



## E-DATA & RESEARCH

### Special Issue 2014

Newsletter on data and research in the Social Sciences and Humanities.

E-data & Research is published three times a year. This Special Issue is sponsored by CentERdata, CLARIAH, DANS and NIDI.

## CONTENTS

- 2 Data and tools for language research
- 3 CLARIN and DARIAH blend into CLARIAH
- 4 ESS studies a changing Europe
- 5 Innovative ways of data collection



- 6 A digital revolution in humanities



- 7 Family relationships primary focus in GGP
- 8 Collaborations improve research infrastructures



Use your smart-phone to scan this QR-code and visit our website [edata.nl](http://edata.nl)

Neelie Kroes, European Commissioner for Digital Agenda:

# 'Potential for new research is enormous'

Neelie Kroes, Vice President of the European Commission responsible for Europe's Digital Agenda, has strong views on the role of the social sciences and humanities (SSH) in today's society. *Erica Renckens*

"Society is not inert, waiting for technological innovation. Rather, it is an active innovation agent engaging with technologies in a way that often goes far beyond the intentions of the engineers, and that's when innovation really happens! Experts in social sciences and humanities can help policy makers and engineers to harness this vibrant societal intelligence." That's why Kroes supports e-science in the SSH: "Under the upcoming H2020 funding program for Research and Innovation we encourage the integration of SSH considerations in ICT research. We even have a specific cross-cutting topic 'Human-centric Digital Age' where we aim to support future work on integrating SSH knowledge with ICT innovation in all areas."

### Strong focus

"Today we have the technology to enable universal availability and connectivity of all the cultural artefacts of the past in unprecedented ways. Imagine every piece of paper, every bit of text on any medium, from the 19th century backwards, just a few clicks away. This will allow us to bring to life and explore our shared past in a breadth and depth never seen before, which has clear implications for Europe and the



## INTERVIEW

Neelie Kroes is responsible for Europe's Digital Agenda photo Peter Elenbaas / Hollandse Hoogte

work on our European identities." "The Netherlands is already very active in research infrastructures, particularly e-infrastructures. To mention only a few examples: Amsterdam hosts the EGI, the European Grid Initiative. SURF is a highly innovative National Research and Education Network that provides Dutch scientists with extremely fast access to research resources, and recently the Cartesius Supercomputer was installed at SURFsara, which demonstrates once again how Dutch

CONTINUED ON PAGE 3

### Neelie Kroes

Neelie Kroes studied Economics at Erasmus University in Rotterdam. She was a Member of the Dutch Parliament and served as Minister for Transport, Public Works and Telecommunication before she was appointed President of Nyenrode University. Since 2004, she has been a member of the European Commission, currently as Vice President responsible for Europe's Digital Agenda.

## Why the Social Sciences and Humanities need research infrastructures

**Until recently, museums, archives and libraries were the only large facilities used by researchers from the Social Sciences and Humanities (SSH). But the digitisation of SSH research has brought about new needs.**

*Peter Doorn*

Research infrastructures have become mainstream in the natural and life sciences: particle accelerators for physicists, telescopes for astronomers, nuclear icebreakers for polar researchers. Since the SSH came to rely heavily on information technology, however, the 'virtual laboratories'

for these disciplines have become part of the same league as the real-world labs in chemistry and physics: digital places where large numbers of researchers share their data and tools.

### No 'one size fits all'

An important factor in this transition was the formulation of a Roadmap by the European Strategy Forum on Research Infrastructures (ESFRI) in 2006; for the first time, this Roadmap included proposals with regard to the SSH. In the Netherlands, the 2005 NWO BIG programme and the first Netherlands Roadmap for Large-Scale Research Facilities (2008)

also embodied a breakthrough by including SSH proposals. Why are one or two research infrastructures not enough for the SSH? The silliness of this question becomes apparent as soon as we turn it around: why not have one research infrastructure for the natural sciences? You cannot look at the stars with a nuclear icebreaker and you cannot break ice with a telescope. The SSH are just as heterogeneous as the natural and life sciences, and therefore one tool or virtual lab does not fit all demands.

### Jungle of acronyms

Still, the SSH were among the first

to understand the logic of collaboration. In the Netherlands the two European-wide infrastructures CLARIN and DARIAH decided to join forces and they even merged their acronyms to form CLARIAH. The same goes for SHARE and LISS/MESS in the social sciences. Even collaborations embracing the entire SSH domain have emerged, such as DASISH, which is there to solve common challenges.

In the brave new world of research infrastructures the reader may get lost in a jungle of acronyms, like ESS, GGP and CESSDA. This special edition of E-data & Research aims to make them a bit more familiar.



# Online gateway to language resources

**Interviews, novels, newspapers, speech recordings: language resources play an important role in the humanities and social sciences. CLARIN makes these data available to scholars.**

*Jan Odijk*

CLARIN is a European initiative to build an infrastructure for social sciences and humanities researchers who make use of language resources. In the Netherlands, the focus of the national CLARIN project is on humanities researchers working with textual resources. CLARIN-NL will offer scholars the tools to allow computer-aided language processing with a view to addressing one or more of the multiple roles language plays in the humanities. Examples of these roles are ‘carrier of cultural content and knowledge’, ‘instrument of communication’, ‘a component of identity’ and ‘an object of study’.

## Ensuring accessibility

A key aspect of the infrastructure is that both the resources and the tools to work with them are easy to find and accessible. Moreover, the infrastructure should be tailored to the average humanities scholar, meaning that the required ICT knowledge to use the tools and resources should be moderate. To achieve this goal, resources and tools are being standardised so that they can be used together seamlessly; the tools are equipped

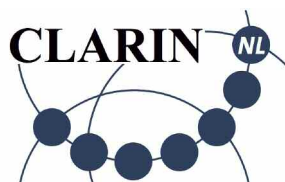


*Jan Odijk, Programme Director, and Arjan van Hessen, member Executive Board: “The infrastructure is tailored to the average humanities scholar” photo Inge Angevaere*

with user-friendly interfaces. In a range of demonstration and curation projects, humanities scholars, ICT experts and data providers collaborate to make resources available in the CLARIN infrastructure and to show the potential of certain technologies or curated data. Curation implies standardisation of resources in accordance with CLARIN require-

ments, addition of CMDI metadata, ensuring findability and accessibility and making provisions for long-term storage.

