Title	**WP7 Final report: Widening CESSDA: Inclusiveness and Comprehensiveness of the upgraded European research infrastructure (D7.1)**
Work Package	WP7
Authors	Brigitte Hausstein (GESIS); Janez Stebe (ADP); Adrian, Dusa (RODA); Zoltan Fabian (TARKI); Peter Hegedus (TARKI); Karl Müller (WISDOM)
Source	WP7 Papers on tasks 7.1-7.13, cooperation with WP3 and WP6
Dissemination Level	PU (Public)

Summary/abstract

This is the final report of work package 7 (WP7), which focused on plans for widening an upgraded CESSDA. The work in WP7 focused on two overarching objectives aimed at ensuring that an upgraded CESSDA will be both more inclusive and comprehensive in terms of the research community that it services. Accordingly, the report is divided in two parts.

The first part deals with strategic measures to bring in new national data resources into an upgraded Research Infrastructure. It presents the main findings of an analysis of existing potential and possibilities for the extension of the RI. Based on these findings a roadmap for widening CESSDA was developed. Within the framework of an action plan specific supporting and training activities as well as the implementation of a Virtual Centre of Competence (VCC) are recommended to the Management Board.

The second part focuses on strategies for the inclusion of data resources that currently reside in organisations and repositories outside of the existing CESSDA network. A list of relevant resources is provided and a specific programme for interoperating with identified agencies and organisations is recommended.

Both parts include recommendations that are relevant for the overall strategic plan for the upgrading of CESSDA.

Contents

Executive Summary

Besides the need for strengthening and developing capacity in the existing CESSDA organisations, which is the focus of WP6 of the CESSDA PPP, there is also the need for widening participation in the upgraded cessda-ERIC (European Research Infrastructure Consortium) both directly by fostering membership in new countries and indirectly by deepening involvement and extending the RI to agencies and organisations which remain outside of the current network yet continue to host important data collections. Thus the work in WP7 focused on two overarching objectives aimed together at ensuring that an upgraded RI is both more **inclusive and comprehensive** in terms of the research community that it services.

1. Inclusiveness of an upgraded RI

The report presents the main findings of the analysis of the existing potential and possibilities in order to extend the RI. Twelve countries from Eastern Europe were included in the analysis. It turned out that most of the countries, except Lithuania, will not be able to participate in the first round of the membership application.

In response to this it is suggested that a roadmap for widening the upgraded RI be implemented. It should cover a period of ten years (2010-2020) and guide the process of future widening.

In order to address the identified needs of potential new data resources the implementation of an action plan is recommended. The action plan comprises specific supporting and training activities. Special emphasis is laid on the installation of training and exchange programmes and the setting up a cessda-ERIC Virtual Centre of Competence (VCC). Furthermore a concept of minimum requirements that new data resources should meet has been developed.

Recommendations 1-23

Recommendation 1

The upgraded RI should develop a common strategy on data sharing in order to support young and emerging data archives in their role as an agent of the change in their national data sharing culture. It should promote and support the implementation of national data sharing policies and initiate a European resolution on data sharing.

Recommendation 2

The upgraded RI should provide active support for potential new data resources through the mobilisation of national support to secure sufficient and long term funding for infrastructure services.

Recommendation 3

The upgraded RI should develop a concise policy statement about pathways concerning the implementation of standards. Best Practice rules, minimum

requirements of setting up and running a data archive, guidelines for quality assessment, and a quality label should be provided. This should include a clear statement on the future requirements and development of the Data Portal.

Recommendation 4

The upgraded RI should provide training facilities that consider the specific circumstances of new and emerging data resources: training for data archivists, training on specific tools and standards, workshops for beginners, staff exchanges, and summer schools.

Recommendation 5

We recommend implementing a roadmap for developing the membership of the upgraded RI. The proposed period of time should be 2010-2020 and the process should be accompanied by a biennial update.

