8 A Common Agenda for the European Research Infrastructures in the Social Sciences and Humanities

Adrian Duşa (SCI-SWG), Claudia Oellers, and Simon Wolff (RatSWD)

Research in the social sciences and the humanities (SSH) in Europe is currently facing a historic turning point. The social sciences and the humanities have been included in the European Union's new framework program for research and innovation Horizon 2020 and embedded across the societal challenges the program seeks to address. The integration of both disciplines is a sign for the increasing recognition and the essential role SSH plays in addressing the societal challenges and the great number of pressing issues Europe faces today and in the near future. These include employment, demographic change and ageing populations, migration, poverty, climate change, food and energy security, European cohesion and cultural diversity.

The pan-European research infrastructures (such as the ESFRI SSH infrastructures CESSDA, CLARIN, DARIAH, ESS, SHARE) were a necessary precondition for SSH to take on this important role. Today, these infrastructures provide the necessary means for innovative research. The existing wealth and diversity of European research infrastructures and the complex issues they address is reflected in this book's contributions.

The "Facing the Future" conference was held to reflect upon what has been achieved, and what the requirements for the European SSH research infrastructures will be in the future, not only to meet the ambitious challenges set out by Horizon 2020, but also to stay internationally competitive and to strengthen pan-European research initiatives in the long-run. Moreover, it was an important aim of the conference to identify present and future challenges for European research infrastructures from the perspective of their users, the SSH research community, and to formulate a common agenda, for both the social sciences and the humanities with regard to the advancement of research infrastructures at the European level.

The conference participants from the social sciences and humanities (researchers, policy makers, funding agencies, scientific research infrastructure coordinators) were able to find much agreement and common ground regarding future challenges and infrastructure needs. In the course of the conference, five essential challenges were identified as key towards strengthening European infrastructures and thus ensuring that the social sciences and humanities will be able to fulfill their important role in the future. 1. Ensuring sustainability and establishing permanent/sustainable institutions. 2. Facilitating research cooperation and interdisciplinarity. 3. Tapping new sources of (big) data. 4. Safeguarding data protection at all levels of research. 5. Increasing the visibility of SSH research infrastructures in their respective fields and for the general public. Clearly, these challenges are inherently interrelated and the following synopsis cannot consider every aspect in full detail.

1. Ensuring Sustainability of Research Infrastructures in the Social Sciences and Humanities

Across the board, sustainability of research infrastructures was considered a crucial issue. In the current European framework, sustainability of research infrastructures is at risk due to the lack of sufficient and long-term funding (funding only for one survey wave, or funding that only covers implementation/set-up costs) as well as the lack of cross-national cooperation, and the fragmentation between the European member states (diverging time frames and funding rounds, differences in national science policy). The introduction of the ERIC legal framework created the basis for pan-European projects; however, it did not install a sustainable European funding scheme. Solutions need to ensure sustainability of European research infrastructures by establishing them as permanent institutions at the European level.

While there are differences between infrastructures, some focusing more on providing data, while others focus more on providing tools and services, the common challenge is turning them into sustainable institutions. By definition, research infrastructures in SSH must be designed to be durable, stable and reliable. At the same time, they operate at the forefront of scientific innovation and must constantly evolve and adapt to technological progress, advances in methodology, and, most importantly, the needs of their users in science and research. In this dialectic of continuity and innovation, as Peter Farago puts it in his article, research infrastructures must continue to find the right balance.

Social science research infrastructures that provide primary research data, for example major, cross-national and longitudinal surveys, rely on broad participation and long-term funding to achieve their full potential of comparability over time and across national borders. While the initial costs for setting up these infrastructures are relatively small, they have to be consistent in the long-run. Large-scale social surveys have become invaluable sources for monitoring social change which can only fulfill their function when built on a foundation of valid, complete and comparable data. Similarly, the preservation and long-term archiving of Europe's cultural heritage in Europe and the provision of these high-quality data to research form the backbone of humanities research that requires serious long-term commitment. Long-term preservation of European cultural heritage and the leading role of Europe in the humanities will depend on the capacity to continue to build and sustain these infrastructures (see Niccolucci in this book) over time.

The experiences of the Survey of Health, Ageing and Retirement in Europe (SHARE) project (see Börsch-Supan in this book) illustrate the problems of fragmented funding schemes for pan-European projects and the dropping-out of participants due to the lack of funding at the national level. Ageing is a very different process in various European regions, and research of this issue demands high quality data from all those regions. Typically, it is those countries in which research on political, economic and social changes is particularly urgent that are not included in the European research effort. Failure of European projects or individual countries dropping out due to the fragmentation and lack of funding is a crucial issue that has to be prevented in the future in order to achieve a real European added value for these research infrastructures.

