CONTENTS

PART2

LANDSCAPE ANALYSIS - SECTION1

SOCIAL & CULTURAL INNOVATION

SOCIAL & CULTURAL INNOVATION

Research Infrastructures that support research across and within the Social & Cultural Innovation domain are among the first known infrastructures: libraries, museums and archives are the most obvious examples of this legacy. In today's digital age, Research Infrastructures in the Social Sciences and Humanities (SSH) enhance research into the historical, social, economic, political and cultural contexts of the European Union, providing data and knowledge to support its strategies.

The data collected and provided by SCI RIs contribute to research that offers new insights into Europe's cultural heritage, its creative industries, education, health and well-being of its citizens, as well as the workings of social and economic policies and societal trends in and across Europe. These insights are fundamental to understand European society and to answer to emerging challenges moving forward.

Data from social sciences and structures that create, collect, assemble and curate such relevant data are fundamental to the further development of the social science research community of Europe. The statistical literacy and research potential of the next generation of social scientists is nurtured using the resources of networked social science data archives and cross-national surveys. The research outputs of Europe's social scientists have impact on Europe's politics and map the social and economic conditions of the continent. Developing better measures of well-being and progress is supported, at the international level, by the Organisation for Economic Cooperation and Development (OECD), the Global Science Forum (GSF) and the European Commission (EC), with important contributions arising from the social sciences.

Public statistics, major scientific surveys, management data and data from opinion polls represent essential sources of knowledge for the social sciences. The development of European indicators on society via longitudinal surveys is a challenging goal to provide a key instrument for constructing Europe.

Research in the humanities provides a better understanding of our society, both diachronically through historical research to answer questions of how we become what we are – by which mechanisms was this driven – and how we can elaborate this knowledge for shaping the future development of our society; and synchronically by monitoring the media to detect what is currently happening in our society – how we react to the major challenges – how societal challenges beside us can be addressed. There is economic impact from this knowledge, as well as a solid base for developing politics and society at large. The increased availability of digital resources in the humanities, and the development of advanced digital methods for research, have prompted remarkable changes in the scale and scope of research in these disciplines. In the area of the humanities, collaboration between the RIs and the GLAM sector – galleries, libraries, archives and museums – will lead to an enhanced impact on culture and society.

Future surveys of the impact of SSH research and RIs will make use of the Social Impact Open Repository (SIOR)[1] allowing researchers to upload qualitative and quantitative testimonies of the social impact of their research, making use of specific indicators[2]. This system also enables funding agencies and citizens to monitor the way in which research impacts society positively, for instance through the reduction of early school leaving. It shows especially the directions of innovation in SSH research are going and how these are integrating with other sciences. The availability of this information will also provide guidelines for SSH research groups and projects to increase political and social impact, as well as a body of evidence to demonstrate the results.

The SSH target community is substantial: according to available data, SSH area accounts for over 40% of the students in Europe. Eurostat[3] reports that 34% of the students are in social sciences – including journalism, business and law – and 11% in arts and humanities. The ISSC World Social Science Report 2016 estimates that more than 30% of the researchers in the higher education system are in SSH disciplines, however with substantial variation among countries. This accounts for more than 500.000 researchers employed – Full Time Equivalent (FTE) – in higher education in the European Union. In contrast, the spending for SSH is substantially lower than 30% of the overall research spending and often lower than 20% in many countries, according to figures given in the same report[4].

[1] Social Impact Open Repository
http://sior.ub.edu/jspui/

[2] Social impact: Europe must fund social sciences. Flecha R., Soler-Gallart M., Sordé T. – Nature, 2015, Vol. 528 (7581): 193.
http://www.nature.com/articles/528193d

[3] Eurostat – Tertiary education statistics
http://ec.europa.eu/eurostat/statistics-explained/index.php/Tertiary_education_statistics#Fields_of_study

[4] ISSC World Social Science Report 2016
http://unesdoc.unesco.org/images/0024/002458/245825e.pdf