Recommendation 6

We recommend allocating a dedicated budget for the update of the roadmap, which should include funding for travel, the compilation of national data audits/data screening reports and meeting organisation. Besides funding we suggest that resources should be allocated for a Working Group on Widening. The working group should be led by a coordinator and assist the new ERIC's director in organising the audits and developing future strategies on widening the organisation.

Recommendation 7

We suggest putting the following countries and potential data resources on the Roadmap for widening: Latvia, Lithuania, Estonia, Poland, Slovakia, Bulgaria, Croatia, Serbia, Macedonia, Russia, and Ukraine. (See table 2).

Recommendation 8

In order to assist and support potential new national data resources in achieving the proposed membership criteria we recommend implementing an action plan for the next three years (2010-2020). (See table 3)

Recommendation 9

Based on the recommendations of the CESSDA-PPP Training Programme we suggest including potential new data resources in the proposed Summer School for Data Scientists, Workshops for Beginners, Visiting/Staff Exchange Programme, and Expert Seminars.

Recommendation 10

The establishment of a Staff Exchange and Visiting Programme should be undertaken in a considered manner. The programmes should be structured, and backed by financial and organisational conditions favourable for both guests and hosts.

Recommendation 11

In addition to the financial sources of the Training Programme we recommend establishing a dedicated Capacity Building Fund within the ERIC's central budget in order to facilitate full participation of new personnel in the proposed Training Programme.

Recommendation 12

In accordance with the planned Training Programme we recommend setting up two training teams and five expert groups, which consist of experts who should contribute to the planned Training Programme.

Recommendation 13

The VCC should become an integral part of the planned training programme in order to provide an information and management platform for future data resources, which consider their special needs.

Recommendation 14

The management board of the upgraded CESSDA RI should ensure that members/service providers will contribute relevant information to the VCC. We recommend including this as a requirement in the service level agreement.

Recommendation 15

The VCC should support all levels of membership and serve the process of working on fulfilling the membership criteria by full members and designated members, as well as supporting non-members (by providing general information).

Recommendation 16

The VCC should be accessed via the Portal and should be incorporated in the current Content Management System (CMS).

Recommendation 17

The login system should be based on the same middleware as currently used by CESSDA, and details regarding operating system, web server, CMS and database system should be agreed upon construction.

Recommendation 18

The new CESSDA RI should set up a VCC working group with duties ranging from updating web pages with fresh information to answering specific questions in the dedicated forums. This working group should be supported by the proposed expert groups.

Recommendation 19

The Forum should employ a user-friendly search engine in order to help users find information of interest.

Recommendation 20

The VCC should employ a teleconferencing system in addition to the forum web pages.

Recommendation 21

Minimal requirements for new data resources should be:

(1) Data should be discoverable over the Internet. Service providers should ensure network connectivity through appropriate national, regional and local area networks;
(2) Information on data holdings should be published in a structured catalogue, available for harvesting by the Portal, and data should be downloadable through common data gateways;
(3) Service providers should comply with agreed interoperable metadata standards. Specifically, to document their data with at least the mandatory items in the DDI standard and publish the metadata files in XML format;
(4) 'Local' language should be included and maintained in the multi-lingual thesaurus. Index terms from ELSST should be used to enhance resource discovery;
(5) ASCII format should be employed for the long-term preservation of data files.

Recommendation 22

Where cost is an issue, new data archives should turn to open-source software and where the ease-of-use is an issue, they should turn to commercial equivalent software.

Recommendation 23

In order to support potential new data resources in fulfilling minimum requirements an upgraded RI should provide guidelines, best practices, a toolkit and specific training.

