstrate the added value of such data integration. In addition, collaborations with other SHARE national teams conducting data linkage will be sought with a specific view to conducting comparative administrative data linkage with SHARE. This would require a secure means of analysing linked administrative data held in multiple countries. This work will examine whether the ODISSEI Secure Supercomputer or distributed computation methods could be used for this, potentially opening up ODISSEI for international collaboration and use.

The second way in which ODISSEI will align national and international data infrastructures is by creating a *production pathway between the ODISSEI Laboratory and international survey infrastructures* (subtask 2.3b). The ODISSEI Laboratory, including the LISS Panel at CentERdata, is already an internationally recognised centre of excellence in survey infrastructure. However, the tools and services offered need to be integrated into survey infrastructures at an international level in order that the innovations can have the broadest possible impact. These tools generate income for Dutch research centres and strengthen the Netherlands as a hub of social science methodological and technical excellence. To support this, the ODISSEI Observatory will publish production-ready services and tools from the Laboratory on the European Open Science Cloud and implement them within Dutch based international survey infrastructures.

ODISSEI will also conduct a *full assessment of international data commitments* (subtask 2.3c) and evaluate participation on a biannual basis (2021 & 2023). This process is currently disjointed and inefficient. Statistics Netherlands is mandated to provide certain data to Eurostat under European regulation and there are also statistical requirements under the United Nations Agencies such as the International Labour Organization and the United Nations Population Division. There are also requirements as part of membership of the OECD and participation in studies such as PISA (Programme for International Student Assessment) and PIAAC (Programme for the International Assessment of Adult Competencies). Then there are commitments of the research community to international research infrastructures such as SHARE-ERIC, ESS-ERIC, GGP, EuroCohort, LIS, and EVS. ODISSEI will develop an evaluation framework for the sustainability of the data, examine the data landscape, and report to the ODISSEI Supervisory Board on its recommendations for data coordination.

The work across this task will be coordinated by a senior scientist at Leiden University (€ 155,000) and supported by a non-scientific data steward manager at EUR (€ 340,000). The material costs are for LIS (€ 90,000), SHARE (€ 600,000) and GGP (€ 350,000) [Total € 1,526,000].

2.4.4   Work stream 3 - The ODISSEI Laboratory

The ODISSEI Laboratory develops the innovations that will shape the infrastructure for the future and ensures that new forms of data can be analysed by researchers.

Task 3.1 Operational Innovations in ODISSEI [Lead: VU]

Techniques in analysing data are advancing rapidly, yet many are not yet mature. Furthermore, they have either not been implemented in the social sciences or are not yet suitable for the wider social science community. ODISSEI will spend energy exploring these developments and work to ensure that the social sciences can apply future analytical approaches which better capture the complexity and multi-dimensional nature of social phenomena. To do this, innovations will be examined, adapted, and developed within the Laboratory before they are put into wider operation within ODISSEI at a later date. There are two specific areas in which ODISSEI will seek to examine the potential for application in the social sciences.

The first area is the use of distributed computation through federated learning algorithms. This has been applied in the Personal Health Train[47] which is a concept that was developed in the medical sciences. In the *Verantwoordelijke Waardecreatie met Big Data* project, funded by the NWA (project number: 400.17.605), Maastricht University together with The Maastricht Study and Statistics Netherlands are implementing the Personal Health Train on health data. Its basic concept is that several stakeholders want to collaborate in data analysis, but do not want or are legally not allowed to share the data with one another – often due to privacy

---

[47] www.dtls.nl/fair-data/personal-health-train/

considerations. Rather than appointing a Trusted Third Party, the researchers develop an algorithm (in this analogy: the train) that visits each stakeholder ('station'), analyses the data on site and goes to the next station with the analysis results, but without the data. The concept of distributed computation has been widely studied in the medical and computer sciences (Sun et al., 2018). Though having great potential in the social sciences, to date, no research has been done about its applications in this field. ODISSEI will *examine the possibilities for use of distributed analytics techniques in the social sciences* (subtask 3.1a). For example, educational data which is highly federated and stored securely in hundreds of individual schools might be analysed without the need to collate and centralize data. The use of such an approach is not without risks and extensive testing and prototyping will be required. Other ODISSEI data collections with access and technical restrictions within the ODISSEI data facility architecture will also be identified and the potential of distributed computing will be examined along with potential data providers in collaboration with the Data Scout team in the Observatory (see task 2.2). *This work will be conducted by a two-year senior scientist at Maastricht University [Total € 155,000].*

A promising, though high-risk area, in the field of computer science is *automated data linkage* (subtask 3.1b). This has the potential of strongly reducing the time needed by researchers to manually link datasets while performing analysis. Linking datasets is one of ODISSEI's core aims. The social data within ODISSEI is highly structured and persistent identifiers and standardised codes are used pervasively. However, the degree of automated linkage is restricted by the lack of infrastructure that exists for such linkage. Computer science techniques are advancing through the use of deep learning methods which help identify and link more opaque constructs within data that are less well defined and structured such as families, social networks, neighbourhoods or even cultural groups. The potential of such techniques in the social sciences is considerable but these approaches require high levels of expertise. In this subtask, these high risk/high reward approaches will be explored and examined to see if they can complement the more functional and established manual and semi-automated linkages which will be made in the development of the ODISSEI Portal. Work will start with the construction of a knowledge graph of ODISSEI data collections and data sources in order to assess the scope and potential for automated linkage. *This work will be conducted by a senior scientist at VU Amsterdam [Total € 318,000].*

Task 3.2 New Software for New Data Paradigms [Lead: UvA]

ODISSEI will support and harness innovations in collecting new forms of data through three tasks which assist researchers in the collection and analysis of such data. They will be piloted, developed, and become operational using an agile model of software development with coordination, cooperation, and shared expertise across the three tasks.

The ODISSEI Media Content Analysis Lab (subtask 3.2a) will offer systematic analysis of large corpora of digital media content. The Media Content Analysis Lab tackles the major challenge of digital text analysis whereby copyright and GDPR restrictions make it very hard to share media content data by providing a workbench for researchers where they can share data and analyses with strict rights management, without the need to read or export the data themselves. Currently, several Dutch initiatives exist that facilitate retrieval, storage, and analysis of media content (i.e. the infrastructure for content analysis [INCA] at the University of Amsterdam (Trilling et al., 2018), and the Amsterdam Content Analysis Toolkit [AMCaT] at VU Amsterdam (Van Atteveldt et al., 2014). Their use, however, requires considerable programming skills and is currently tailored to a limited set of applications. Through the Media Content Analysis Lab, the large amount of longitudinal media content data (i.e. online news data; social media data) currently available within ASCoR at University of Amsterdam will be made available to researchers at ODISSEI member organisations. It will also be possible to deposit new data, or to use specific tools on data stored elsewhere. First, an initial set of pre-existing tools will be identified - ranging from scrapers to collect content from specific sources, to the extraction of meta information to scripts to conduct topic modelling, sentiment analysis, various forms of machine learning, deep learning (for a concise overview see Boumans & Trilling, 2018). Next, suitable existing textual corpora will be added to the database. An interface will be developed that offers researchers the opportunity to register as users, and to apply existing tools on existing datasets. Moreover, the opportunity to apply the content analytical tools on one's own datasets will be added. Researchers from different social science disciplines will be actively invited and encouraged to do so and will receive technical support when needed. *The Media Content Analysis Lab will hire a four-year senior scientist specializing to develop the system [Total € 270,000].*