EMERGING DRIVERS

BIG DATA

The speed and versatility of electronic communication and the growth of digital media and tools, as well as their accessibility, is underlying the success of the five ESFRI Landmarks in the SCI domain. From large-scale databases to virtual museums, the tools that these developments have fostered are changing the way in which research is carried out. From the old model of *theorise/hypothesise/collect data/test/refine/conclude,* scientific enquiry has now become much more data-driven and, at the same time, more theory- and method-dependent. The ability to rapidly access large bodies of texts in different languages, to examine music archives, to compare three-dimensional images, to analyse census and survey data from around the world provides research possibilities that were inconceivable 20 years ago and redefine Social Sciences and Humanities research.

The term *Big Data* started to be used to describe assemblages of data – data files, datasets, databases or data streams – that, in terms of their volume, their variety, and the velocity of creation, pose severe challenges for many conventional analytical and computational methods in SSH. Such data may be generated by machines through the operation of sensing and imaging devices – e.g. Radio-Frequency Identifiers, imaging equipment; by robotic analysis – e.g. genome wide scans; by social media interactions – e.g. Twitter feeds; mass-recordings of video magnetic tapes – e.g. video cassettes from last centuries art projects; or from the recording of administrative processes – e.g. hospital records, tax and benefit claims. In 2012, digital content grew to over 2.8 Zettabytes (ZB, 10^{21} bytes) to 8.5 ZB by 2015[5]. *Big Data* technologies, tools, and services that turn this information overload into information gains are the next opportunity for competitive advantage, and Language Technology (LT) is a core *Big Data* technology. Growth in the volume and variety of data is mostly due to the accumulation of unstructured text data; in fact, up to 80% of all data is unstructured text data[6]. Moreover, the translation technology segment will continue to dominate the European LT market. RIs in LT are indispensable in breaking new ground. A common characteristic of *Big Data* in SSH is that they have significant research value in terms of the information contained either in its own right or when linked to other data sources. They can, for example, be used to extract information about preferences, or undiscovered relations between people, and therefore provide important snapshots of human activities and orientations. When data are collected over time, such collections will also contain information about how culture and society develop.

While data types are many and varied, their value for research relates to the depth of their content and the extent of their coverage and the possibility to link data from different sources, which in turn is a function of the processes by which such data are generated. For example, supermarket store card data derived from specific and self-selecting customers who shop at particular stores and feel motivated to use a store card to gain a loyalty bonus. With data from millions of shoppers, in particular when linked to social surveys or administrative sources, the information generated can be used to explore dietary patterns and to relate these information to geographical indicators of social deprivation. Data generated by social media interactions can be used to gauge the mood of users, their political affiliations, or to document popular interpretations of significant events – e.g. migration, riots, and virus outbreaks. Biosocial data, such as a genome-wide scan linked to longitudinal life course survey data, represent a special form of *Big Data*, with the potential to demonstrate the links between our health, well-being and lifestyles. These data are evidence bases as well as indicators of the effects of public policies.

Before the research value of *Big Data* for SSH can be realised, three important conditions must be met. First, the data must be accessible for research purposes, often through their availability, digitalisation and normalisation; second, the best possible metadata and methods need to be used to extract and interpret the information; and third, there should be clarity about how the data have been generated. While the first condition seems obvious to researchers, data holders may place restrictive conditions on research when individual-related data or commercial interests are involved. In addition, linking data from different sources substantially increases their scientific potential but raises many practical issues to be addressed: technical connected to data integration, and legal related to data protection.

[5] The Forum for Europe's Language Technology Industry
https://www.lt-innovate.org/sites/default/files/LTIpresentation_European_Data_Forum100413_0.pdf

[6] Unstructured Data and the 80 Percent Rule
http://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/

NEW MEANS FOR COMMUNICATING AND DISSEMINATING RESEARCH

The mechanism for dissemination of research results emerges as one of the most important predictors of extra-academic impact. Open Access has gained momentum with the involvement of Governments from different countries and the support of funding agencies for research in order to create a strong Open Access Landscape in Europe.