Moreover, there is a strong need for scientific leadership beyond single committed individuals in order to be able to advance research infrastructures. In a funding environment where resources are allocated according to competition between various stakeholders, public funding institutions require strong partners in the scientific community who represent their constituency and formulate needs and requirements.

2. Facilitating research cooperation and interdisciplinarity

Interdisciplinarity is essential to meet the complex challenges of the 21st century such as labor markets, ageing populations, climate change, food and energy security, or cities and well-being. Such cutting-edge research will increase the need for interdisciplinary research by default in order to achieve robust results because they transcend academic disciplines as well as national borders. Addressing the "real life problems" which make up the societal challenges described by Horizon 2020 requires an integrated, interdisciplinary approach between very diverse scientific domains.

Research infrastructures facilitate and foster interdisciplinarity by design, promoting cooperation and harmonization as well as strengthening cross-national projects. An essential prerequisite to making interdisciplinarity work is the comparability of data over time, across national borders, and between scientific disciplines. Comparative analysis lies at the heart of the European infrastructures in the social sciences. Research infrastructures have been successfully working on increasing comparability between regions and countries through common methodological frameworks, international harmonization platforms and international data portals. But there is still much effort required to enhance interdisciplinary comparability of data on a trans-European level.

There is unanimous agreement on the benefits of cooperation. However, it remains a huge challenge in light of the extreme heterogeneity of the existing research infrastructures. Most research infrastructures are established in response to a demand for specific data on a national level and are not coordinated with other infrastructures from the onset. More coordination between research infrastructures is needed at all levels – not least to avoid a loss of efficiency and a waste of resources. In order to fully exploit the potential of this diversity, by creating synergies between infrastructures instead of doubling institutions, coordination among them has to be improved at the national and international level. Some efforts have started with the DASISH project, funded by the European Commission through the FP7, which seeks to reach compatibility between the five projects of the ESFRI roadmap for the Social and Cultural Innovation area, in order to find common solutions to common problems (see also Marker in this book).

The project Data without Boundaries (see Bender in this book) has been successful in by coordinating existing infrastructures, the Council of European Social Science Data Archives (CESSDA) and the European Statistical System (ESS), and is making access to official research microdata easier than ever by overcoming the need for multiple accreditations when requesting access to comparable datasets.

Another pervasive challenge in this area is establishing common standards between infrastructures and disciplines. A prerequisite for linking different kinds of data are common, harmonised standards for data documentation. In the humanities, especially, there is still much room for improvement regarding standardization. In the future, these efforts will have to be extended further to facilitate linking different types of research data from different sources and disciplines.

3. Big Data: Tapping New Data Sources for Research

New forms and sources of data are emerging everywhere in the Digital Age. Today, wittingly or unwittingly, individuals produce vast amounts of data just by being online, going shopping, or using a mobile phone; personal data are shared in social networks such as Facebook or Twitter. The quantity as well as the quality of these new data sources is challenge in itself.

However, while not all of these data are scientifically significant, there are many hidden sources of data that could potentially have enormous value for innovative social research and which previously could not have been collected or were not available (see Maynard and Strohmaier/Zens in this book). The potential of these new data sources is higher when they are linked to ("traditional") survey data. Linking data from various sources to other data, such as private sector data (commercial data, tracking data, and satellite imagery), internet data (social media) or biological data (mental/physical measures, biomarkers, genomics), will be of growing importance and has to be facilitated by infrastructures in the future. This will also require science to increasingly turn to partnerships with private actors for gaining access to data that is not produced by science itself. Issues that have to be addressed are how to provide access to this data under simultaneous consideration of the legal issues related to personal information contained in the data, the question of consent for reuse and the safeguarding of data protection, quality control for unprepared and undocumented data, replicability and durability (see Marker in this book for detailed requirements).

On the whole, it will become more important to establish access to relevant new data sources for research. Where specific and relevant research data does not yet exist, it is important to fill in the gaps. However, efforts should be continued to open up and connect already existing data repositories to research infrastructures. An important step would be to set up a central, international (European) centre for cross-disciplinary research on new forms of data and the establishment of European data service centres. Especially in the UK, pioneering progress has been made in making a wealth of administrative data available to research which could serve as a model to other countries. Administrative data have an enormous potential for social science research because they cover a majority of the population in great detail, are already prepared in many ways for scientific use, and are relatively up-to-date (see Woollard and Bega in this book).