The second piece of new data infrastructure will be *the ODISSEI Mass Experiment Online Lab* (subtask 3.2b). The Mass Experiment Online Lab will facilitate experiments in which large numbers of subjects simultaneously interact under controlled conditions. These population-level experiments cannot be conducted in the traditional laboratory

as they require scale in which network structure is systematically varied across multiple large-sized experimental populations (Centola, 2007; Bail, Merhout & Ding, 2018). The Mass Experiment Online Lab overcomes (1) the significant infrastructural and logistic challenges associated with simultaneous networked participation of many subjects (Salganik, 2018) and (2) removes barriers-to-entry by providing methodological and organisational research facilities to interested but otherwise ill-equipped domain experts through a series of open calls. The Lab will be piloted iteratively using an agile model, with increasing functionality to service a gradually expanding community of beta users. Data generated will be archived, findable via the ODISSEI Portal and linkable within the ODISSEI Secure Supercomputer. *The Mass Experiments Online Lab will be piloted by a one-year non-scientific programmer [Total € 77,000].*

The final piece of new data infrastructure will be the *ODISSEI Citizen Science Platform* (subtask 3.2c) which will improve data quality in citizen science by applying existing expertise from the field of social science methodology. Citizen science projects include those which rely on ordinary citizens to collect scientific data at a large scale on for example air pollution[48], the backyard bird count[49], or the history of marriage[50]. Underestimated aspects of data collection through citizen science are issues of selectivity and measurement error. For example, what type of citizen participates in science, and does that affect the conclusions? How reliable are the measures collected, and is it possible to estimate the measurement errors and correct for any detrimental effect they might have? How does one make sure that member organisations collect data in such a way as to minimise these errors? Social methodologists are used to developing solutions to such questions for social data collection at the service of social science. ODISSEI will pilot a platform for the benefit of all fields of science that use citizens as data collectors. This platform will perform three functions: (1) trusted and convenient data collection interface for fast development of Citizen Science applications through a web-based and mobile app platform to facilitate data collection; (2) link to spatial and demographic information from Statistics Netherlands to allow post-stratification to adjust for selectivity of the citizen scientist population and investigate the sensitivity of the conclusions; (3) allow double-coding and validation so researchers can estimate and correct for classification errors. The platform will be built open-source and will be modular so that teams can easily add components to the study or change the look-and-feel. Existing components will be recycled as much as possible. Over the project, the ODISSEI Citizen Science Platform will collaborate with existing citizen science projects and invite groups to use the tools and receive support through open calls. *The Citizen Science platform will hire a one-year non-scientific programmer to pilot it [Total € 77,000].*

Task 3.3 Innovative Survey Experiments [Lead: CentERdata]

ODISSEI will strengthen and develop the LISS panel in several ways. Firstly, ODISSEI will provide an upgrade of the existing LISS panel through a *panel refreshment* (subtask 3.3b) of a further 1,000 households in the first, third, and fifth year of the project. This will increase the representativity of LISS and capture subpopulations that the current panel is not adequately resourced to capture. Secondly, ODISSEI will provide extensive *space on the LISS panel for researchers* (subtask 3.3c) at ODISSEI member organisations for the deployment of their own research designs and experiments. This recently launched initiative is a hugely successful and inspiring element of ODISSEI. Researchers of any seniority can design their own survey research that is then deployed by the LISS technical team at CentERdata at no cost. These research projects are selected through an open call administered by the ODISSEI Coordination Team. Within the project, ODISSEI will grant space on LISS to between five to ten projects per year via open calls and will evaluate projects on the degree to which they integrate with and enhance the wider ODISSEI infrastructure through data linkage and enrichment. All collected data will be available to the entire academic community at no cost.

The work in this task for CentERdata includes a panel refreshment (€ 1,275,000) and LISS calls (€ 475,000) [Total € 1,750,000].

---

[48] www.samenmetenaanluchtkwaliteit.nl/zelf-meten
[49] www.tuinvogeltelling.nl/
[50] www.collective-action.info/ja-ik-wil

2.4.5   Work stream 4 - The ODISSEI Hub

The ODISSEI Hub develops a community and supports the usage and usability of the infrastructure through an educational program, community events and as analytic support.

Task 4.1 Community Management [Lead: EUR]

Community management is aimed at building and maintaining a sustainable ODISSEI collectivity. It is led by the Coordination Team based at EUR who have several responsibilities within ODISSEI to ensure smooth operations across the diverse range of activities. The first subtask will be the *general administration of the project* (subtask 4.1a) including financial and technical reporting, the administration of calls and the monitoring of progress. Being a community infrastructure, it is vital that ODISSEI actively monitors researchers' needs as input for future directions and communicates what ODISSEI offers. Community Officers will engage in various *community development* (subtask 4.1b) activities through regular face-to-face meetings as well as mailings, the website, social media, and the aforementioned ODISSEI annual conference. Community Officers are the first point of contact for researchers from member organisations turning to ODISSEI. Their job is to help researchers find their way in the infrastructure. Finally, Data Stewards will help researchers *create FAIR data and execute ethical practices* (subtask 4.1c). Experts from DANS will develop and maintain the user agreement, which focus on data management, data sharing, research ethics, and the principles and practice of FAIR data (also see section 2.3.3.4-2.3.3.6). ODISSEI Data Stewards will provide support to researchers to implement the user agreement, in collaboration with the experts at DANS. *Work across this task will be conducted by the Coordination Team at EUR which scales from three full-time non-scientific staff in the first year, up to five staff in the final year of the project. This includes the Managing Director, the Chief Technical Officer, the External Relations Officer, and community officers and a data steward who also conducts work within the Observatory (a further 1 FTE) (€ 1,126,000). Material costs are for communications and staff training (€ 150,000) [Total € 1,276,000].*

Task 4.2 Educational Programme [Lead: EUR]

The ODISSEI Hub's External Relations Officer will operate an extensive educational programme to facilitate the use of the infrastructure and the uptake of computational and innovative methods in the social sciences. First, there will be an *annual ODISSEI Conference* (subtask 4.2a) where researchers can showcase their projects and engage in broad and open discussion about developments in the infrastructure. This annual event will be hosted by a different member organisation every year and will be coordinated by EUR. Second, there will be a *series of workshops* (subtask 4.2b) aimed at promoting the usage of the infrastructure and the methods and skills needed to do so. These workshops will be coordinated by the EUR but will be hosted by various member organisations. There will be approximately six workshops per year. The program of workshops includes but is not limited to: The ODISSEI Data Facility users Bootcamp; An introduction to Statistics Netherlands microdata access; An introduction to the LISS Panel; Ethics and GDPR in ODISSEI; Machine learning for social science;. Third, there will be regular communications with the educational officers at member organisations to *synchronise and coordinate existing educational programs* (subtask 4.2c) on computational social sciences.

Work across this task will be coordinated by a non-scientific education officer based at EUR (€ 170,000). Material costs for communication are € 175,000 [Total € 345,000].

Task 4.3 Stimulating state-of-the-art social analytics [Lead: UU]

The ODISSEI Hub will actively work with social scientists across the country in employing state-of-the-art techniques to leverage ODISSEI infrastructure for new discoveries in the social sciences.