Open Access is just one component of Open Science – the movement to give access to data, research and publications and open up the whole research cycle for participation and collaboration. Initiatives such as the European Open Science Cloud (EOSC)[7], or the Open Access mandate that goes along with the Horizon 2020 Framework[8], are strong messages to the whole scientific community. Yet, the development of Open Access for publications in SSH seems to lag behind other scientific disciplines. For example, the Directory of Open Access Journals indicates that only 41% of the journals[9] – representing 23% of the articles – belong to the SSH area. One of the reasons explaining this divergence could be that a large part of the scientific production in SSH disciplines is published in books and not journals. Even if quantitative social research is often coming closer to other sciences in terms of methodological approaches, books still play an important role in the dissemination of knowledge for those disciplines. Open Access for academic books in SSH develops under different conditions than those we know for articles in the natural or hard sciences. The challenges are different in technical and economic terms as well as in usage, and there are many initiatives in different European countries, by the publishers, the libraries or the scientific communities themselves, which need to be better coordinated in the future.

The dissemination is increasingly not limited to the publications but to the data underlying publications. The request to make primary sources of SSH data available is increasing, both from funding agencies and publishers. This trend in open science not just for publications, but for data as well, requires new and enhanced capabilities to store and access different type of data in the future.

The EOSC aspires to become the European stakeholder-driven infrastructure for science and innovation. It will not only be a data repository, but will also comprise technical elements of connectivity, hardware, repositories, data formats and Application Programming Interfaces (APIs) and it will offer access to a wide range of user-oriented services, data-management, associated HPC analytics environments, stewardship services and, notably, expertise.

NEW FORMS OF INTERDISCIPLINARITY

Recently there has been an increase in the value and practice of interdisciplinary SSH research.

INTERNAL INTERDISCIPLINARITY. The traditional fragmentation of the area is being overcome: social sciences and humanities makes way for promising interactions. Disciplinary boundaries are gradually fading to make room for integrative and transversal research methods concerning the entire field of Social Sciences and Humanities. On the one hand, a large body of digitised texts allows Humanities to use quantitative methods that were previously confined to the Social Sciences. On the other hand, a *linguistic turn* within the Social Sciences, makes room for new types of discourse and conversation analysis. Media Studies, which connect the Social Sciences and Humanities, are an eloquent example of that evolution. In particular, the scientific study of the web, which has become an integrated part of society, culture, business, and politics, is a burgeoning field of research activity, with enormous potential for contributing to societal challenges related to the evolution of communication, solidarity or security issues.

EXTERNAL INTERDISCIPLINARITY. The increase of the interaction between SSH and other sciences is one of the most salient features of the recent period. There is now a more acute perception that many causal chains that are the object of natural sciences have their determinants in human action and behaviour. To cite just one example, the extraction of oil from bituminous sands and shales in Canada is expected to move every year more than two times the total mass of annual river sediments in the whole world. While the environmental impact of such extraction can be estimated by natural sciences, it requires the social sciences to analyse and understand the decision-making processes that lead to or can avoid such massive changes in the environment. This change has been accentuated by recent developments in the way of managing science. Horizon 2020, which is not structured by disciplinary fields, but by societal challenges – e.g. health and well-being, climate changes – is the paradigmatic example of this transformation of the science system in Europe. This new approach poses the question of hybrid infrastructures, aggregating data arising from different domains or, alternatively, new forms of collaboration and interchange between existing infrastructures. A good example of this hybridization is provided by the **ESFRI Project E-RIHS** (European Research Infrastructure for Heritage Science) which combines material science methods with interpretative schemes of history of art to rejuvenate the field of heritage studies.

[7] European Open Science Cloud
https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

[8] H2020 Programme Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

[9] Directory of Open Access Journals
https://doaj.org/

CURRENT STATUS

The digital aspect of the currently most prominent ESFRI Landmarks and ESFRI Projects is outlined below to testify the progress in the use of digital techniques throughout SSH research methodologies.