[clarin.nl](http://clarin.nl)



## MIMORE A microcomparative morphosyntactic research tool

The MIMORE tool enables researchers to investigate morphosyntactic variation in Dutch dialects by searching three related databases (DynaSAND, DiDDD and GTRP) using a common online search engine. The search results can be visualised on geographic maps and exported for statistical analysis.

With the MIMORE search engine these three databases can be searched simultaneously, using text strings, part of speech tags and syntactic variables. The researchers can combine categories and features into complex tags or use predefined tags. All categories and features are based on the ISOCAT standards. Since all sentences have a location code, the morphosyntactic phenomena found in a set of sentences resulting from a search can be automatically plotted on a geographic map. It is possible to include more than one morphosyntactic phenomenon in one map, thus visualising potential correlations between these phenomena. Also included is a user-friendly export function for external data use, e.g. in a statistical application. *Sjef Barbiers*

ritual and religious dynamics for a long time. In the PilNar project a corpus of modern pilgrimage narratives is constructed. It consists of Dutch texts written after ca. 2000 that present the thoughts and impressions of pilgrims to Santiago de Compostela.

The pilgrimage to Santiago is used as an example of current ritual and religious dynamics. This source has hardly, if ever, been used for contemporary research in the cultural sciences. Previous exploratory research has made it clear that the corpus of stories intended here is an excellent source for research into the profile (or, better, profiles) of the modern pilgrim. *Paul Post*

## D-Lucea Database of the longitudinal utrecht collection of english accents

At University College Utrecht (UCU), students and staff speak a wide variety of native languages, but they all use English as the lingua franca on campus. How will the English accent evolve over time, the accent of English native speakers and of native speakers of other languages (Dutch, Italian, Spanish, etc.)?

To answer this question an existing database of speech recordings of L1 and L2 speakers of English is being curated. The recorded speakers are students from the UCU community. These students are being recorded longitudinally throughout their 3-year period on campus, using read and spontaneous speech in their L1 and in L2 English (or in L1 English only). The resulting database is of interest for research and development in linguistics, language education, speech technology, and sociophonetics. *Hugo Quené*

## PilNar Pilgrim narratives

Churches are being closed and religion is moving to the margin. Paradoxically, we see religion and ritual flourishing and emerging in Europe. One example that stands out here is pilgrimage: the pilgrimage to Santiago de Compostela in particular has become unprecedentedly popular. Pilgrimage narratives, especially travel accounts, have been used as a favourite source for research into

## COLUMN

It will take our standard university server (2 Quad cores, 8 GB RAM, Linux) 175 years to process the daily batch of over 2 million news articles stored by information brokers such as Lexis Nedis. My research group at the Faculty of Arts of VU University Amsterdam is building systems for ‘deep reading’ of natural-language text to relate today’s news to news processed in the past. What happened where and when, and who was involved? But also: who is the source, what is the opinion expressed, is it a factual statement, a denial or speculation? Ideally, our cascade of more than 15 natural-language processing modules should be able to process this batch before the next day’s batch arrives.

That’s why we use the HPC cloud of the SURFSara infrastructure to have batches processed in parallel by Virtual Machines (VMs) on which the full range of modules has been installed. This makes it easier to deploy more VMs if needed – as long as the infrastructure allows it. Using this configuration, we recently



## 175 years of processing - every day

processed 66,000 news articles in one week. We are experimenting with parallelisation of the process and optimisation of the infrastructure at SURFSara. The system also includes a Knowledge Store (KS) for sources and processing outcomes. The KS is installed on Hadoop and Hbase and includes a triple store. For storing the Terabytes of results, we use specific storage units at SURFSara.

Our group is also heavily involved in making research results available in such a way that they can be replicated and reproduced by others. To this end, we use version-control systems

such as Github, websites with releases, documentation of modules, processes and data formats, descriptions of experiments and tutorials. We try to standardise the systems and the formats against widely used practices, while at the same time developing our own standards for new types of information.

Eventually, we hope to demonstrate that we can handle the news streams in different languages and provide a proper scientific platform both for developing natural language processing modules and for creating data structures for

researchers to explore news streams as datasets. The latter will shed light on how many changes in the world are actually reported, how much duplication there is across sources, how much they agree or disagree about the information provided, what opinions and perspectives are provided. Our data structures should provide valuable information and knowledge about the history of the changing world as provided by a wide range of media sources. In the future, we will expand this range beyond written sources to include various multimedia sources, among them structured databases, sensors, audio-visual data, and images.

*Piek Vossen*

*photo Riechelle van der Valk*



## Kroes: ‘Challenges ahead’

CONTINUED FROM PAGE 1

policies are geared towards a knowledge-based society. I am also very pleased with the policy on eScience and the creation of an eScience centre that serves a number of application areas. Finally, data archiving at DANS is an important service focusing on the social sciences and humanities in particular, which are areas with a strong

potential and acute needs for future data storage.”

The natural sciences, social sciences and humanities have developed many different research infrastructures in the course of time. Will they all stand the test of time?

“There are many challenges ahead. The best research infrastructures support researcher collaboration in virtual research communities where

knowledge sharing between the best brains is combined with open access to research results and state-of-the-art computing systems to increase the efficiency and creativity of research in Europe. This requires a strong focus on e-infrastructure (i.e. digital infrastructures such as data and computing systems, communication networks and virtual research environments) to ensure solutions for global high-speed networking, access to research data and instruments wherever they

are, sufficient computing capabilities and software that responds to the needs of tomorrow’s researchers.” Kroes sees opportunities for the SSH when big data tools and research methods are put to efficient use. “Open research data could help combine and share works of different research groups, thereby creating new collaborations and tackling new issues for solving common challenges. Furthermore, social sciences and humanities have an enormous new research potential

in the form of data created and emerging from online social networks and social media activities. It would therefore be very important to include the topics of curation methods and data collection & analysis in SSH curricula, so as to help future researchers and scholars benefit from such new resources. I would advise promoting and supporting the inclusion of both topics; this could lead the way to new multidisciplinary approaches.”