First, the *Social Analytics team* (subtask 4.3a) will bridge the gap between applied social scientists and the data, computing, and analytical infrastructure provided by ODISSEI. They will actively seek out short-term collaborative projects with social scientists across ODISSEI member organisations to uniquely leverage ODISSEI's infrastructure to answer novel substantive questions and to build expertise in computational social science. These projects will result not only in 'traditional' publications, but also in open source reusable code that can be used for teaching purposes or as a starting point for other projects. *The team will be coordinated by Utrecht University and trained by NLeSC. Over the project, the team will be scaled up to consist of three computational social scientists (senior scientific employees) with a background in social science methodology and strong skills in coding, modern data analytics, supercomputer usage, and machine learning [Total € 791,000].*

Second, the ODISSEI Hub will pilot *workshops and hackathons on computational benchmarking in the social sciences* (subtask 4.3b). In computational sciences, researchers often have to choose between different computational approaches. Benchmarking uses standard datasets and parameters to systematically compare

computational approaches and establish best practices in the community. Since translating research problems into benchmark studies is a novel concept in social sciences, workshops will be organised on their design and to build consensus on the set-up for a specific task. This will eventually allow ODISSEI to monitor the state-of-the-art algorithms for a specific task on the EYRA (Enhance Your Research Alliance) Benchmark Platform[51] developed by NLeSC and SURFsara and will be a community effort. Donoho (2017) ascribed a large part of the astounding recent successes in machine learning to the application of the 'common task framework' (CTF), in which (1) a training dataset derived from ODISSEI infrastructure is publicly shared, (2) teams of experts compete for a common task pertaining to prediction on the training data, and (3) an impartial referee reports a performance score for each team based on held-out testing data. Fields in which this competition framework has been systematically applied, such as automated driving, machine translation, and medical imaging, have reported impressive and objectively quantifiable progress (Donoho, 2017). *A pilot of the platform will be developed by a one-year non-scientific programmer at NLeSC [Total € 80,000].*

Third, the ODISSEI Hub will provide *grants for computational social science* to support the use of the data infrastructure developed across the other three Work streams (subtask 4.3c). Following a successful practice employed at the NLeSC, these grants will provide hours from eScience Research Engineers employed at the NLeSC to collaborate with social scientists to enhance their research, by exploiting digital technology. The grants will be made available via an annual open call which will be assessed by experts in the field of computational social science. Applicants can apply for three different grant types, depending on the amount of eScience engineering hours required to achieve their goals: 1) Five small pathfinder grants (€ 20,000 = 336 engineering hours) will be available per year (in year 2-4). These pathfinders are pilot projects suitable for researchers just starting in the field to explore the use of digital technology for their research. 2) One medium grant (€ 50,000 = 840 engineering hours) per year will be available in year 3-4, which is suitable for researchers that passed the pilot phase and would like to extend their experiments. 3) One large grant (€ 100,000 = 1680 engineering hours) will be available in year 4 to consolidate previous technical advancements for social sciences. The application process will be lean and ODISSEI will allow for flexibility during execution of the project. All calls will be organised together with the NLeSC, who will provide specialised staffing when needed. The grants will be administered by the ODISSEI Hub team and evaluated by independent committees. *The budget is dedicated to the calls [Total € 550,000].*

2.4.6   Program of work

The four Work streams will consist of the subtasks identified above. ODISSEI subtasks will be developed in a continuous and iterative product development lifecycle that is designed and controlled by subtask leaders. This agile approach is widely used in highly complex projects and based on the principle that, in order to create products and services that end users want, they should be involved as early as the conceptual stage. To aid with reporting, management and coordination across these tasks and subtasks, a shared monitoring framework covering four simple phases has been developed: concept, prototype, beta, and mature. At the end of every phase, the deliverable is the product that has been tested by potential end users; the quality and completeness of the end result increases with every phase.

At the end of the **concept** phase of any subtask, a conceptual design must have been evaluated by potential end users and the design and specification report must be submitted to the task leader, Managing Director and Chief Technical Officer for approval. This will detail the products or services, their specifications and a development roadmap. During the **prototype** phase, products or services are tested on one or two use cases. At the end of prototyping, subtask leaders are required to report to the Managing Director and Chief Technical Officer on the functionality of the prototype in comparison to the conceptual design and include usability reporting from the research team within the use case. Based on this, the development roadmap will be adjusted, and the necessary budgetary and technical provisions agreed for beta testing. During the **beta** phase, the product or service will increase the number of use cases to 3-5 and diversify the nature of research being conducted and increase the functionality offered. During this phase, demonstrations of the service must be provided to the Management Board for approval. Based on this, the development roadmap will be adjusted, and the necessary budgetary and technical provisions agreed for the mature product or service. Any product or service that is **mature** must conduct an evaluation every year and report on usage, problems, planned improvements and a delivery plan for the coming year. Deviations from the program of work set out below must be submitted to the Managing Director

---

[51] www.eyrabenchmark.net

and approved by the Management Board. A sustainability plan must also be included which sets out how the product or service is to be sustained beyond the lifetime of the project.

It is the responsibility of the Managing Director to identify, manage and mediate the interdependencies between tasks, and implement contingencies when delays occur and mitigate spill-over effects. To facilitate this, task leaders are required to report progress to the Managing Director on a quarterly basis. Furthermore, it is the responsibility of the Managing Director to identify potential synergies between tasks. Where synergies or interdependencies occur, which could result in significant delays or require substantial changes to the resource allocation within the project, the Managing Director is responsible for bringing it to the attention of the Management Board at the next monthly meeting.

Task			Responsible	Year 1–5 (Quarters 1–20)
Data Facility				
Widening Microdata Access	1.1b	Integrate tools into the RA environment	Stat. NL	
	1.1c	Provide data stewardship	Stat. NL	
Secure Supercomputer Scaling	1.2a	Create a scalable OSSC	SURFsara	
	1.2b	Diversify and increase OSSC accessibility	SURFsara	
Developing the ODISSEI Portal	1.3a	Extend the dataset coverage	DANS	
	1.3b	Extend and improve search functionality	VU	
	1.3c	Manage Data access policy	SURFsara	
Observatory				
Optimising Data Collection	2.1a	Develop a cost reduction Strategy	NIDI	
	2.1b	Develop protocols for online surveys	EUR	
	2.1c	Coordinate community data collections	NIDI	
Data Scouts	2.2a	Scout data	NIDI	
	2.2b	Integrate historical data	IISG	
	2.2c	Incorporate biomedical studies	VU	
International Integration	2.3a	Link SHARE to administrative data	EUR	
	2.3b	Create production pathway innovation	Uni Leiden	
	2.3c	Assess the landscape	Uni Leiden	
Laboratory				
Operational Innovations	3.1a	Explore distributed analytics techniques	Uni Maast.	
	3.1b	Develop automated data linkage	VU	
Software for Data Paradigms	3.2a	Develop Media Content Analysis Lab	UvA	
	3.2b	Pilot Mass Experiments Online lab	Uni Utrecht	
	3.2c	Pilot Citizen Science Suite	Uni Utrecht	
Innovative Survey Experiments	3.3b	Refresh the LISS panel	CentERdata	
	3.3c	Improve access to the LISS panel	CentERdata	
Hub				
Community Management	4.1a	General administration of the project	EUR	
	4.1b	Community development	EUR	
	4.1c	Create FAIR data	DANS	
Educational Programme	4.2a	Organise ODISSEI conference	EUR	
	4.2b	Organise ODISSEI workshops	EUR	
	4.2c	Coordinate teaching programmes	EUR	
Stimulating Social Analytics	4.3a	Social analytics team	Uni Utrecht	
	4.3b	Social science benchmarks	NLeSC	
	4.3c	Administer grants for analytical support	NLeSC	

Deliverables:

Concept
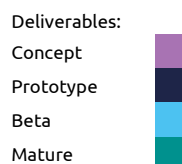
Prototype

Beta

Mature

Figure 6 – Program of work

2.4.7 Risk analysis

Like any other enterprise, ODISSEI will be confronted with some risks, which can be particularly significant at the interface between implementation and operational phase. Table 5 provides a list of such risks. A relatively simple risk management system will be implemented, checked quarterly, and kept up to date by the Management 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 alerts the Managing Director if there is a probability or impact of a risk and proposes mitigating actions. The Management Board will perform bimonthly audits of the risk register and flag problems to the Managing Director. If a risk increases in probability or impact it will be reported to the Management Board and Supervisory Board at the next scheduled meeting.

