

# **OPEN ACCESS TO DATA: A NEW ROLE FOR CESSDA**

John Kallas

## **1 Open access to data across Europe**

In the context of Globalization and European integration which are based on information exchange, comparative research and open access to data become issues of great importance for all the European countries. Information for research purposes is not a scarce resource in Europe for the social sciences. Well developed official statistical systems combined with a variety of academically driven data gathering programs and activities are producing a wealth of data and information about various aspects of European societies. This also includes collections of multimedia content. However the majority of these resources are country or nation specific. They are produced to meet national requirements and collected by means of nation and language specific instruments based on local methodologies and classifications. They are normally documented in local languages only and rarely published for use outside the country of origin. On top of this, nation specific access restrictions will often prevent information to travel abroad.

To support open access to data across Europe we need a European research infrastructure. But we need also common research programs and practices across Europe.

## **2 The need for a European research infrastructure**

To encourage and nurture research within social sciences at a European level, research infrastructures offering a series of crucial functions and resources must be established. Many of these are extensions of functions and resources already existing at a national level in some European countries. In some cases the model service might even be present at a European level, but would need to be extended in scope, strengthened or multiplied to other disciplines. The development of research infrastructure must at least support the following

### **2.1 Data production**

For the social sciences production of data is essential especially to comparative research. Long established data gathering projects like the Eurobarometers, ISSP and the World Value Studies have for decades been the only available data sources designed to produce comparative empirical evidence. The newly established European Social Survey (ESS) has recently set new standards for this type of operations and demonstrated what can be achieved when data are collected according to the most stringent scientific methods.

The development of a distributed pan European research infrastructure, will support comparative research, not only by enabling data dissemination but mainly because it will support the researchers with a frame for secondary data production and analysis, data harmonization and supplementary documentation

Data protection within official statistics is in many cases an obstacle for the social researchers. Statistical confidentiality guarantees that individual data will not be made known to third parties. Agencies of official statistics are allowed to process non-personalized data, mainly because the statistical offices are effectively able to ensure the integrity of their computers and buildings (so nobody can steal data) and their integrity of their personnel (by means of

professional code of ethics and special legal enforcement in their work contracts). Scientists do not have the same legal status as official statisticians. They are not allowed to analyse statistical data provided by official statistics which are "personalized". Anonymisation is therefore a very important issue. Also CESSDA must discuss the possibility of establishing a legal status of "research data confidentiality" both on the national and international levels. One objective of such a legal status is to ensure that scientists can be especially trusted to process sensitive data under specific conditions.

Our modern digital societies are also producing a wealth of new data that up to recently hardly has been the object of scientific research. These data after the appropriate processing may also be preserved and or disseminated by research archives.

## **2.2 Interoperability**

One major objective is to provide seamless access to data across repositories, nations and research purposes. Data generated for one purpose should be open for use in many ways. They should consequently be created, described, and preserved in ways that facilitate use for a variety of purposes. The use of standards is crucial to the sharing of data. The culture of standards has shown to be weak in the social sciences.

To encourage data sharing and interoperability across communities and software systems, standardization of metadata and data are needed and should be encouraged. For social sciences data important initiatives and efforts are however underway on an international level, most notably the Data Documentation Initiative (DDI) carried out by an international alliance of data producers and data providers.

It is important to note that although the creation of standards is very important is not enough. The new standards must be based on new data models capable for the support of comparative research. Some important steps towards the development of new data models have been made by EU-funded projects, like the MetaDater project. The development of a conceptual data model for comparative research will support further developments in research infrastructures.

## **2.3 Central Data access/ location services**

In the majority of European countries data are preserved and made available for the research communities by national repositories. It is of crucial importance that these data resources can be found and even accessed through central web-based cataloguing and location services. This does not imply that data as such should be centralized in a "European Data Archive" – only that the knowledge of similarly structured data and metadata are made available through a virtual and central access point.

## **2.4 Harmonization of data access policies**

One major obstacle to access to empirical data in Europe is produced by the multitude of data access policies and regulations implemented by the national governments. To make it easier to bring together data from various countries mapping and even harmonization of regulations are needed. CESSDA plays an important role in the harmonization of data access policies.

### **3 A new role for CESSDA**

Social sciences are hampered by a raft of problems caused by the current state of research infrastructures at European level. This has severe consequences for the possibilities of conducting comparative European empirical research. Against this background a European strategy is needed that addresses the current problems. The goal of this strategy is to establish a virtual research infrastructure for social sciences, which will build upon existing resources, and both actively and systematically promote synergy and coherency. The general idea underlying this new strategy is to progress from the current situation characterized by isolated patches of research infrastructures to a situation where research infrastructures and data at European level are knit together to form a coherent and integrated system.

One important issue is to address the main actors that will realize such a strategy. CESSDA as a consortium of European data Archives will play an important role for two main reasons: The first is that CESSDA members cooperate for quite a long time towards the goal of building a distributed research infrastructure. The second is that CESSDA members already play an important role in their home countries serving open access to data for a number of research organizations at each country. But CESSDA has also one disadvantage. The development of a distributed social science infrastructure is a multidiscipline scientific attempt. Not only data providers but data producers and data analysts must be strongly involved on a permanent base. This means that although CESSDA is the only European actor capable for the realization of this strategy, it must change. Data producing and analysing organizations must also become CESSDA partners, and CESSDA must act in the future as a research organization.

The vision of the new strategy is rather straightforward, yet nonetheless highly ambitious: to provide the optimal conditions for creating and circulating data and services within social scientific research at European level and thereby facilitating scientific excellence.

This vision is to be materialised into a concrete research program to create a coherent, comprehensive, integrated and distributed research infrastructure for the social sciences – with a view to enhancing the existence, accessibility and comparability of data at European level.

The development of the new infrastructure will be guided by four main principles:

- The facilitation of access to and sharing of existing European and national data, thereby more efficiently and effectively linking data resources already available.
- The development of improved data models, standards and documentation relating to existing European and national data in order to enhance the scientific quality of data and their potential for interoperability.
- The generation of new and genuinely European data. This will involve both the collection of new data and the digitalisation of materials not currently computerised.
- The provision of research training programmes for the next generation of researchers.

The new research program must be implemented in a number of research projects. It will strengthen interdisciplinary and cross-border collaboration and comparative research on a European dimension. Further, it will enhance the building of research infrastructure capacity in the less resourced European countries. Finally, it will increase the opportunities for improving knowledge on social processes and thus holds great potential in terms of advising European and national policy-makers on how to manage the challenges currently being faced by the societies of Europe.

I hope that the works of this expert seminar will help us to make one more effective step towards our goals. It will help us to open access to the European data to the European research community.

Thank you