**The Dutch projects CLARIN-NL and DARIAH-NL have merged and evolved into a new infrastructure for data and tools in the humanities: CLARIAH.** *Patricia Alkhoven*

Libraries, museums and archives conserve and manage large collections of data (texts, images, structured data, audio-visual materials) which are gradually being made available in digital form for re-use. Until now there has not been an integrated approach to dealing with digital data and tools for the humanities; existing datasets are not interconnected, and researchers and students experience a lack of training in applying digital methods to analyse large datasets. CLARIAH aims to deal with these issues and find integrating solutions. The CLARIAH infrastructure will give researchers access to large collections of digital data and innovative, user-friendly applications for the processing of these data. Both data and tools will be managed in a sustainable way, i.e. they will be easily accessible to humanity scholars such as linguists, historians and media scientists.

CLARIAH is the natural continuation and extension of two national projects: CLARIN-NL (linguistics, texts) and DARIAH-NL (socio-economic history, structured data), complemented with a third pillar: Media Studies (audio-visual data). CLARIAH combines the European CLARIN and DARIAH infrastructures in the Netherlands.

### Beyond topic and time

Until recently, addressing questions about culture and identity depended on experts’ ability to identify potentially relevant pieces of information in archives, libraries and museums.

An integrated approach for data and tools in the humanities

# Searching beyond the borders of topic and time



*Baroque Library Prague photo Bango*

Because such research was extremely time consuming, it was hardly possible to look at all the data or test alternative analyses. In the CLARIAH infrastructure digital data will be made accessible to a large number of researchers, who will be able to process them using digital tools. They will be well equipped to verify data selections and other researchers’ interpretations. CLARIAH will open up their world by enabling them to search beyond the boundaries of topic and time, to tackle the vast universe of data and study the development of culture, cultural identities and ideas.

On 1 October 2013, the CLARIAH proposal for a

‘National Roadmap for Large-Scale Research Facilities’ was submitted to the Netherlands Organisation for Scientific Research (NWO). A decision about the project proposal is expected by June 2014.

If funding is awarded, the project will be launched on 1 January 2015. The five demo projects described here show the current and future possibilities of CLARIAH.

[clariah.nl](http://clariah.nl)



### TROVe Transmedia Observaory

Different kinds of media, such as newspapers, TV broadcasts, blogs and other online social media, are getting more and more entangled. It therefore makes sense to use all news sources available if one is interested in studying the diffusion of news items across media channels and over time. TROVe has built a search engine that allows scholars from the humanities to analyse multi-media content over time. Specific roles of the various media, their interrelationships and their mutual influences can be examined in one interface. TROVe not only allows the identification of key players, but also analyses the roles they play in discursive constructions. Eventually it will be able to show how the various media and actors (people, organisations) interact, in other words: how they influence public debate. [clariah.nl/en/trove/summary](http://clariah.nl/en/trove/summary)

*Johan Oomen*

### OHT Oral History Today

Oral History collections are a rich basis for various types of scholarly research. However, scholars need tools to explore these collections in a way that is suitable for answering research questions. In the first phase of research, scholars need to select relevant collections. After that, they will want to analyse the selected materials thoroughly. The tool developed in OHT aids the scholar during the selection phase by means of keyword search, content visualisation using word clouds, and various filter options based on information available in the metadata, such as time periods and location. For the analysis phase, the OHT tool provides features like a personal folder to store interesting segments selected using a virtual cutter. The segments can be annotated and shared with colleagues. The tool also recommends related interviews or interview segments in the collection. [oht.staging.dispectu.com/#](http://oht.staging.dispectu.com/#)

*Roeland Ordelman*

### HLZ (HSN Links Zeeland)

Microdata on ‘changes in life courses’ provide a unique insight in how societies change over time. The Historical Sample of the Netherlands (HSN) collects this kind of data for the Netherlands and is largely based on municipal registers. However, some registers are incomplete, do not provide information on stillbirths and sometimes provide conflicting information. LINKS (Linking System for Historical Family Reconstruction) is based on civil certificates and aims at the reconstruction of all nineteenth and early twentieth century families in the Netherlands.

HLZ demonstrates the value of combining differently structured datasets by extending and enhancing the HSN database with data from the LINKS database for the province of Zeeland.

[iisg.nl/hsn/index.html](http://iisg.nl/hsn/index.html)

*Kees Mandemakers*

### CLIO-DAP

The purpose of CLIO-DAP is to improve communication in research by asking leading scientific journals in the field of economic and social history to make available the research data associated with articles they publish. The software should facilitate the process where data are deposited in reliable digital data archives; ideally, these data collections are also published and reviewed, and subsequently linked to the corresponding publications.

CLIO-DAP has built a demonstrator service for CLARIAH with enhanced publications to showcase existing output from socio-economic history journals with a Data Availability Policy (DAP) and to illustrate the workflows needed. It thus contributes to convincing stakeholders in other journals to adopt similar policies.

[cliodap.dans.knaw.nl](http://cliodap.dans.knaw.nl)

*Leen Breurs*

### Nederlab

Nederlab’s goal is to enable scholars in the humanities to find answers to new, longitudinal research questions. For this purpose Nederlab aims at setting up a user-friendly tool-enriched web interface, allowing researchers to simultaneously search, analyse, and enrich the digital historical texts made available by scientific libraries and institutes, at the text and metadata levels. The CLARIAH demonstrator is a working Nederlab prototype. It consists of a website offering a Virtual Research Environment for diachronic research, with a transparent and efficient user interface that is designed to meet the specific needs of end users. In this respect, Nederlab differs from existing user interfaces such as those currently offered by CLARIN, which mainly focus on the supply of tools and data.