Risk factors	Risk owner	Probability	Impact	Mitigating actions
Financial				
Financial Commitments limited to term of project	MB	Moderate	Moderate	Renewal discussions to commence in 2021
Lack of Budgetary Compliance	CT	Low	Moderate	Pre-proposal FLAT check at EUR
Organisational				
Dysfunctional Governance	SB	Moderate	High	Annual evaluation by SB & AB
Dysfunctional Coordination Team	MB	Low	High	Quarterly Evaluation by MB
Insufficient Resources for tasks	CT	Low	Moderate	Bi-Monthly updates on budgets and progress by task leaders
Insufficient commitment from task leads	CT	Moderate	High	Consortium Agreement signed
Poor integration within international initiatives	MB	Low	Low	Responsible MB member assigned
Technical				
Delay in delivery of platforms or tools	CT	High	Low	Each task to develop contingency plans at project outset
Diverse metadata impedes Portal development	CT	Moderate	Low	DANS & Statistics Netherlands to appoint leads for mapping metadata standards
Linkage limited by the use of Persistent Identifiers	CT	Moderate	Moderate	Utilisation of standard codings and classifications where possible
Low demand for analytical support	CT	Low	Low	Community consultation and awareness raising in first 6 months
Legal & Ethical				
Inadvertent Personal Data Disclosure	MB	Very Low	High	Transparent and robust disclosure checks by Data Controllers
Security breach in the ODISSEI Data Facility	MB	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 ODISSEI	MB	Low	High	Code of Ethics embedded in the user agreement signed by all users

Table 5 – Risk analysis[52]

---

[52] SB = Supervisory Board, MB = Management Board, CT = Coordination Team

2.4.8   Financial feasibility

All the costs detailed in this analysis have been estimated and collated in cooperation with ODISSEI partners and are based on provisional estimates regarding the tasks outlined above. The estimates have been assessed and approved by financial controllers at EUR and adhere to the guidelines of the call. The budget is financed by four streams of revenue. First, the existing commitments of ODISSEI member organisations through 2020-2024. This has been committed in the form of cash contributions from all member organisations, paying an annual contribution fee, totalling to € 3.18 million over the five years of the project. These commitments are supported by written affirmations in the annex of this proposal. Second, by cash contributions from ODISSEI's equity and from an earlier NWO contribution. These cash contributions combined total to € 1.46 million and are also supported by written affirmations in the annex of this proposal. Third, € 9.30 million are requested from NWO as part of this proposal. The funds will be spent over five years (2020-2024). The costs are split into the four Work streams and spread more or less equally over the years. Fourth, for the years 2020-2029, key partner organisations have committed to the continued operation of the Hub and Data Facility which represent the core operational activities of the infrastructure, as in-kind contributions. These in-kind contributions, totalling to € 7.24 million, are also supported by written affirmations in the annex of this proposal.

All staff and material costs associated with each task and subtask are detailed in the task descriptions in sections 2.4.2 – 2.4.5[53]. The Work streams, excluding in-kind contributions, sum up to Data Facility (€ 4,832,000), Observatory (€ 3,418,000), Laboratory (€ 2,647,000), Hub (€ 3,043,000). The total project budget for 2020-2024, excluding in-kind contributions, is € 13,940,000.

Since ODISSEI will not be building a physical infrastructure but a federated digital and human one, disentanglement, legacy and decommissioning costs are not applicable. In-kind commitments of € 994,000 per year on have been guaranteed by key partners for the period 2025-2029 for the maintenance of the ODISSEI Portal, the ODISSEI Secure Supercomputer, Microdata Services Facility and the core of the Coordination Team. The total capital and running costs for 2020-2029 inclusive of in-kind contributions are € 21.18 million. During the project, ODISSEI will also seek to extend and expand the commitments for cash contributions through membership fees from 2025 onward. It is estimated that by 2025 the ongoing cash contributions to ODISSEI would total approximately € 1 million per year annually. Taken in combination with in-kind commitments, ODISSEI will therefore be able to cover the operation of the existing infrastructure from 2025 onward by covering the ODISSEI Secure Supercomputer, ODISSEI Portal, Microdata services, fieldwork, grants, social analytics team, community officers, and educational officers.

Day-to-day financial management lies with the Managing Director, supported by the Chief Financial Officer. The task and subtask leaders financially report yearly to the Work Stream lead. The Work Stream leads are responsible for quarterly reporting to the Management Board on their stream's financial and technical progress. The Management Board yearly reports to the Supervisory Board via the Annual Report. All are supported by the Chief Financial Officer, working in the Hub at the host member organisation. The Managing Director is also responsible, alongside the Chief Financial Officer, for reporting to the NWO project officer.

2.4.9   Roles & responsibilities

The project consists of 13 partners who execute the subtasks. All of these partners are ODISSEI member organisations. All partners have signed a consortium agreement. This constitutes a formal agreement on obligations as described in this proposal and details payment conditions, legal liability, intellectual property, and accessibility.

2.4.9.1   Erasmus University Rotterdam

The Erasmus School of Social and Behavioural Sciences (ESSB) at the Erasmus University Rotterdam (EUR) acts as the coordinator for ODISSEI. ESSB is responsible for managing the project and implementing the work plan as agreed upon by the Management Board of ODISSEI (see section 2.3.1.3). EUR has been the host of ODISSEI since its inception, under the leadership of the Scientific Director, Pearl Dykstra. EUR provides legal support including the consortium agreement between all partners, tenders the fieldwork described in the Observatory Work stream, and bears responsibility for the financial and legal administration of the project. They will also host ODISSEI and employ the staff that constitutes the coordination team. EUR's chair of the Executive Board approved the hosting of ODISSEI at EUR and the financial, legal, administrative and tax elements of this proposal. They

---

[53] In-line with the guidelines, in the budget non-VNSU staff are classified in the budget Excel file as material costs. The task descriptions however include staff estimates for both VSNU and non-VSNU staff.

ensure that the interests of this diverse range of stakeholders are all recognised and accounted for whilst ensuring that the scientific potential of the data infrastructure can be realised. ESSB provides a solid base for ODISSEI given its expertise with respect to:

- Identifying and linking datasets from academia, cities, and non-governmental organisations, as well as the assessment of the quality of these datasets (e.g. BOLD cities);
- A comprehensive range of traditional and new analytic techniques for analysing big, open and linked data that follow Fair, Accurate, Confidential, Transparent (FACT) principles by design;
- A thorough understanding of how the 'data-revolution' is challenging academic, policy and administrative conventions and of the changes in governance it requires, especially regarding: (a) interorganisational multidisciplinary collaboration, (b) shared understanding and procedures for Findable Accessible Interoperable Reusable (FAIR) data management, (c) compliance to the EU General Data Protection Regulation, and (d) sensitivity to ethical and affective concerns regarding data (e.g. RISBO);
- Complex stakeholder management, reconciling the different pace and interest of governmental and academic institutions, as well as respecting the resulting differences in work styles and culture[54].

Pearl Dykstra has a chair in Empirical Sociology and is director of research of the Department of Public Administration and Sociology at the Erasmus University Rotterdam. She is the director of the Netherlands Kinship Panel Study (NKPS), and a member of the Consortium Board of the Generations and Gender Programme (GGP). Dykstra is a member of the Royal Netherlands Academy of Arts and Sciences and served as vice-president of that organisation. She is a fellow of the Gerontological Society of America, member of the Royal Holland Academy of Sciences, and a member of the Academia Europaea. In 2015, she was appointed in the group of Chief Science Advisors of the College of European Commissioners, and currently she serves as Deputy Chair. Dykstra also serves on the Board of the Social Sciences and Humanities Division of the Netherlands Organisation for Scientific Research. Professor Dykstra is the **Scientific Director** of ODISSEI.

Lucas van der Meer is the Managing Director of ODISSEI and oversees the day-to-day management of the ODISSEI coordination team based at EUR. Van der Meer has a master's degree in computer science and is co-founder of Landscape Data Science and of CodeCombat Nederland.

Marlon Domingus is the current Data Protection Officer for Erasmus University Rotterdam and will act as the ODISSEI Data Officer for the duration of the project. In his role at EUR, Domingus deploys all means available for securing privacy in research, education and business operations, in collaboration with the EUR Privacy Organisation and the many professionals at Erasmus University Rotterdam. Domingus is an expert in research data management, privacy by design and the implementation of GDPR.