Scientific databases are a crucial part of the pan-European infrastructures and more generally in the global science system. Effective access to research data, in a responsible and efficient manner, is required to take full advantage of the data and the possibilities offered by the rapidly evolving digital technology. Accessibility to research data is an important condition for maximising the research potential of new digital technologies and networks. An open and democratic access policy not only provides scientific advantages to the whole academic community, it also provides greater returns from public investments in research activities.

The **ESFRI Landmark CESSDA ERIC** (Consortium of European Social Science Data Archives) is a distributed Research Infrastructure that provides and facilitates researchers' access to high quality social science data and supports their use of this data. The CESSDA Work Plan 2017 informs about significant improvements in information retrieval across the range of relevant service providers and other sources. In this regard, the CESSDA Product and Service Catalogue (PaSC) will be made fully operational in 2018, and a new retrieval tool is under development. Other improvements are in the areas of metadata management, technical architecture, PID policy, and outreach to new scientific cohorts through workshops on data discovery, collaborative data management, etcetera. CESSDA does not disseminate data itself, it coordinates the activities of the national data service providers across Europe. In total it holds, curates and provides access to several thousands of separate data collections, supporting a European-wide user community. CESSDA operates within a global data environment, with reciprocal data access arrangements and agreements established with other data holding organisations worldwide.

The **ESFRI Landmark ESS ERIC** (European Social Survey) is an academically driven long-term pan-European survey designed to chart and explain the interaction between Europe's changing institutions and the attitudes, beliefs and behaviour patterns of its diverse populations. ESS recently announced 100k registered users. The ERIC updated its Multilevel Data resource in June 2017. Italy, Slovakia and Spain were the final countries to confirm that they will take part in Round 8 (2016/17) of the European Social Survey alongside 21 other European countries. ESS has a broad scientific network and is widely used for academic publications: 109.063 users were registered by the end of August 2017; 76.677 users have downloaded ESS data from all over the world by end of August 2017. Bibliometric data from a google scholar analysis shows 3.140 ESS publications and citations based on ESS data recorded in the period 2003-2015; 2.821 outputs and publications are registered in the ESS online bibliography as of 5 September 2017[10]. In addition, HypeStat[11] states: europeansocialsurvey.org receives about 650 unique visitors and 2.210 (3.40 per visitor) page views per day.

The **ESFRI Landmark SHARE ERIC** (Survey of Health, Ageing and Retirement in Europe) is the upgrade into a long-term Research Infrastructure of a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of about 110.000 Europeans aged 50 or over. SHARE covers 27 European countries and Israel. The main data collection of Wave 7 is under way in all these countries. As the largest pan-European Social Science Panel Study, SHARE broke a new country record – there have never been more countries in one wave. 6.400 SHARE users were recorded in February 2017. Furthermore, new country teams in Romania, Cyprus and Slovakia were established between 2016 and 2017. SHARE had more than 7.000 users by July 2017. Overall 1.836 publications were produced with SHARE data, 815 were peer reviewed articles[12].

The **ESFRI Landmark CLARIN ERIC** (Common Language Resources and Technology Infrastructure) provides easy and sustainable access to digital language data – in written, spoken, video or multimodal form – and advanced tools to discover, explore, exploit, annotate, analyse or combine them, wherever they are located. In 2018, CLARIN has 20 Members, two Observers and one third country institution (Carnegie Mellon University, USA). CLARIN has a global vision, although the focussed user community is European (EU and Associated States). In 2018, CLARIN has more than 40 centres and more than 160 institutes. Europeana[13] has 800.000 resources listed in CLARIN's Virtual Language Observatory (VLO). The number of monthly visits to the CLARIN website is slowly growing since January 2015. In addition, there is a steady increase in the number of visits to the Discovery Service – used during single-sign on to CLARIN service providers. This can most probably be explained by the increase of the number of services available and their growing popularity.

The **ESFRI Landmark DARIAH ERIC** (Digital Research Infrastructure for the Arts and Humanities) is aiming at enhancing and supporting digitally enabled research and teaching across the Arts and Humanities. During 2017, DARIAH launched its Teaching Platform in Digital Arts and Humanities.