[www.nederlab.nl](http://www.nederlab.nl)

*Nicoline van der Sijs*



# Great opportunities for comparative research: the European Social Survey

Despite ever closer European integration, EU countries show persistent and remarkable differences in a variety of fields. The European Social Survey (ESS) covers a wide range of topics that tap into key challenges facing Europe today. Gerbert Kraaykamp and Kees Aarts



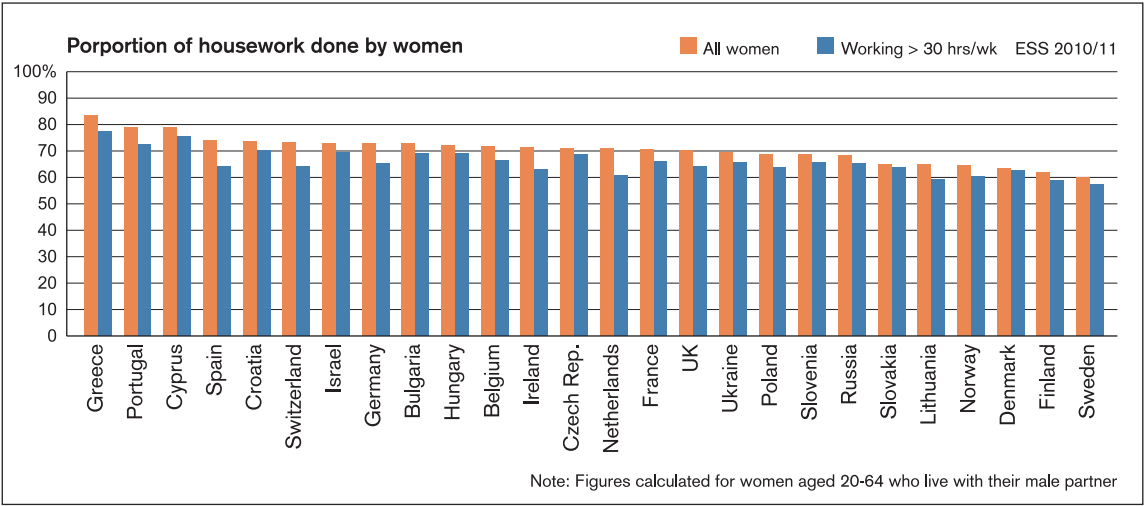
Women in full-time paid work are responsible, on average, for around two-thirds of the total time heterosexual couples spend on housework photo Everett Collection

The ESS studies a broad range of attitudes and behaviours in a changing Europe in a two-yearly sequence. Topics addressed in the main questionnaire of the ESS vary from moral opinions, health issues, trust and social capital to all kinds of aspects of a person's socio-economic position and household circumstances. Other topics change from round to round, enabling the ESS to cover a wide thematic range and adapt to changing demands. The ESS is widely used by academics, researchers, politicians, policymakers and journalists interested in patterns in public opinions and behaviours which vary over time and across

countries. The ESS data, collected in more than 30 countries in six rounds since 2002, thus may be used to shed light on issues of comparability and/or differentiation. For instance, significant differences in political engagement and social attitudes persist between Eastern and Western Europe, underpinned by differences in history, culture, institutional and legal conditions. There is also evidence of a growing economic and political divide between Northern and Southern Europe, fuelled by different experiences with the Eurozone crisis. Over the past decade, the ESS has

become an authoritative source of reliable data about Europe's evolving social, political and moral condition with high standards of rigour in cross-national research with respect to sampling, data collection, reduction of bias, and reliability of questions. europeansocialsurvey.org

This article is based on the ESS report 'Exploring public attitudes, informing public policy. Selected findings from the first five rounds', 2013.



## Removing barriers to facilitate cross-border research

**A wider use of data leads to an improved understanding of societal problems.**

Whether they are researching, say, family changes in modern society, social effects of primary schools or attitudes towards data, social scientists want to use and compare data from as many different European countries as possible. CESSDA could play an important role in acquiring these data.

**Cross-border access**

The main goals of CESSDA are to facilitate and promote the wider use of data in social, economic and political research and, by doing so, improve our understanding of ongoing societal problems as well as available solutions. CESSDA aims to achieve these objectives through the development and coordination of standards, protocols and professional best practices in data management and technical infrastructure. CESSDA will provide relevant training programmes, which are also explicitly aimed at countries

where national infrastructures still have to be set up from scratch. One of CESSDA's key objectives is to facilitate access to data resources regardless of the location of either the researcher or the data, in Europe or elsewhere. A real challenge for international research today is cross-border access to these resources. CESSDA intends to break down the barriers created by specific national requirements. The EU-funded Data without Boundaries project, in which CESSDA is involved, has already started work on this. CESSDA has a solid base as it contains the oldest digital archives of Europe. It was founded in 1976 as the Council of European Social Science Data Archives. As a European Research Infrastructure Consortium (ERIC), CESSDA has now become a permanent legal pan-European research infrastructure in 2013. Heiko Tjalsma cessda.net



## COLUMN

### Personal preferences matter, if culture allows

Why do some women work full-time, when others work only part-time and some not at all? The number of hours that women work is affected by their personal preferences on the one hand, and by national culture on the other. In a recent study conducted with colleagues at the University of Twente, we have analysed this interplay between individual work preferences and social conditions. We focused on the extent of gender egalitarianism in 24 European countries. The European Social Survey (round 2, 2004) provided suitable data for testing existing theories about the influence of work preferences on women's decisions concerning work and family. Using the European Social Survey turned out to be a prudent choice. First of all, round 2 of the European Social Survey comprises a dedicated set of questions on family, work and wellbeing. With these questions we could measure not only individual work preferences, but also the extent of gender egalitarianism in the countries.

Secondly, the strong methodological basis of the European Social Survey ensures cross-national comparability of the data, an issue that makes the ESS the most important survey of its kind. Thirdly, the European Social Survey provides important information on the respondents' socio-economic characteristics: indispensable information for scholars aiming to understand social phenomena in their national context.