Estimated resources needed for the Action plan (PM/per year)

Director	1PM
Widening Coordinator	6PM
PR Manager	3PM

IT Manager	3PM
Training Coordinator	2PM
Working Group on Widening	10PM
Training groups	2PM
Expert groups (training, VCC)	60PM/first year 30PM/second year
External experts	5k Euro
Budget: PR, travel, audits, surveys etc.	10k Euro
Capacity Building Fund (outreach)	10k Euro
Training facilities	5k Euro

2. Comprehensiveness of the upgraded RI

In building up an integrated and comprehensive upgraded RI it is necessary to develop a strategy whereby data holdings from within the member organisations and those held externally can be simultaneously accessed and integrated. This report gives recommendations for a programme for interoperating with these resources.

The proposed strategy and actions have been developed for a term of five years. **It is recommended** that these be split into a construction phase (2010-2011) and an operational phase (2012-2014), to implement the programme.. However, both the strategy as well as the actions should be reviewed and updated periodically.

Based on a comprehensive list of potential resources four groups are identified. According to these groups a specific interoperating programme has been suggested. The recommendations include various modes of interoperation: offering (affiliated) membership, applying for membership in data providing organisations (e.g. the Interuniversity Consortium for Political and Social Research, USA: ICPSR), the exchange of metadata, cross references, data redistribution licence agreements on a contractual basis, the purchase of licenses (national, European), facilities for self archiving and training, special cooperation with national research councils, and support for setting up national cooperation bodies for the purpose of associating all national data providers with the proposed ERIC.

In order to realise the programme it is recommended that a dedicated budget is allocated for interoperating and planning capacities for a number of coordinating tasks. These tasks should be carried out in the framework of the permanent working group “Interoperating with external data resources” and comprise the management of the relationship to external resources (including legal issues), technological development of an ERIC Gateway to external data, setting up of a web data server, implementation of an authentication system as well as the organisation of training facilities. In order to devise the acquisition strategy a sample acquisition policy for external sources was also developed.

Recommendations 24-30

Recommendation 24

Based on the overall aim of interoperating with external data sources we suggest implementing a set of actions in order to address the existing barriers (see table 5).

Recommendation 25

In order to implement the strategy of interoperating we recommend a specific programme of interoperating (see table 6).

Recommendation 26

We recommend planning a dedicated budget for interoperating with external data sources.

Recommendation 27

Besides funding we suggest allocating dedicated capacity for a number of specific coordinating tasks. These tasks should be carried out within the framework of a permanent Working Group on Interoperating.

Recommendation 28

The proposed strategy and actions have been developed for a medium length term of five years (2010-2014) whereby we recommend a construction phase (2010-2011) and an operational phase (2012-2014) for implementing the strategic programme.

Recommendation 29

The strategy and proposed actions should be reviewed and updated periodically. We recommend a yearly update.

Recommendation 30

We recommend implementing an Acquisition Policy with external data sources.

Estimated resources needed for interoperating (PM/year)

Director	0.5 PM
Coordinator for interoperating	12PM
IT manager	6PM (first year) 3PM
Working Group on Interoperating	6 PM
Budget for interoperating	10k Euro
Training facilities	5k Euro

Introduction

Besides the need for strengthening and developing capacity in the smaller existing CESSDA organisations, which is the focus of WP6 of the CESSDA-PPP, there is also the need for widening the participation in an upgraded RI, both directly by fostering membership in new countries and indirectly by deepening involvement and extending the RI to agencies and organisations which remain outside of the current CESSDA network yet continue to host important data collections. The main focus of the work of WP7 was strategic planning to bring in new national data resources into the upgraded RI and the development of a strategy whereby data holdings from within the member organisations and those held externally can be simultaneously accessed and integrated.

Thus WP7 focused on two overarching objectives both aimed at ensuring that an upgraded CESSDA is both more inclusive and comprehensive in terms of the research community that it serves:

Objective 1 - To extend the existing CESSDA RI and to foster the development of national data archiving initiatives in those countries which are not currently part of the CESSDA network, in order to create and maintain a 'complete' pan-European RI, including representation from emerging and candidate countries.

Objective 2 - To extend the network to agencies and organisations, which remain outside of the CESSDA RI, yet continue to host important data collections.