The **ESFRI Landmarks CLARIN ERIC** and **DARIAH ERIC** have collaborated on the relaunch of the Digital Humanities Course Registry. Recent workshops have focused on Data Management, Software Sustainability, Research Ethics, Access, Staff Training, Impact, etc. Numerous universities and libraries joined as cooperating partners in the first half of 2017.

The **ESFRI Project E-RIHS** supports research on heritage interpretation, preservation, documentation and management. It connects researchers in the humanities and natural sciences and facilitates a trans-disciplinary culture of exchange and cooperation. E-RIHS enables the provision of state-of-the-art tools and services to cross-disciplinary users and communities. It aims at the advancement of knowledge about heritage and the division of innovative strategies for its preservation. Based on the preliminary work done in the framework of the H2020 IPERION-CH project started in May 2015, the **ESFRI Project E-RIHS** is currently in the Preparatory Phase which will be used to address legal status and governance/management organization. This will lead to the foundation of an ERIC – or another suitable legal form – to be launched in 2020. Further developments are planned for connecting and including partners and facilities outside the EU, and gradually reaching the status of a global distributed Research Infrastructure.

[10] ESS User Statistics
http://www.europeansocialsurvey.org/about/user_statistics.html

[11] HypeStat
https://hypestat.com/

[12] SHARE Project
http://www.share-project.org/

[13] Europeana
https://www.europeana.eu

GAPS, CHALLENGES AND FUTURE NEEDS

The development of high speed data connection together with storage capacities and information processing software has provided access to massive amounts of data, as well as new ways of analysing analogue resources of cultural heritage. The disciplines in Social Sciences and Humanities are thus confronted with a momentum that is transforming the entire profession of the researcher. Research Infrastructures in this area must enable the creation and manipulation of large and very heterogeneous bodies of data, of a qualitative or quantitative nature, opening up new research possibilities and encouraging interdisciplinary work. RIs contribute to the valorisation of scientific and cultural heritage.

Data storage and digital interactivity, have opened up new opportunities in terms of appropriation and handling of research resources. Consequently, we have seen a diversification in the locations of digital resource production which have resulted in the creation of many platforms. They form clusters for bringing together disciplinary and technological skills that offer many services to support researchers in the humanities who use ICT either directly because the research data is digital or as an environment allowing for access to new processing tools.

Exponential growth in the amount of data, their increasing use by SSH scholars, as well as the rapid evolution of technology, opens up new opportunities for SSH research. The use of *Big Data* also bears new methodological challenges with implications for empirical research: the implementation of surveys on emerging social trends in longitudinal perspectives can lead to important advances in epistemological and methodological fields. In particular, *Big Data* raises some important issues for the SCI domain, among which we point out the following.

THE PREDICTIVE CAPACITY OF BIG DATA. The challenge is to understand if and how a large amount of data can improve and enhance predictive capacity regarding social phenomena. One concern is that *Big Data* might lead to causality being overlooked, since this perspective relies on correlations and trends whose underlying cause may not be clear.

BIG DATA AND THE ROLE OF THEORY. The advent of *Big Data* has often been accompanied by concerns about the end of theory and about a kind of knowledge that is only data-driven. The question is to understand if the use of *Big Data* can contribute to the generation of interpretative patterns of past, present and future social, political or cultural reality, or identification of the limits of machine learning for humanist and sociological knowledge.

BIG DATA AND DATA PROTECTION ISSUES. The use of *Big Data* in research does not only refer to the issues of privacy protection, but also concerns about access and possession of such data and in many cases also copyright issues. Solutions at national and European level must be found in order to enable researchers to use *Big Data*, given the ethical and legal challenges.

METHODOLOGICAL CHALLENGES OF BIG DATA. The scientific community has some concerns about the validity and the reliability of *Big Data*: the discussion is how to use them in a controlled way in order to produce scientifically relevant inferences.