The main conclusion of our study is that the number of hours worked is neither exclusively determined by women's personal work preferences nor exclusively determined by a country's culture

of gender egalitarianism. It is the interplay of personal preferences and prevailing cultural values that seems to matter: the individual preferences appear to matter most in the European countries, where the populations at large share more gender-egalitarian attitudes. The European goal of increasing women's labour-market participation thus cannot be achieved by just 'activating women'. Indeed, the population at large should also embrace values of gender equality – a challenge for the policy-makers of the 21st century. Our dive into the European Social Survey proved to be rewarding.

Andringa, W., Nieuwenhuis, R. and Gerven, M. van (to appear). Women's employment: The interplay between individual work preferences and country liberalism in 24 European countries.

Minna van Gerven photo Charles Kuiper





The social sciences are becoming increasingly multi-disciplinary as new technology opens up vast new opportunities for measuring and linking data from very different fields. SHARE-MESS will position Dutch social scientists at the forefront of their field and form the hub of an international network of advanced data-collection efforts. *Marcel Das*

Technology is changing society itself, particularly the way we communicate and interact. Simultaneously, Western societies are ageing fast, with many inherent economic, social and medical challenges. The Roadmap projects SHARE and MESS make it possible to study such significant changes in society using state-of-the-art methods. Both SHARE and MESS provide a large-scale social science facility, each of a different kind. SHARE collects similar data in 20 countries. Researchers can use them to study the impact of different institutions, policies and policy reforms on health, employment, and wellbeing in different countries. In MESS, researchers from around the world can conduct their own experiments using the LISS panel or immigrant panel and test new measurement devices. The two approaches are now combined in a new large-scale social science facility: SHARE-MESS.

### Attractive and cost efficient

The merged facility will build on the results that these two facilities have achieved separately and capitalise on the synergies between them to create a set of state-of-the-art resources for cutting-edge multi-disciplinary work in the social sciences. The SHARE-MESS team will:

- collect longitudinal data for substantive research in the social sciences;
- develop, test, and deploy technological innovations in survey tools and measurement techniques;
- conduct experiments;
- incorporate multi-mode interviewing into the SHARE-MESS facility;
- link survey data to various forms of administrative data;
- increase linkages with other population surveys.

The first mission of SHARE-MESS is to develop an infrastructure that exploits new technology, is attractive and cost efficient for collecting survey data in general, and will accommodate numerous major data-collection efforts in the future. Here, the framework is essentially survey methodology. A second mission of SHARE-MESS will be to collect rich longitudinal and experimental data so that researchers may better understand individual and household decision-making, and how decisions and circumstances affect many aspects of life.



*MESS uses innovative ways of data collection, for instance accelerometers to objectively measure physical activity photo Dirima*

## SHARE and MESS join forces in social science facility

# Data collection to a new level

### MESS Boosting research in the social sciences and related disciplines

**Progress in the social sciences relies on high-quality data. MESS provides researchers with an active laboratory and forefront environment for collecting data and conducting innovative experiments. The facility is open to the global scientific community, free of charge.**

The Advanced Multidisciplinary Facility for Measurement and Experimentation in the Social Sciences (MESS) is a large-scale research infrastructure open to academic researchers worldwide. The project is designed to stimulate and integrate research in the social sciences, life sciences and behavioural sciences. MESS is run by CentERdata (Tilburg University). MESS, a Dutch initiative, is embedded in an international research network that includes the American Life Panel, a similar facility in the US, and comparable enterprises in France and Germany. The central resource in the facility is the

Longitudinal Internet Studies for the Social sciences (LISS) panel. This panel of about 5,000 households is representative of the Dutch-speaking population in the Netherlands aged 16 and over. The recruitment of households is based on a probability sample drawn by Statistics Netherlands from municipal registers. Each month, panel members answer questions online for approximately 30 minutes. Another element of the MESS project is the additional special immigrant panel. The immigrant panel is comprised of approximately 2,400 individuals, 1,700 of which are of non-Dutch origin. The sample was also drawn by Statistics Netherlands from the municipal registers, stratified by country of origin.

Researchers can propose new modules, free of charge, ranging from a few questions to longer questionnaires or tailored experiments.

All data collected in MESS are published online and are freely available to academic researchers.

*lissdata.nl*

*Marcel Das*



### SHARE An indispensable tool for research on ageing issues

**SHARE is part of a global scientific effort involving researchers from around the world working in various disciplines. The synchronised cross-national SHARE data have led to new insights into how public policy affects issues like health, poverty and social exclusion among the older part of the population.**

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a large-scale, longitudinal survey of socio-economic status and health among people over 50 years of age in 20 countries, including the Netherlands. The survey covers key areas of life, including household demographics, physical and mental health, health behaviour, socio-economic status, and family and social networks. The survey data include data from physical tests which the respondents complete during the interviews, such as grip strength and lung force.

Recently, the first experiments to collect dried blood spots using a finger prick have started to facilitate more precise disease risk assessment.

SHARE provides an indispensable tool for research on ageing-related issues. It aims at increasing knowledge on the sustainability of pension systems, the labour market for older workers, preparation for retirement and savings behaviour, mental and physical health in relation to socio-economic status, and healthcare use. Today, SHARE is part of a consortium of harmo-nised ageing studies in many countries, including the US, Mexico, Japan, China, South Korea and India. This makes SHARE part of a global scientific effort involving researchers from around the world working in diverse disciplines. In March 2011, the EU selected SHARE as the first ever European Research Infrastructure Consortium (ERIC), with a legal status that gives it the advantages of a major international organisation.

*share-project.nl*

*Marcel Das*



## Innovative forms of data collection

**MESS facilitates a wide range of experimental approaches to link social sciences to life sciences.**

### Accelerometers

Self-reports on physical activities are usually limited to certain aspects of daily activity such as structured exercise or walking. Daily activities and sedentary behaviour are usually missing. In addition, responses to questionnaires may not be accurate because of the cognitive challenge of estimating frequency and duration of activities, and social desirability bias. The development of accelerometers has opened up new possibilities for studying all intensity levels of physical activity, from completely sedentary to vigorous activity, over periods of several days.