4. Data Protection, Confidentiality, and Research Ethics

One of the most pervasive challenges to research, particularly in the social sciences, is safeguarding data protection at all levels of research without obstructing innovative research. The different regulations of data protection standards on a European level remain a challenge for cross-national cooperation. The planned general European data protection regulation will be a step towards harmonizing national standards in the EU.

In the age of big data and increasing voluntary disclosure of private information, researchers have to reflect on the changing nature of privacy and confidentiality. Who do you ask for consent when tapping sources of big data? As mentioned in the last paragraph dealing with new data sources, data protection in this changing environment will require an international research agenda to reflect on these issues and find new ways of dealing with a new data situation (see Lane et al.).

In the future, it will be essential to find the right balance between data acquisition and data protection which is always a question of finding the right balance. Research infrastructures will continue to play an important role in establishing best practice of data protection and research ethics.

5. Promoting Research Infrastructures and Increasing Visibility

Among the greatest challenges to the SSH research infrastructure community is to increase the visibility of research infrastructures, integrating them further into the daily work of researchers, and enhancing the way their benefits are showcased towards the general public.

First, visibility of research infrastructures has to be further improved by promoting their benefits to their potential users, the researchers in the social sciences and humanities, and to encourage sharing, using, and finding research data by using existing infrastructures. The notion of "one researcher, one project, one dataset" is gradually being superseded by a culture of sharing, cooperation and re-use of data, but there is still much work to be done. While there has been considerable progress in promoting a culture of data sharing in the social sciences, there is still a lack of knowledge and acceptance among researchers in the humanities. Especially in the latter, the development towards setting up infrastructures and the methodological transition to Digital Humanities has been met with some scepticism. However, the main reason for reluctance to engage in data sharing is the same for both disciplines: missing incentives for researchers for investing time and energy into preparing their data for secondary use. The hard work involved with producing, and also sharing, high-quality data is currently not being appropriately rewarded by journals, universities and funding agencies with professional credit. Researchers involved in setting up research infrastructures perform a service of great value to the entire scientific community and should not fall behind the reputation of colleagues due to this important commitment. Technical difficulties, however, are also an obstacle to sharing data. This should be facilitated by making data management and documentation as easy and user-friendly as possible. The same applies for research tools which should be designed in close cooperation with their users in research to ensure usability. Further developing international standards for documentation and the use of persistent identifiers will also lead to greater recognition of data production as an important scientific achievement of its own.

However, a simple lack of knowledge about what is being offered is also often a reason why researchers do not make use of the resources research infrastructures provide. These have to be better integrated into the daily work of researchers. To this end, finding suitable data for a research project should be made as easy as finding other information on the web. This need for efficient

and easy-to-use tools needs to be addressed in the near future. Moreover, data portals should be organised in a more centralized fashion in order to avoid searching across multiple sources.

The second aspect of visibility is that SSH research infrastructures need to be more visible towards funders and policy-makers. The integration of the social sciences and the humanities into the framework of Horizon 2020 is a great step forward, but, more generally, SSH are usually not deemed as essential as the natural sciences. Moreover, SSH will soon be required to produce concrete data on the impact of publicly funded research infrastructures in view of future project assessments, as Žic Fuchs points out in her contribution to this book. Gathering detailed information on the added value produced by research infrastructures, while necessary, might also enable SSH to showcase their success on the basis of specific information.

The third visibility challenge is to promote the benefits of research infrastructures beyond the field of science. Towards the greater public, the natural sciences have been very successful in justifying enormous investments in e.g. the Large Hadron Collider or space programs by showcasing their merits and touching upon people's imagination. Without doubt, the social sciences and the humanities are producing research results that are equally relevant to many, but ways have to be found to better communicate and showcase their results to the public. Since most research is publicly funded, citizen participation is essential. Especially in the social sciences, where vast amounts of personal information from citizens are gathered for research, it will be increasingly necessary to make the public aware of how and to what extent personal data are gathered, how the data is processed and put to use, and, ultimately, why this will advance progress in science and society.

Not least due to the European-wide establishment of research infrastructures, the European social sciences and humanities are world-leading and will continue to play a crucial role in analyzing the societal challenges facing Europe in the future. Their capacity to do so will depend on the capacity of the scientific community, research funders, and European policy makers to find good and innovative ways to meet the challenges for European research infrastructures in the future.