BIG DATA ANALYSIS. Especially in the field of opinion mining and sentiment analysis, treatment and analysis of *Big Data* rely on automatic procedures more or less monitored by the researcher and on algorithms of machine learning through which the software is able to classify a large amount of textual information. This raises classic methodological problems in a new form: the inspection and thus the control and evaluation of data analysis procedures.

In light of the changes outlined in the preceding section, new forms of Research Infrastructures combining storage and state-of-the-art information extraction methods and services are required if the research community is to utilise all potential research opportunities. This section identifies a number of areas in which the changing research landscape needs to foster new research opportunities in SSH and at the disciplinary boundaries with other scientific communities.

INTEGRATION OF BIO-SOCIAL DATA

Interdisciplinary research cutting across SSH has the potential to supply increasingly rapid insights into the influence of socio-economic and environmental conditions on biological changes. These have long-lasting consequences for our behaviours, health and socioeconomic well-being through the life cycle. There is a need to understand the pathways and mechanisms involved in these reciprocal feedbacks over the range from cells to society. Bringing together interdisciplinary teams to address these research issues and ensuring that our longitudinal and cohort studies are augmented and enhanced to enable such research, provides new opportunities for scientific discovery.

Bringing together data from diverse sources, spanning genomics, blood analyses and biomarkers, health and other administrative records, and business or transaction data, and linking all of these into the rich longitudinal cohort and panel studies presents several major challenges to ensure accessibility and usage. Many researchers would benefit enormously from having complex data pre-processed and summarised in useful ways.

Existing RIs, such as the **ESFRI Landmark ELIXIR** (A distributed infrastructure for life-science information, H&F) and **ESFRI Landmark SHARE ERIC**, indicate that there is significant potential at the pan-European level to integrate a range of biomedical and socio-economic data resources during the lifecycle. In the same way, the surveys and data provided by the emerging project GGP will contribute to the analysis of generational differences in values and gender roles that are highly relevant for policy debates.

Further development in this area of the European research landscape will provide fertile ground for trailblazing research with huge potential benefits for the health and well-being of populations.

PROMOTING AN INTERNATIONAL APPROACH TO REAL-TIME DATA ANALYTICS

Historically, data that have been used for research in the social, economic and behavioural sciences have been designed and/or collected specifically for that purpose. In recent years, however, new forms of data not originally intended for research use, such as transactional and administrative data, internet data (derived from social media and other online interactions), tracking data (monitoring the movement of people and objects), and image/video data (aerial, satellite and land-based), have emerged as important supplement resources and alternatives to traditional datasets. In quantitative humanities, however, the analysis of resources that were not created for specific research questions has a long tradition. The problem of non-tailored resources arises newly since large amounts of digital resources are available to researchers. The troubling gap is emerging in our ability to capture and explore these new forms and amounts of data for the purposes of research. This gap is arising because:

- the prevalence of new forms of data will increase exponentially as technologies and digital capabilities evolve and it is imperative for the SSH community to take a leading role in establishing a robust, quality assuring, secure and sustainable infrastructure for utilising them;

- technological and methodological advances must be made in order to realise the potential of real-time analytics for research in SSH;

- much of the value of new forms of data lies in the potential for linkage and calibration with other data and the derived opportunities for addressing novel research questions aa well as re-examining open questions through a new lens;

- current training and capacity-building provisions are insufficient to meet the growing demand from researchers at all stages of their careers to utilise new forms of data;

- new forms of data and their subsequent uses pose novel ethical, quality and privacy questions that must be explored to ensure that these technologies are deployed in a responsible way.

Europe has to implement standards with respect to privacy and security issues in Research Infrastructures. The recently established General Data Protection Regulation (GDPR) could bring clarity and benefit to research across the SSH and associated disciplines – e.g. medical sciences, health research etc. The new GDPR is an opportunity to develop and establish a lead in regard to privacy and security issues concerning Research Infrastructures.