### Bathroom scales

Advanced bathroom scales are used to objectively measure the weights and fat percentages of LISS panel respondents. The scales have a wireless internet connection, which immediately transmits all measurements to the database without the need for respondents to report anything. The accurate and steady flow of measurements allows the monitoring of fluctuations in weight and fat percentage over time and in addition provides a unique insight in the reliability of self-reported measures.

### Smartphones

Smartphones make it easy to keep a diary of activities multiple times a day simply because people tend to carry their smartphone with them all day. Another advantage of using smartphones for time use data collection is that it enables the collection of additional information on the respondents' reporting behaviour. These so-called 'para-data' include, for example, time of reporting, correction of activities, or the effect of sending reminders to respondents. The time use diary is also combined with GPS registrations of the locations where the reported activities took place, and questions about the mood or emotions of the respondent during the day.

Mobile devices allow researchers to gain insight into the movements of respondents throughout their daily routines. A mobile application is used in the LISS panel to automatically detect detailed trip information of a group of respondents by making use of different location measures, such as GPS, cell tower triangulation and WiFi geo-location. The collected information includes the starting and end points, modes of transportation and in some cases even the purpose of the trip.



DARIAH connects Europe's resources, tools and scholars

# Advancing a digital revolution in the arts and humanities

**DARIAH, the Digital Research Infrastructure for the Arts and Humanities, is committed to advancing the digital revolution in the arts and humanities across Europe by connecting resources and tools with the emerging next generation of digital scholars.**

*Laurent Romary and Sally Chambers*

As more sources become digital, more digital content is being produced and more digital tools are being deployed, we see a next generation of digital scholars in the humanities emerge. DARIAH aims to connect these resources, tools and scholars, ensuring that the state of the art in research is sustained and integrated across European countries.

## Collaborations

DARIAH works closely with research communities in an expanding network of affiliated projects, for example with archaeologists (via ARIADNE, the Advanced Research Infrastructure for Archaeological Data Networking in Europe), medieval and modern historians (via CENDARI, the Collaborative European Digital Archive Infrastructure) and Holocaust researchers (via EHRI, the European Holocaust Research Infrastructure). This collaboration fosters the production of digital reference materials of all kinds (archival descriptions, transcriptions, annotated images, musical simulations, 3D models, etc.) and makes it essential to offer a data platform combining generic and specific hosting facilities. Although some digital humanities projects such as the TEI (Text Encoding Initiative) have already achieved high levels of maturity and reach, many new projects are funded where scholars need to collaborate and require help and support. It is important to remember that the wide majority of scholars has little or no background in digital methods. The vision for DARIAH is therefore to offer a portfolio of infrastructure-orientated activities centred around research communities across the broad spectrum of the arts and humanities, no matter how far advanced they are in the digital revolution.

[dariah.eu](http://dariah.eu)



Watercolour showing perspective of main elevation of Craighend House by David Bryce credits RCAHMS

## ARIADNE Excavating the past, building the future

There is a large availability of archaeological digital datasets spanning many periods, domains and regions. However, due to access barriers and non-homogenous perspectives, their potential has been constrained. ARIADNE brings together and integrates existing archaeological research data infrastructures to allow researchers to use the various distributed datasets as well as new and powerful technologies in archaeological research. ARIADNE aims to enable researchers to access data centres, tools and guidance from various countries; intends to create new web-based services based on common interfaces to data repositories; and will promote the availability of reference datasets and the use of innovative technologies. ARIADNE is set to contribute to the creation of a new community of researchers ready to exploit the contribution of Information Technology and to incorporate it in the body of established archaeological research methodology.

[ariadne-infrastructure.eu](http://ariadne-infrastructure.eu)  
*Marieke Polhout and Valentijn Gilissen*

## NeDiMAH Network for Digital Methods in the Humanities

NeDiMAH is one of the so-called Research Networking Programmes of the European Science Foundation (ESF). Sixteen European countries participate in the initiative, which is scheduled to run for four years until May 2015. NeDiMAH is responsible for a series of activities and networking events that will allow the examination of the practice of, and evidence for, digital research in the arts and humanities across Europe. It is promoting collaboration and networking among the community of European scholars in this field as well as those engaged in creating and curating digital collections of scientific and cultural heritage. Five working groups have been set up in which scholars can exchange and share insights and results. NeDiMAH will enable arts and humanities researchers to develop, refine and share research methods that allow them to create and deploy digital methods, collections and infrastructures. NeDiMAH collaborates closely with the DARIAH and CLARIN e-research infrastructure projects and other national and international initiatives. All NeDiMAH activities are open to the entire European community of scholars.

[nedimah.eu](http://nedimah.eu)  
*René van Horik*

## CARARE Bringing content for archaeology and historic buildings to Europeana users

CARARE is a Best Practice Network, funded under the European Commission's ICT Policy Support Programme. It is designed to involve and support Europe's network of heritage agencies and organisations, archaeological museums, research institutes and specialist digital archives in the following activities:

- opening up – through the cultural network of Europeana – the digital content on archaeological and architectural heritage which they hold;
- aggregating content and delivering services; and
- enabling access to 3D and Virtual Reality content through Europeana.

CARARE is one of a suite of projects aimed at further developing the Europeana cultural network. It will play an important role in involving Europe's network of organisations responsible for investigating, protecting and promoting unique archaeological monuments, architecturally important buildings, historic town centres and industrial monuments of World, European and national heritage importance alongside the existing national, regional and local content providers.

[carare.eu/eng](http://carare.eu/eng)  
*Marieke Polhout and Valentijn Gilissen*

## EHRI Digital infrastructure as well as human network

The European Holocaust Research Infrastructure (EHRI) aims to support the Holocaust research community by providing –online – access to dispersed sources relating to the Holocaust and by encouraging collaborative research. EHRI helps researchers overcome one of the major challenges facing Holocaust research: the fact that the research materials are dispersed across Europe. Beyond this unique attempt to bring information about collections together, EHRI activities range from promoting new research methodologies to conducting analyses of research needs, e-Science standards and technologies, and research support. EHRI's fellowship programme enables researchers to work on-site in a number of leading institutes. EHRI organises summer courses on Holocaust research and promotes a European-wide approach. An important condition for EHRI to succeed is its interdisciplinary structure. The collaboration between historians, archivists and e-infrastructure specialists is a key contributing factor to EHRI's success.