Tasks - Within the project Erasmus will act as coordinator of the project and will lead and execute task 4.1 (Community Management) and 4.2 (Education) in the Hub. EUR will also outsource the fieldwork in the Observatory. EUR have committed to the long-term hosting of ODISSEI in the five years after the conclusion of the project by supplying the Scientific Director and the Managing Director in-kind for € 162,000 per year.

2.4.9.2 SURFsara

SURFsara is the Dutch national high-performance computing centre. SURFsara is a SURF organisation providing an integrated ICT research infrastructure for science and the knowledge economy that works together with the academic community (including researchers, educational institutions, and academic medical centres), industry, and SMEs. SURFsara offers a complete package of services in the field of High-Performance Computing (HPC), Data Services, Visualisation, and Cloud Services, and is, amongst others, the home of the national supercomputer Cartesius. The system is used for research areas, such as sustainable energy, climate change, water management, product and process optimisation, reduction of noise pollution, and improvement of medical treatments. SURFsara is also a partner in national and international projects, such as LOFAR, EGI, WLCG/LHC, EUDAT and PRACE. Within these projects, the most powerful European supercomputers and data storage systems for petabytes of scientific data are used. SURFsara offers a full range of services, expertise and support in the field of high-performance computing, data services, visualisation, cloud services, and networking. SURFsara is ISO 27001 certified, meaning it complies with the high requirements of this international standard in the field of information security.

Irene Nooren is a community manager at SURFsara which involves managing strategic roadmaps, the common interest and initiating innovation and research & development projects. Nooren is a member of the ODISSEI

---

54 www.eur.nl/en/essb/essb-society/networking-and-collaboration

Management Board. Her expertise lies in life sciences and health e-infrastructure and working with privacy sensitive data.

Narges Zarrabi is a data consultant and provides general and technical advice on data management. She contributes to development and support of IT services to facilitate research and education. She is involved in national and international projects such as CompBioMed and EOSC-hub, and contributes to technical requirements, gathering, defining workflows, documentation and training materials.

Tasks - Within the project SURFsara will be responsible for the development and extension of the ODISSEI Secure Supercomputer as part of task 1.2. SURFsara are also involved in the implementation of the ODISSEI Data Node (task 1.3 - ODISSEI Data Facility). SURFsara have committed to supplying ODISSEI Secure Supercomputer support for the five years after the conclusion of the project through an in-kind contribution of € 200,000 per year.

2.4.9.3 Statistics Netherlands

Statistics Netherlands provides access to administrative data for scientific and social policy driven research through its Microdata Services facility[55]. Whilst the data collection is financed by the Statistics Netherlands structural budget, the Microdata service to external researchers is not, and the costs of the department have to be fully covered by its users. On an annual basis, the facility is used by typically a thousand researchers from authorised institutions like universities, from both inside and outside the Netherlands. Statistics Netherlands constructed and hosts the so-called System of Social Statistical Datasets (SSD). The SSD is an integrated, longitudinal system of registers and surveys, containing, among others, important socio-economic, demographic, health, education, crime, income, and wealth variables of the integral 17 million population of the Netherlands (Bakker et al., 2014). In addition, Statistics Netherlands hosts integral employer-employee data and a wide range of business economic data, both largely integral, register-based. The data typically stretch back over 10-25 years. All data can be linked at record level via robust persistent linkage keys for persons, addresses, enterprises, etc; these meaningless linkage keys exist only within the domain of Statistics Netherlands, but can conceptually be considered as pseudonymised equivalents of identifiers such as the Citizen Service Number, Chamber of Commerce Number, etc. Statistics Netherlands spends in the order of € 35 million per year collecting their data.

Access to the data is governed by Statistics Netherlands law with the leading principle that data may only be used for statistical purposes and that no information may leave the domain of Statistics Netherlands that could possibly related to individual persons, households, or organisations[56]. Therefore, access is organised in a strictly secured network environment, accessible via remote access using secured individual tokens and covered with various legal contracts, and with strict pseudo-anonymisation at all levels. Research projects are based on the need-to-know principle, and all outputs are checked for non-disclosure before being released to the researchers.

Ruurd Schoonhoven is the Senior Account Manager within the Microdata Services department Statistics Netherlands. He is also a member of the ODISSEI Management Board. Schoonhoven is responsible for the maintenance of ties with Statistics Netherlands microdata users, including the Microdata Access Program within ODISSEI, and is specialised in aspects like data stewardship, legal conditions, data security, linkage methods, and innovation of the microdata facility.

Wim Schaasberg is the deputy director of the department of Policy Statistics and Data Services and head of the Data Services team at Statistics Netherlands. His team is responsible for archiving and dissemination of data and metadata at Statistics Netherlands and for granting access to Statistics Netherlands microdata to external researchers. This implies responsibility for data security, GDPR and Statistics Netherlands Law compliance and data and metadata quality.

Tasks - Within the project Statistics Netherlands will be responsible for extension of the Microdata Services Facility as part of task 1.1 within the ODISSEI Data Facility. Statistics Netherlands will also execute the integration of metadata into the ODISSEI Portal (subtask 1.3a - Data Facility). Statistics Netherlands have committed to supplying their collection of administrative microdata to the research community in the five years after the conclusion of the project through € 455,000 per year.

---

[55] Statistics Netherlands Microdata Services: www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research
[56] Statistics Netherlands Law wetten.overheid.nl/BWBR0015926/2019-01-01

2.4.9.4 DANS

DANS (Data Archiving and Networked Services) is the Netherlands Institute for permanent access to digital research resources. DANS encourages researchers to make their digital research data and related outputs Findable, Accessible, Interoperable, and Reusable. DANS provides expert advice and certified services. Their core services are: DataverseNL for short-term data management, EASY for long-term archiving, and NARCIS, the national portal for research information. By participating in national and international projects, networks, and research, DANS contributes to continued innovation of the global scientific data infrastructure. Open if possible, protected where necessary. DANS is an institute of the Dutch Academy KNAW and funding organisation NWO.

In addition to these services, DANS guides other archives, research institutes and research financiers to questions relating to data management, certification and subjects such as FAIR, open access and software sustainability.

Herbert van de Sompel is Chief Innovation Officer at DANS. He has previously held positions as head of Library Automation at Ghent University, Visiting Professor in Computer Science at Cornell University, Director of e-Strategy and Programmes at the British Library, and information scientist at the Research Library of the Los Alamos National Laboratory where he was the team leader of the Prototyping Team. Herbert has played a major role in creating the Open Archives Initiative Protocol for Metadata Harvesting, the Open Archives Initiative Object Reuse & Exchange specifications, the OpenURL Framework for Context-Sensitive Services, Open Annotation, Robust Links, and Signposting the Scholarly Web.

Tasks - Within the project DANS will be responsible for the development and operation of the ODISSEI Portal as part of task 1.3 within the ODISSEI Data Facility and for FAIR data support (subtask 4.1c). DANS will be responsible for maintaining the ODISSEI Portal beyond the duration of the project. DANS have committed to supplying hosting and support for the ODISSEI Portal and ODISSEI Data Node in the five years after the conclusion of the project through € 167,000 per year.

2.4.9.5 CentERdata

CentERdata is a non-profit research institute specialised in data collection and applied and methodological research. Staff members are experts in collecting and analysing panel data and putting these at the disposal of academic researchers. The institute manages and maintains several panels, including the LISS panel, that regularly complete questionnaires through the Internet. A professional helpdesk and panel management staff support the fieldwork. Data scientists at CentERdata work on techniques and applications related to massive amounts of complex data from various sources. They employ machine and deep learning, data and text mining, and visualisation techniques. By doing so, new insights are obtained, and societal value is created from all sorts of data, such as sensor data, social media data, or data scraped from the Internet. CentERdata also has many years of experience in developing software to support panel research, both online and offline. Academically skilled IT-staff have ample experience with a great diversity of information systems and have designed and developed tailor-made software for large-scale international data infrastructures such as SHARE, ESS, and EVS. This allows for language-independent questionnaire development in order to reach out to international target groups. CentERdata is located on the campus of Tilburg University and mostly works for universities, research platforms, Dutch ministries, and the European Commission.