RIs FOR SOCIAL MEDIA – ARCHIVING WEB

Since the mid-1990s the web has become an integrated part of society, culture, business, and politics, and national web archives have been established to preserve this part of the digital cultural heritage. But for the scholar who wants to study the web across borders, national web archives become an obstacle since they delimit the borderless information flow on the web by national barriers. Thus, a transnational Research Infrastructure should be established with a view to: i) developing a more efficient and attractive European Research Area; ii) ensuring the researcher free access to the digital cultural heritage from different nations; and iii) increasing the potential for fostering innovative partnerships with the software development industry for studies of *Big Data*.

RIs FOR HUMANITIES AND CULTURAL INNOVATION

Contemporary technologies offer great opportunities to revitalize and make available on a large scale cultural items which represent a collective treasure for Europe in terms of identity, citizenship, diversity, cultural growth, and economic potential. The effort in that direction should be conceived in two different ways:

- Cultural items – manuscripts, papyri, books, movies, music, paintings, monuments, etc. – in their material reality, are complex physical objects that are in need of material analysis, dating, preservation and restoration. Viewed in this way, they are relevant for RIs which aim to support the analysis of physical objects in general.

- Making a material object part of our cultural heritage largely depends on the collective awareness of its existence and on the value vested in it. In this respect, RIs devoted to the dissemination (digitisation, 3D-reconstitution, etc.) of those objects are crucially needed to the maintenance of our cultural heritage.

An enormous amount of diverse materials is widely distributed across Europe: they are often difficult to access from outside local communities, and sometimes at risk of deterioration. The main challenge of RIs is to provide users access (educators, museums and exhibition curators, public) to such treasures and heritage, and to the state-of-the-art analysis carried out by experts and researchers, also by exploiting digital media and archives.

National museums and integrating Research Infrastructures such as the European Cultural Heritage Online (ECHO)[14] and Europeana[15] have made important efforts to digitize libraries and collections. ECHO was established in 2002 to create a research driven IT infrastructure for the humanities. It works on digitisation of cultural heritage and develops research driven tools and workflows for analysis and publication of scholarly data linked to primary sources. ECHO features more than 70 collections from more than 24 countries worldwide. Europeana is a European network representing more than 3,300 institutions and aggregators and provides cultural heritage collections to all in the form of more than 30 million digitised objects and descriptive data.

However, making these treasures accessible in digital form is only the first step in ensuring their uptake by the target audience. The vast bulk of the cultural heritage accumulated through centuries of European history is a formidable resource of rich material for new and far-reaching analyses, typically in languages no longer spoken. Therefore, the mere availability of this heritage no longer

[14] European Cultural Heritage Online (ECHO)
http://echo.mpiwg-berlin.mpg.de/home

[15] Europeana
http://www.europeana.eu/portal/

guarantees that current scholars, let alone the general public, will be able to internalise it through conventional methods, i.e., reading about this heritage and annotating it. Thus, new methods of intelligent information mining and text analytics are needed that should be capable of automatically processing the content of the massive amount of cultural heritage treasures and making them accessible to present day audiences. In response, initiatives have been made by setting up projects to record and possibly revitalize endangered languages, in which social media can also play an important role[16]. It may be noted that CLARIN has examples from a very large number of languages, and that more than 1.500 languages are represented with 5 examples or more.

Meeting this challenge requires significant interdisciplinary efforts to integrate competences from different expert fields, bring together the most advanced facilities and make their resources available on a large scale. Infrastructures such as the Cultural Heritage Advanced Research Infrastructure (CHARISMA) contributed to the development of joint activities in the field of conservation of cultural heritage. CHARISMA covers joint research, transnational access and networking of twenty-one organizations that provide access to advanced facilities, develop research and applications on artwork materials for the conservation of cultural heritage and open up larger perspectives to heritage conservation activities in Europe. It has defined and consolidated the background for the **ESFRI Project E-RIHS**.

Additionally, in the archaeological sciences the ARIADNE network developed out of the vital need to develop infrastructures for the management and integration of archaeological data at a European level. As a digital infrastructure for archaeological research ARIADNE brings together and integrates existing archaeological research data infrastructures so that researchers can use the various distributed datasets and technologies. ARIADNE has strong ties to the **ESFRI Landmark DARIAH ERIC**.