[ehri-project.eu](http://ehri-project.eu)  
*Petra Drenth*



**Family relationships have changed a lot over the past few decades. The Generations and Gender Programme (GGP) is intended to answer the questions that come with these changes.**

*Tom Emery and Anne Gauthier*

Today's families differ considerably from those of the 1950s, where the male breadwinner and his housewife were the norm. Families have become less stable, more complex and highly diversified. The rapid ageing of European populations has also contributed to this fast change, while new types of families have emerged alongside new relationships between generations and genders. Understanding these changes will help us meet many of the challenges that societies face today, such as: How do we support and care for the elderly? How is disadvantage inherited? Why are women having fewer children?

The GGP was launched in 2001 and now covers 19 advanced industrialized countries. It improves our understanding of how various factors affect family life by collecting high quality individual-level survey data on topics such as partnership formation & dissolution, fertility and intergenerational solidarity. Respondents are interviewed every 3 years and changes in their family life are recorded. Importantly, the GGP covers the entire lifespan of adulthood between the ages of 18 and 79 and is the only dataset dedicated to the longitudinal and cross-national study of family life and generational relationships from early adulthood to old age. Over time, the GGP follows respondents through relationships, marriages, parenthood, divorces, deaths of loved ones and many of the trials and tribulations that people meet with, tracking the impact and consequences of these life events at the individual and the societal level. These survey data are complemented with regional and national indicators through a contextual database which helps us understand the part that policy and other contextual factors play in family life.

The Netherlands is a front-runner in this research area. Not only has it been involved in the GGP since its inception, it has also hosted the project since 2009. Last year has been a particularly busy year for the GGP in the Netherlands as it saw the collection of Wave 4 of the Dutch GGP involving a new collaboration with CBS (Statistics Netherlands). Using the Wave 4 dataset, researchers can now look back over a turbulent decade and see its impact upon Dutch family life as well as look forward to monitoring the impact and implications of this change in future waves of the GGP.

### Building a nest during a financial crisis

Recent evidence from the GGP shows that starting a family is strongly intertwined with housing quality and home-ownership. Typically, couples will only start a family if they can live together and they will only start to live together if there is good and affordable housing available. But after many decades of increasing average housing quality and levels of home-ownership in



*Grandparents helping out can be important for women who want to return to work after having a child photo Marcel Jancovic*

#### GGP Topics

- Fertility
- Partnership
- Transition to adulthood
- Work-family balance
- Gender relations
- Intergenerational exchanges
- Informal and formal care
- Wellbeing and health
- Grandparenthood
- Economic activity
- Retirement

**GGP follows respondents through trials and tribulations**

# Understanding society through our relationships

Europe things have turned around recently with the housing-market crisis in connection with the global financial crisis. It can now be very difficult for young couples to get a mortgage and start climbing the housing ladder. Given this, future research aims to understand housing-market fluctuations, including the recent crisis, as well as the spatial differentiations in housing opportunities and their interplay with household formation and child-bearing. For example, do couples delay having children because they can't afford to buy a home? Such questions are currently under investigation using GGP data.

### Who has kids outside of marriage?

Recent analysis of GGP data supports the long held belief that having a first child outside marriage is more prevalent among those with lower levels of education. But it also reveals that in some countries, such as France, this is no longer true and it is in fact

those with higher education who are more likely to have a child outside of marriage. Whether or not this pattern will spread to other countries, and whether or not it is influenced by the legal and institutional environment of families are key questions that researchers are looking to answer with the GGP. In answering questions like this we can start to understand what marriage and parenthood mean in the 21st century. Why do we get married? Why do we have kids? And what have the two got to do with each other?

### Care and support across generations

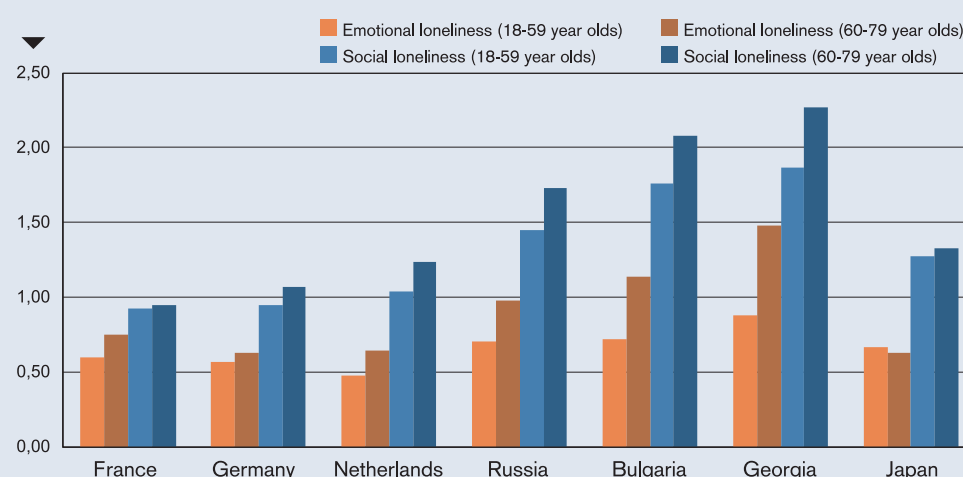
The GGP has led to several key findings about the way generations support each other throughout life. For example, the GGP was used to show how loneliness is more prevalent in Eastern Europe than in Western Europe. This is attributable to the greater health and wealth of older generations in Western Europe and the extent to which this

helps them combat loneliness. But the GGP has also revealed that older generations are not merely vulnerable but also play an important support role for younger generations. GGP data were used to show that in some countries, such as Hungary, grandparents helping out with childcare are important for young women who want to return to work after having a child. Yet in other countries, like the Netherlands, the availability of grandparents for childcare does not affect a mother's decision to work. Future research is set to examine whether this is because of culture, policy or some other factors. These are just two of the ways in which the GGP has demonstrated the complexity and diversity of relationships between generations as well as the need to consider this diversity in different countries.