Marcel Das is Director of CentERdata and Professor of Econometrics and Data Collection at Tilburg University. He is a member of the ODISSEI Management Board. Das is one of the principal investigators of the LISS panel and member of several (inter)national scientific advisory boards of large-scale data infrastructures. In 2016, he was appointed as a member of the Advisory Council on Methodology and Quality of Statistics Netherlands.

Seyit Höcük is a data scientist at CentERdata with a PhD in astrophysics. He works with large volumes of data, involving predictive modelling, utilising machine learning and deep learning, and AI techniques, such as natural language processing.

Joris Mulder is a survey researcher and data scientist at CentERdata with a background in both technical computer science and social and economic psychology. He specialises in experimental research projects with a technical and innovative nature.

Tasks - Within the project CentERdata will be responsible for the development of the LISS Panel (task 3.3).

2.4.9.6 eScience Center

The Netherlands eScience Center (NLeSC) is a joint initiative of the Dutch national research council (NWO) and the Dutch organisation for ICT in education and research (SURF). It serves as the national hub for the development and application of domain overarching software and methods for the scientific community. High-

level research experts develop bridges between the ambitions of scientists from across all disciplines and the increasingly complex modern e-infrastructures. The application of digitally enhanced scientific practices, referred to as eScience, is a fundamental toolbox for all researchers. In support of this goal the NLeSC funds and participates in multidisciplinary projects with academia. Their core expertise is in the fields of optimised data handling, efficient computing and big-data analytics. NLeSC will be responsible for ensuring that computational social scientists have access to the expertise necessary to apply advanced computational methods on social science problems. Furthermore, the NLeSC and SURFsara formed an alliance with the mission to enhance research by providing an online platform[57] to benchmark the performance of research software. This platform will be used to set-up benchmarks for social sciences to boost analytics on research topics that are relevant for social scientists and to enable open assessment of research software on these topics.

Rob van Nieuwpoort is director of Technology at NLeSC, and professor of Efficient Computing for eScience at the University of Amsterdam. He is a member of the ODISSEI Management Board. He initiated the Ibis project[58] on high-performance distributed computing and deployment. His work on the Grid Application Toolkit resulted in an Open Grid Forum standard. He is one of the key players in eScience both nationally and internationally.

Frank Seinstra is the Director eScience Program at NLeSC. He has received several international awards (including a Most Visionary Research Award at AAAI 2007) and has won First Prizes in several international computing challenges (including IEEE SCALE 2008, IEEE DACH 2008, IEEE SCALE 2010). In May 2012, he joined NLeSC, where he supervises the development, implementation, and research of innovative eScience solutions in multidisciplinary research projects. He worked for nearly 15 years as a scientific researcher in the fields of High-Performance and Distributed Computing, and eScience.

Adriënne Mendrik is part of the management team at NLeSC, where she coordinates a wide variety of projects and leads the online benchmark platform (EYRA) project. She has extensive experience with designing benchmarks and challenges for measuring algorithm performance in biomedical image analysis and has organised various workshops and tutorials on this topic at international conferences (IEEE ISBI, IEEE eScience, MICCAI).

Tasks - Within the project NLeSC will be working on subtask 4.3c regarding pathfinder analytic grants and will assist in the organisation of grand data challenges and prizes (subtask 4.3b).

2.4.9.7   VU Amsterdam

Frank van Harmelen is Professor of Knowledge Representation at the VU University Amsterdam. He is a member of the Management Board of ODISSEI. He played a leading role in the development of the Semantic Web, which aims to make data on the web semantically interpretable by machines through formal representations. He is one of the architects of the semantic storage engine Sesame, with over 200,000 downloads worldwide. This work received the ten-year impact award of the Semantic Web community. Van Harmelen is a member of the Royal Netherlands Academy of Arts and Sciences, the Royal Holland Academy of Sciences, and Academia Europea, and fellow of the European Association for Artificial Intelligence.

Jacco van Ossenbruggen leads the Information Access group at Centrum Wiskunde & Informatica (CWI) in Amsterdam, and the User Centric Data Science research group at VU Amsterdam. His research interests include assessment of data quality in large, heterogeneous research datasets in the digital humanities and computational social sciences, large scale web data integration and linking in cultural heritage and other data rich domains. He has been involved in many EU projects, including VRE4EIC, Fish4Knowledge and EuropeanaConnect.

Dorret Boomsma is Professor of Biological Psychology at the VU Amsterdam. Her research focuses on the causes of differences in behaviour and health as a function of genes and environment. She established the Netherlands Twin Register (NTR), which since 1985 recruited more than 200,000 members from twin families and is a resource for studies of human complex traits. Her work has resulted in more than 1,100 papers and multiple awards, including the Spinoza prize and the Dobzhansky Award for lifetime achievement in behaviour genetics. She is a member of the Royal Netherlands Academy of Arts and Sciences and the Academia Europaea.

Eveline de Zeeuw is an Associate Professor in the Department of Biological Psychology at VU Amsterdam. She is responsible for the data collection of data from teachers of twins and their siblings in the Netherlands Twin Register (NTR). She led the ODISSEI pilot project in which NTR data were linked to data from Statistics Netherlands in a high-performance computing environment.

---

[57] www.eyrabenchmark.net
[58] www.cs.vu.nl/ibis/projects.html

Tasks - Within the project VU Amsterdam will be responsible for several tasks. In the ODISSEI Data Facility, VU Amsterdam will be responsible for the design of the ODISSEI Portal in task 1.3 (overseen by Van Ossenbruggen). Related to this work, they will also lead the development of operational innovations within task 3.1 of the Laboratory (overseen by Van Ossenbruggen) and execute subtask 3.1b on automatic data linkage. In the Observatory, VU will be responsible of the integration of the NTR within ODISSEI and the development of protocols for integrating biomedical cohorts (subtask 2.2c - overseen by Boomsma).

2.4.9.8 Utrecht University

Tanja van der Lippe is Professor of Sociology at the Department of Sociology of Utrecht University, head of the Department of Sociology and chair of the ICS research school. Van der Lippe is a *member of the ODISSEI Management Board*. Her research interests are in the area of work-family linkages in Dutch and other societies, for which she received many large-scale grants from Dutch and European Science Foundations. She received an ERC Advanced Grant for her research into 'Investments in a sustainable workforce in Europe'. She is a member of the Royal Netherlands Academy of Arts and Sciences, the Royal Holland Academy of Sciences and the European Academy of Sociology.

Arnout van de Rijt is Professor of Sociology at Utrecht University. He received his PhD. in Sociology from Cornell University in 2007 and worked until 2016 as Assistant and Associate Professor of Sociology at Stony Brook University in the USA, where he co-led the Center for Computational Social Science. For his work, Van de Rijt received the Lynton Freeman (2010) and Raymond Boudon (2017) early career awards. His work has been funded by the National Science Foundation.

Daniel Oberski is Associate Professor in Data Science Methodology at Utrecht University. He is a co-founder of the university-wide focus area in Applied Data Science and coordinates the focus area's special interest groups on Machine Learning. Oberski is a member of the Young Academy of the Royal Netherlands Academy of Arts and Sciences. In 2019 he was awarded a personal VIDI grant.

Peter Lugtig obtained his PhD in 2012 at Utrecht University on measurement and non-response errors in panel studies. He is currently Associate Professor at the Methodology and Statistics Department at Utrecht University. His research focuses on improving the design and analysis of surveys and the use of sensor data in smartphones.