INCREASING THE GLOBAL REACH

Given that RIs in the Social Sciences and Humanities will be stably anchored in Europe in the future, further actions have to be undertaken to make them attractive and compatible on a global scale.

The accessibility of digital research data – e.g. survey data in the Social Sciences, digitized and annotated cultural heritage in the humanities – is obviously the key driver for increasing global research not only in the Social Sciences and Humanities but in the whole scientific system. It can be stated that not only is there still a possible gap between different standards for certain kinds of data – depending on the source from which they were derived – between European infrastructures and non-European infrastructures, there are also rather difficult challenges to meet that are inherently connected to the content types of data and research traditions between different parts of the world.

Digital tools and functions have to be potentially newly programmed when applied to sources in languages, scripts or symbols more recently encountered by the technology. The underlying understanding of text types, art classification systems and semantics would potentially have to be adjusted, complete methodologies would have to be newly negotiated. The same is true for political and sociological research terms and classification systems. In order to integrate and interconnect heritage and knowledge from and about societies and cultures from all over the world, their history and self-conception a lot of work has to be done to enable global infrastructures to offer a certain degree of consistency between these data and concepts.

SUSTAINABILITY AND GEOGRAPHICAL COVERAGE

There is an increased sustainability of the research data due to the fact that all RIs provide archives for storing data and state-of-the-art methods to analyse and interpret them. This is an important difference with respect to ten years ago where data could disappear when a researcher retired. Currently not only data are stored in sustainable, long-lasting and secure archives, but the current RIs – e.g. the **ESFRI Landmarks CLARIN ERIC** and **CESSDA ERIC** – also use innovative methods such as Persistent Identifiers for resources and data collections, so that the same version can always be retrieved and so that research based on their data can be replicated or extended. We also need to consider the sustainability of the RIs themselves. Research Infrastructures need to be sustainable: i) financially and organisationally; ii) technically; and iii) in terms of human resources. These three dimensions of sustainability are heavily interlinked and therefore require adequate financial resources. The organisational sustainability is supported through the use of the ERIC and other legal structures. The financial sustainability of the central and national operations may still be an issue worth considering. For all of the SSH Infrastructures, geographical coverage is crucial for the quality of the research they support and hence for their sustainability. Data from one country is not only of interest for the researchers of this country itself, but also for everyone else in Europe for comparison. To compare attitudes to different aspects of society, it is not enough to have one part of Europe if other parts are missing. For those Infrastructures where language plays an important role, it is obvious that a very good geographical coverage is needed, so that all types of languages, and preferably all languages, are described and will be the basis for the research developments. European data collecting infrastructures only have European Added Value, if they are able to provide data from all over Europe. Technical sustainability has to do with upgrading to new versions, following and updating standards, including new tools and possibilities, following international developments. All current SSH Research Infrastructures are heavily involved in and committed to continuous technical development.

[16] The Endangered Languages Project
http://www.endangeredlanguages.com/

Sustainability in terms of human resources is at the heart of our infrastructures. There are three classes of activities where human resources are crucial:

- building and operating the infrastructure and keeping it up-to-date in the light of technological and methodological developments and evolving user needs (this is treated above under technological sustainability);

- instrumentation and population of the infrastructure with community specific data and services;

- education, training and research support for existing and future users.

There are various instruments to make these things happen in a sustainable way, and they are all implemented to some extent by the current ESFRI Landmark SSH RIs. For example, building knowledge about the availability of RIs within standard university curricula is a good, sustainable long-term investment. In the shorter term the obligation for infrastructures to build and maintain what could be called a *Knowledge Sharing Infrastructure* is important. Knowledge Sharing Infrastructure is a formalized way of recognizing and sharing knowledge among members. It is an acknowledgement that not all useful knowledge can be concentrated at the central level, and that the knowledge present at the national level is crucial for sustainability and has to be made visible and shared. This is particularly true for distributed Research Infrastructures like the SCI RIs.