[ggp-i.org](http://ggp-i.org)



De Jong Gierveld short scales for emotional and social loneliness (0-3)



*Loneliness on the De Jong Gierveld Short Scale across 7 Countries. (De Jong Gierveld, J., & Van Tilburg, T. (2010). The De Jong Gierveld Short Scales for emotional and social loneliness: tested on data from 7 countries in the UN generations and gender surveys. European Journal of Ageing, 7(2), 121-130.)*



**DASISH****Data Service Infrastructure for the Social Science and Humanities**

DASISH is an umbrella project coordinating activities for all of the five existing European infrastructures in the social sciences and humanities (CLARIN, DARIAH, CESSDA, ESS and SHARE). Its goals are to identify possible synergies in infrastructure development and to work on some concrete joint activities. These activities involve improving data quality, data archiving, data enrichment and shared data access, as well as legal and ethical issues. Dutch project partners are CentERdata, based at Tilburg University, and DANS.

[dasish.eu](http://dasish.eu)

**EUDAT****European Data Infrastructure**

The EUDAT project has some similarities with DASISH, as it brings together data service providers and aims at developing common standards and services. EUDAT is, however, directed towards a user community which is active in other areas. The research communities involved in EUDAT are EPOS (earth sciences), ENES (climate sciences), LIFEWATCH (environmental sciences), INCF (neuroinformatics) and CLARIN (linguistics). The latter is also part of the DASISH initiative. The EUDAT consortium comprises 26 European partners from 13 countries. The Netherlands is represented by SURFSARA. The development of common data services is organised by establishing common views on a number of topics. These are (1) support for real-time Dynamic Data, (2) Workflow Support to orchestrate data processing chains, (3) Semantics to check correctness of incoming data against trusted ontologies, and (4) Data Access and Re-use at community and service provider levels.

[eudat.eu](http://eudat.eu)

**APARSEN****Alliance for Permanent Access to the Records of Science Network**

The aim of the APARSEN project is to establish a Virtual Centre of Excellence (VCoE) on digital preservation. The 31 project partners, operating in the cultural heritage sector, the research data community and the corporate world, share their expertise in the field of digital preservation on four topics: 'trust', 'sustainability', 'accessibility' and 'usability'. The outreach activities, such as consultancy, training and tools, will form the basis of the VCoE. The Dutch partners in the APARSEN project are Data



*Collaborations are the best way to initiate, improve and extend research infrastructures picture Alphaspirt*

## Common standards, procedures and tools to improve research facilities

# Joining forces for the future of research infrastructures

**The best way to initiate, improve and extend research infrastructures is by means of collaboration among people and organisations.**

**In several projects, partners collaborate on a wide range of subjects related to research infrastructures, such as standards, procedures and tools.**

*René van Horik*

Archiving and Networked Services (DANS) and the National Library of the Netherlands (KB).

[aparsen.eu](http://aparsen.eu)

**Europeana Cloud**

Europeana Cloud is a best practice network related to the Europeana service, which provides access to millions of European cultural heritage objects. The aim of the project is to establish a cloud-based solution for the Europeana system and its aggregators. Europeana Cloud will provide new content, new metadata, a new linked storage system, new tools and services for researchers and a new platform: Europeana Research. The 'Researcher Needs' work package is dedicated to the identification and assessment of the needs of researchers in the social sciences and humanities. Among the activities carried out within the

framework of this work package is the execution of a web survey to collect information on digital research practices, tools and content, focusing on the potential use of content from Europeana. The project has 35 partners of which six are Dutch or Netherlands-based. [pro.europeana.eu/web/europeana-cloud](http://pro.europeana.eu/web/europeana-cloud)

**OpenAIREplus**

This project deals with promoting an Open Access infrastructure for research in all scientific disciplines in Europe. Providing Open Access to research – both research papers and underlying datasets – not only benefits the general public, it is also good for researchers: several studies indicate that openness increases citations. Openness also improves reproducibility of research results – and it might introduce new and perhaps

unexpected audiences to existing work. The project, with 33 participating partners, implements, supports and monitors the European Commission's Open Access pilot. Guidelines for the registration and validation of repositories containing Open Access publications are being compiled by the project. [www.openaire.eu](http://www.openaire.eu)

**4C****Collaboration to Clarify the Cost of Curation**

The 4C project is a 'Coordination and Support Action' supported by the EU. It focuses on a specific aspect of digital curation and preservation, namely cost and benefit issues related to it. The resulting project will help organisations across Europe to invest more effectively in digital curation and preservation activities by helping them understand the accruing costs and benefits. Apart from project management, the project includes four work packages comprising expert groups collaborating on specific tasks and questions. 'Engagement' deals with communication with a wide range of stakeholders. 'Assessment' concerns the evaluation of existing methods to estimate and compare the cost of digital curation. The 'Enhance-ment' work package is

based on a model that contains all relevant economic determinants of digital preservation; it will be further developed and refined during the project. A 'Roadmap' work package will arrive at coherent and evidence based recommendations for further actions. 4C has 13 partners (among them DANS) in 7 countries.

[4cproject.eu](http://4cproject.eu)

**RDA****Research Data Alliance**

A truly global initiative in the field of research infrastructures is the Research Data Alliance (RDA). It was founded by three research funding organisations from Europe, Australia and the United States. The RDA aims to accelerate and facilitate research data sharing and exchange. The work of the RDA is primarily undertaken through its working groups. Participation in working groups and interest groups, starting new working groups, and attendance at the twice-yearly plenary meetings is open to all. The working groups and interest groups cover a wide range of topics relevant to improve the quality of research data infrastructures. A spirit of cooperation and a low threshold for joining the groups are key features of the RDA.

[rd-alliance.org](http://rd-alliance.org)