Tasks - Within the project UU will be involved in several tasks. In the Laboratory, UU will execute the development of the Mass Experiments Online Lab (subtask 3.2b - overseen by Van de Rijt) and the Citizen Science Platform (subtask 3.2c - overseen by Lugtig). In the Hub, UU will lead task 4.3 on stimulating social analytics and will provide analytics support to the whole social science community (subtask 4.3a – overseen by Oberski).

2.4.9.9 Maastricht University

Michel Dumontier is the Distinguished Professor and Director of the Institute of Data Science at Maastricht University (UM) and the co-founder of the FAIR data principles. His research focuses on the development of computational methods for scalable and responsible discovery science. He is a principal investigator in the Dutch National Research Agenda, the VSNU Digital Society Initiative, the European Open Science Cloud, the NIH/NCATS Biomedical Data Translator programme. He is the co-chair for the W3C Community Group on Semantic Web for health care and life sciences, the editor-in-chief for the journal Data Science, and an associate editor for the journal Semantic Web.

Lianne Ippel is a Postdoctoral researcher at the Institute of Data Science at Maastricht University. She received her PhD degree from Tilburg University for her thesis "Multilevel Modelling for Data Streams with Dependent Observations", for which she won 'Best Thesis Award' at the General Online Research conference in Cologne (2018). Her research interests focus on ethical and responsible use of machine learning and machine learning models in relation to methodological issues such as response style, measurement invariance, and missing data.

Tasks - Within the project UM will be leading the subtask on distributed computing (subtask 3.1a) in the Laboratory.

2.4.9.10 Leiden University

Marike Knoef is Professor of Empirical Microeconomics at the Department of Economics at Leiden University (UL) and Director of the Network for Studies on Pensions, Aging and Retirement (Netspar). She has extensive experience in engaging diverse stakeholders such as pension companies and government ministries with research via NETSPAR. In 2017, she received the World Cultural Council Special Recognition Award for leading scientists and in 2018 she was nominated for the Huibregtsen Prize. She is successful in bridging the gap between research and practice.

Tasks - In the project, LU will oversee task 2.1 and 2.3 in the Observatory which seeks to coordinate the international survey commitments within the Netherlands and advance the strategic position of ODISSEI partners within ESFRI and the European landscape of social science.

2.4.9.11 University of Amsterdam

Rens Vliegenthart is a Professor of Media and Society and Scientific Director of the Amsterdam School of Communication Research, University of Amsterdam (UvA). He received his PhD in Social Sciences from VU Amsterdam in 2007 (cum laude) and held visiting positions at the University of California Irvine, and the University of Southern Denmark. His research focuses on media content and effects, both on citizens and politicians. His research has received various large grants (e.g., NWO VENI and VIDI). Vliegenthart is also editor of Acta Politica and member of the advisory council of the SSH section of the NWO.

Tasks - Within the project UvA will be responsible for the new software for new data paradigms task (3.2) and execute the development Media Content Analysis Lab (subtask 3.2a) as part of the Laboratory. UvA will also be responsible for maintaining and sustaining the Media Content Analysis Lab after the project has been completed.

2.4.9.12 Netherlands Interdisciplinary Demographic Institute (NIDI)

NIDI is the national demographic institute of the Netherlands. With its leading scientific research of population issues the mission of NIDI is to contribute to the solution of societal challenges with a focus on migration and integration, labour market and social security, ageing and family relationships. A central characteristic of NIDI is the inter-disciplinary and inter-national approach, and the fostering of young research talent. NIDI is a research institute of the Royal Nether-lands Academy of Arts and Sciences (KNAW) and is affiliated with the University of Groningen (UG).

Tom Emery is the Deputy Director of the Generations and Gender Programme and is a member of the ODISSEI Management Board. Emery has a PhD in Social Policy from the University of Edinburgh and an MBA in Research Infrastructure Management from the University of Bicocca. He has been involved in major international research projects on European Social Science Infrastructure including SERISS and SSHOC.

Tasks - Within the project NIDI will lead the Data Scout task (2.2) and execute subtask 2.2a within the Observatory which includes the coordination of the Data Scout team which seeks to engage data providers from the public and private sectors to allow their data to be securely linked within the ODISSEI Data Facility.

2.4.9.13 International Institute for Social History (IISG)

The IISG serves science and society on a global scale. At an international level, IISG generates and offers reliable information and insights about the (long-term origins, effects and consequences of social inequality. To promote this, IISG forms an international hub for social historians worldwide, offering and producing historical sources and data, facilitating social-history research and collaborating internationally in ground-breaking research projects. Moreover, by preserving the heritage of often oppressed social movements, the Institute serves the quality of the world's memory.

Kees Mandemakers is senior research fellow, directing the Historical Sample of the Netherlands (HSN) and the LINKS database. He is professor of Large Historical Databases at the Erasmus School of History, Culture, and Communication in Rotterdam. He is chair of the European Historical Population network and former president of the International Commission of Historical Demography. Main research interest: Methodology of large historical databases, Family and Demography, Social mobility.

Tasks - In the project, IISG will oversee the integration of HSN within ODISSEI as part of subtask 2.2b. This will include the linkage of HSN data with the Social Statistical Datasets held by Statistics Netherlands and the integration of the HSN within the ODISSEI Portal.

2.5 Literature references

Allcott, H., and Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, 104(10), 3003-3037.

Arntz, M., Gregory, T., and Zierahn, U. (2016). The risk of automation for jobs in OECD countries: A comparative analysis. *OECD Social, Employment, and Migration Working Papers* 189.

Bail, C.A., Merhout, F., and Ding, P. (2018). Using Internet search data to examine the relationship between anti-Muslim and pro-ISIS sentiment in US counties. *Science advances* 4, no. 6: eaao5948.

Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P. Haohan Chen, Fallin Hunzaker, M.B., Lee, J., Mann, M., Merhout, F., and Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences 115, no. 37, 9216-9221.

Bakker, B.F.M., Van Rooijen, J., and Van Toor, L. (2014). The system of social statistical datasets of Statistics Netherlands: An integral approach to the production of register-based social statistics, *Statistical Journal of the IAOS*, 30(4), 411–424.

Bauman, Z. (2013). Liquid modernity. *John Wiley & Sons.*

Benoit, K., Conway, D., Lauderdale, B. E., Laver, M., and Mikhaylov, S. (2016). Crowd-sourced text Analysis: Reproducible and agile production of political data. *American Political Science Review, 110*(2), 278-295.

Boas, T. C., Christenson, D. P., Glick, D. M. (2018, online first). Recruiting large online samples in the United States and India: Facebook, Mechanical Turk, and Qualtrics. *Political Science Research and Methods.*

Bol, T., Ciocca Eller, C., Van de Werfhorst, and DiPrete, T.A. (2019). School-to-Work Linkages, Educational Mismatches, and Labor Market Outcomes. *American Sociological Review*.

Borghans, L., Gielen, A.C., and Luttmer, E.F.P. (2014). Social support substitution and the earnings rebound: Evidence from a regression discontinuity in disability insurance reform. *American Economic Journal: Economic Policy* 6, no. 4, 34-70.

Boumans, J.W., and Trilling, D. (2016). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital journalism* 4, no. 1, 8-23.

Bréchon, P., & Gonthier, F. (Eds.) (2017). European values: Trends and divides over thirty years. Leiden: Brill.

Castells, M. (2011). The rise of the network society: The information age: Economy, society, and culture. Vol. 1. *John Wiley & Sons.*

Centola, D., and Macy, M. (2007). Complex contagions and the weakness of long ties. *American journal of Sociology* 113, no. 3, 702-734.

Colodro-Conde, L., Couvy-Duchesne, B., Whitfield, J.B., et al. (2018). Association between population density and genetic risk for schizophrenia. *JAMA Psychiatry*, 75(9), 901-910.

