



Data collection, dissemination and access issues

Work Package 10 of the 2008-2009 CESSDA PPP under the 7th Framework Programme
(See <http://www.nsd.uib.no/cessda/project/index.html> for information on the CESSDA PPP.)

Accessibility to data for research purposes is essential to promote a science-based understanding of society and an informed discussion about policy issues at the local, national, and European levels. To improve access to the data resources that social scientists need for high quality research, CESSDA members have, for many years, offered networked mediation services between data producers and the research community. On behalf of CESSDA, the WP10 team is working to create the conditions necessary to reinforce CESSDA's capacity to act as data 'brokers', negotiating European-wide data access agreements and providing wide-ranging data services for researchers.

In order for an upgraded CESSDA to act as an effective and comprehensive conduit for data access, it is necessary to assess the expectations and needs of all stakeholders, including data producers and providers, member Data Archives, and researchers. It is also important to take into account the complexity and unevenness of today's data landscape, largely arising from the legal, institutional, cultural, and linguistic differences that exist among European countries.

It is the primary purpose of WP10 to deal with relationships between CESSDA member Archives on the one hand, and external stakeholders (data producers and providers as well as researchers) on the other hand, in the context of the current European data situation. WP10 is aiming to assess existing relationships and to make recommendations for subsequent policy action. To achieve this we are focusing on three issues:

1. Access to governmental statistical microdata, also taking into account concerns, often raised by National Statistical Institutes and other public-sector data producers, about statistical disclosure control;
2. Needs and requirements for archives' data collection strategies;
3. Data dissemination and preservation problems posed by multiple data formats and data types.

Three working teams have been formed within WP10 to address the specificities of each of these issues separately; regular exchanges between these teams ensures information-sharing and coordination, in view of the common objective of exploring the potential and possibilities for an upgraded CESSDA acting as a central intermediary for data access in Europe.

Access to governmental statistical microdata

Research in the social and economic sciences, as well as public policy evaluation, increasingly require access to data from official sources –including Eurostat, National Statistical Institutes, and other governmental producers of data, both at country level (ministries, central registries, etc.) and at sub-national level (regional statistical offices, counties, towns, etc.). In particular, researchers' demands for governmental microdata has surged in recent years, boosted by methodological advances. Data producers and providers are addressing these demands nationally, and bilateral negotiations with the scientific community have resulted in significant improvements in some countries, often through the mediation of a CESSDA Archive. However, many problems are common and could be better managed through a collaborative framework at the European level.

In this context, WP10's efforts are directed at fostering improvements in the relationships between governmental data producers, Data Archives, and the research community, at both national and European levels. As a starting point in working towards this goal, WP10 is conducting a country-by-country study of present-day access arrangements to official national and European microdata, distinguishing non-disclosive data from confidential disclosive data, and aiming to identify barriers to access for researchers in each case. Recommendations to facilitate greater access will then be prepared, with the focus on how an upgraded CESSDA can contribute to providing viable solutions. The necessary information is mainly obtained from two surveys, one targeted at Data Archives and the other at Eurostat, National Statistical Institutes, and other governmental data producers.

The team is also involved in the preparation of a workshop on Microdata Access hosted by Eurostat, in an effort to bring together National Statistical Institutes, Data Archives, and representatives of researchers to discuss existing access arrangements and availability of official statistical data in European countries, so as to identify recent progress, best practices, and areas for possible improvement. On this basis, the meeting aims to explore directions for further development and future policy measures.

Data collection strategies

Social scientists' demand for data has soared in recent years, with rising expectations in relation to both quantity and quality of data. In order for CESSDA to address these demands effectively, it is important to make sure data collections cover the needs and requirements of researchers in various disciplinary fields. To achieve this, it is essential to encourage deposit by multiple data producers, and also to explore ways through which CESSDA may facilitate researchers' access to data collections that exist outside of it.

As a preliminary step in this direction, WP10 will review current CESSDA data collections in order to identify strengths, weaknesses, and areas of expertise across the various member organisations, and how they relate to external data collections. A detailed map of what CESSDA organisations are in terms of their data collections requires identification of data sources (not only official statistical agencies but also international and European organisations, academia, private-sector companies, etc.), of types of data holdings (micro/macro, qualitative/quantitative, etc.) and of the fields of study to which these data relate (economics, demography, sociology, political science, etc.), as well as information on deposit conditions and deposit licences. To gather information, the working team is collecting and organising a large amount of material from the World Wide Web, and is now analysing the results of a survey distributed to Data Archives. Results from this study will then be used to produce guidelines for informing the future data acquisition and collections strategy, including recommendations on how an upgraded CESSDA might interoperate with data collections available from other providers.

Data preservation and data formats

Data preservation is a key issue with significant potential impact on dissemination, by determining the availability of data long after they have been collected. Yet today's multiple and rapidly changing data formats raise new curation problems. A closely related issue is the expanding number of datafile types and information formats, with a surge in non-standard data formats such as text files, spreadsheets, databases, audio and video files.

In order for an upgraded CESSDA to address these issues effectively, WP10 is currently preparing an audit that focuses on the amount of non-standard data files that member organisations are currently receiving, and the strategies that they use to preserve them; e.g. conservation in their original format, migration into a more standard format, or other form of conversion. The necessary

information is mainly being retrieved from a survey distributed to CESSDA members and other archives. On this basis, WP10 will make recommendations relating to the construction and maintenance of a CESSDA-wide file format registry, and will critically assess the measures that will need to be taken in order to adequately acquire, preserve, and disseminate the new forms of data.

Participation in Work Package 10 (contact persons at the national partner archives)

The project team of WP10 is comprised of staff from different CESSDA members.

Contact persons at each organisation:

Paola Tubaro (CNRS-RQ, Work Package Leader); Paola.Tubaro[at]ens.fr

Susan Cadogan (UKDA); susan[at]essex.ac.uk

Dimitra Kondyli (EKKE); dkondyli[at]ekke.gr

Vigdis Kvalheim (NSD); Vigdis.Kvalheim[at]nsd.uib.no

Heiko Tjalsma (DANS); Heiko.Tjalsma[at]dans.knaw.nl

Peter Farago (FORS); Peter.Farago[at]fors.unil.ch