Connelly, R., Playford, C.J., Gayle, V., and Dibben, C. (2016). The role of administrative data in the big data revolution in social science research, *Social Science Research*, 59, 1-12.

Conrad, F.G., and Schober, M.F. (Eds.) (2007). Envisioning the survey interview of the future. Vol. 542. *John Wiley & Sons*.

Conti, G., Heckman, J., and Urzua, S. (2010). The education-health gradient. *The American Economic Review* 100(2), 234-238.

Currie, J. (2013). "Big data" versus "big brother": On the appropriate use of large-scale data collections in pediatrics. *Pediatrics*, 131(Supplement 2), S127-S132.

De Graaf, B., and Rinnooy Kan, A. (2017). Hoe zwaar is licht? Meer dan 100 dringende vragen aan de wetenschap (in Dutch), *Amsterdam: Uitgeverij Balans/KNAW.*

Donoho, D. (2017). 50 years of data science*. Journal of Computational and Graphical Statistics*, 26(4), 745-766.

European Commission (2018). Cost of not having FAIR research data: Cost-benefit analysis for FAIR data research. Luxembourg: Publication Office of the European Union.

Galesic, M., Tourangeau, R., Couper, M.P., and Conrad, F.G. (2008). Eye-Tracking data new insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly* 72(5), 892-913.

Groves, R. M. (2011). Three eras of survey research. *Public Opinion Quarterly*, 75(5), 861-871.

Hagestad, G. O., and Dykstra, P. A. (2016). Structuration of the life course: Some neglected aspects. In M. J. Shanahan J. T. Mortimer, & M. Kirkpatrick Johnson (Eds.), *Handbook of the life course, Volume II*. New York: Springer, 131-157

Halford, S., and Savage, M. (2017). Speaking sociologically with big data: Symphonic social science and the future for big data research. *Sociology*, 51(6), 1132-1148.

Hoogendijk, E.O., Deeg, D.J.H., Poppelaars, J., Van der Horst, M., Broese van Groenou, M.I., Comijs, … and Huisman, M. (2016). The Longitudinal Aging Study Amsterdam: Cohort update 2016 and major findings. *European Journal of Epidemiology* 31(9), 927-945.

Kapteyn, A., Banks, J., Hamer, M., Smith, J.P., Steptoe, A., Van Soest, A., Koster, A., and Saw, Htay-Wah (2018). What they say and what they do: Comparing physical activity across the USA, England, and the Netherlands, *Journal of Epidemiology & Community Health,* 72(6), 471-476.

Koens, L., Vennekens, A., Hofman, R., Van den Broek-Honingh, N., and De Jonge, J. (2018). Balans van de wetenschap 2018 [Balance of science 2018]. Den Haag: Rathenau Instituut.

Kooreman, P. and Scherpenzeel, A. (2014). High frequency body mass measurement, feedback, and health behaviours, *Economics & Human Biology*, 14, 141-153.

Lenzner, T., Kaczmirek, L., and Galesic, M. (2011). Seeing through the eyes of the respondent: An eye-tracking study on survey question comprehension.

International Journal of Public Opinion Research, 23(3), 361-373.

Mann, A. (2016). Core concept: Computational social science. Proceedings of the National Academy of Sciences 113, no. 3, 468-470.

Oldehinkel, A.J., Rosmalen, J.G.M., Buitelaar, J.K., Hoek, H.W., Ormel, J., Raven, D., … Hartman, C.A. (2015). Cohort profile update: The tracking adolescents' individual lives survey (TRAILS). *International Journal of Epidemiology* 44(1), 76-76.

Pickard, M. D., Roster, C.A., and Chen, Y. (2016). Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human behaviour,* 65, 23-30.

Reimsbach-Kounatze, C. (2015). The proliferation of "Big Data" and implications for official statistics and statistical agencies. *OECD Digital Economy Papers* 245.

Robinson-Garcia, N., Sugimoto, C. R., Murray, D., Yegros-Yegros, A., Larivière, V., and Costas, R. (2019). The many faces of mobility: Using bibliometric data to measure the movement of scientists. *Journal of Infometrics*, 13(1), 50-63.

Salganik, M. J. (2018). *Bit by bit: Social research in the digital age*. Princeton, NJ: Princeton University Press.

Stone, A., Shiffman, S., Schwartz, J.E., Broderick, J.E., and Hufford, M.R. (2002). Patient non-compliance with paper diaries, *British Medical Journal*, 324, 1193.

Sugie, N. F. (2018). Utilizing smartphones to study disadvantaged and hard-to-reach groups. *Sociological Methods & Research, 47*(3) 458-49.

Sun, C., Ippel, L., Wouters, B., Van Soest, J., Malic, A. Adekunle, O., Van den Berg, B., et al. (2018). Analyzing Partitioned FAIR Health Data Responsibly. arXiv preprint arXiv:1812.00991.

Thorvaldsen, G., and Østrem, N.O. (2018). Migration and the Historical Population Register of Norway. *Journal of Migration History 4*, no. 2, 237-248.

Trilling, D., and Jonkman, J.G.F. (2018). Scaling up content analysis. *Communication Methods and Measures 12*, no. 2-3, 158-174.

Tropf, F.C., Stulp, G., Barban, N., Visscher, P.M., Yang, J., Snieder, H., and Mills, M.C. (2015). Human fertility, molecular genetics, and natural selection in modern societies. *PloS one*, 10(6).

Turner, G., Van Zoonen, L., and Adamou, B. (2014). Research through gaming: Public perceptions of (the future of) identity management. SAGE Research Methods Cases.

Van Atteveldt, N., Murray, M.M., Thut, G., and Schroeder, C.E. (2014). Multisensory integration: flexible use of general operations. *Neuron* 81, no. 6, 1240-1253.

Van de Rijt, A. (2019). Self-Correcting Dynamics in Social Influence Processes. *American Journal of Sociology* 124, no. 5, 1468-1495.

Van der Lippe, T., De Ruijter, J., De Ruijter, E., and Raub, W. (2010). Persistent inequalities in time use between men and women: A detailed look at the influence of economic circumstances, policies, and culture. *European Sociological Review* 27, no. 2, 164-179.

Van Dijck, J., and Van Saarloos, W. (2017). The Dutch polder model in science and research. Amsterdam: Royal Netherlands Academy of Arts and Sciences (KNAW).

Verma, I. M. (2014). Editorial expression of concern: Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the USA, 111*(29), 10779.

Wallach, H. (2016). Computational social science: Towards a collaborative future. In B. Alvarez (Ed.), Computational social science: Discovery and prediction (pp. 307-316). New York: Cambridge University Press.

Wilkinson, M.D., Dumontier, M., Aalbersberg, IJ.J., Appleton, G., Axton, M., …, and Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3.

Willemsen, G., Vink, J.M., Abdellaoui, A., Den Braber, A., Van Beek, J.H.D.A., Draisma, H.H.M., … and Boomsma, D.I. (2013). The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. *Twin Research and Human Genetics* 16(1), 271-281.

Zagheni, E., Weber, I., and Gummadi, K. (2017). Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review, 43*(4), 721-734.

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., and … Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLoS Computational Biology, 13*(3), e1005399.

3 Declaration and signature (by coordinating applicant)

This section is not included in the 50 pages A4 page limit.

3.1 Have you requested funding for this research infrastructure elsewhere?

☒ No

☐ Yes

3.2 Statements by the applicant

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

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

☒   By submitting this document I declare that I satisfy the nationally and internationally accepted standards for scientific conduct as stated in the Netherlands Code of Conduct for Scientific Practice 2012 (Association of Universities in the Netherlands (VSNU)).

☒   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.

☒   I hereby declare that the obligatory letters of commitment of the consortium partners have been uploaded separately in ISAAC.

☒   I have completed this form truthfully.

Name: Prof. dr. Pearl Dykstra

Place: Rotterdam

Date: 4 June 2019