WP7 worked, in close co-operation with WP3 and WP6, on outlining recommendations for the proposed administrative bodies in relation to strategic measurements and potential financial contributions from within and outside the CESSDA network for incorporating new members into the planned cessda-ERIC. The results of the work are contributing to the overall strategic plan for the upgraded RI and the new organisation's membership requirements.

Part I: Inclusiveness: Extending the upgraded RI

1. Identification of relevant organisations

A central focus of WP7 was to explore the potential and possibilities that exist in order to extend the current CESSDA membership, therefore increasing the availability and accessibility of important national data collections.

The work built upon the experience gained in previous co-operation with more than 15 associated partners in many emerging member states initially formed during the UNESCO Workshop on Social Science Data Archives in Eastern Europe 2002 (formerly referred to as EDAN - the East European Data Archives Network) and coordinated by the GESIS Leibniz-Institute for Social Sciences. EDAN acted as an informal network designed to unite data archive projects/initiatives, and data archives, which are at an early stage of their existence and share common problems; as well as ensuring that Eastern European data archives were given opportunities to develop themselves in line with the advanced western data archives. For those archives, which were not formal members of CESSDA at that time, EDAN served as a good organizational framework to co-ordinate and support the efforts for the setting up of archives.

Figure 1: Potential new cessda-ERIC members (in blue/red)

Based on the preliminary work within EDAN the strategic planning of the widening activities in the CESSDA-PPP concentrated on data archives and data archive projects from 12 countries: Latvia, Lithuania, Estonia, Poland, Slovakia, Bulgaria, Croatia, Serbia, Macedonia, Belarus, Russia, and Ukraine (see figure 1). The identified data archives and data archive projects became associate partners in the project (see table 1). Montenegro, Bosnia-Herzegovina, Albania (potential candidate EU member countries) and Moldova (a non-EU country) were excluded as the political and economic situation is rather tense and the circumstances for social research in these

countries are unclear. The same applies to Turkey, Cyprus and Malta. The data archive initiatives in Belgium and Portugal were considered in the information collection phase.

Table 1: Data archives and data archive projects included:

Country	Data archive/project	Institutional Affiliation
Lithuania	LiDA, Lithuanian Data Archive for Social Sciences and Humanities http://www.lidata.eu/en	University of Kaunas
Estonia	ESSDA, Estonian Social Sciences Data Bank	University of Tartu
Latvia	LSZDA, Latvian Databank of Social Sciences	Academy of Sciences, Riga
Slovakia	SASD, Slovak Archive of Social Data http://sasd.konzum.sk/	Slovak Academy of Sciences, Institute for Sociology, Bratislava
Poland	ADS, Polish Social Data Archive http://www.ads.org.pl/	University of Warsaw, IFIS, Polish Academy of Sciences
Bulgaria	Social Science Data Archive Project http://www.reglo-bg.org/	REGLO, Sofia
	IS BAS, Institute of Sociology http://sociology-bg.org/display.php?language=en	Bulgarian Academy of Sciences, Sofia
Yugoslavia	NEDA, National Social Sciences Empirical Data Archive Project http://www.cpijm.org.yu/	Center for Political Studies and Public Opinion Research, University of Belgrade
Macedonia	CRPM, Center for Research and Policy Making, Skopje http://www.crpm.org.mk/	
Russia	RSDA, Russian Sociological Data Archive http://www.socpol.ru/eng/archives/about_proj.shtml	Independent Institute for Social Policy (IISP), Moscow
Ukraine	KIIS-DB http://www.kiis.com.ua/index.php?id=15&sp=1&lng=eng)	Kiev International Institute of Sociology
	Ukrainian Sociological Archive of the Institute of Sociology	National Academy of Sciences of Ukraine and Kiev National Taras Shevchenko University, Ukraine
Belarus	Department of Philosophy and Social Sciences	Belarus State University, Minsk
		State Committee on Science and Technology of the Republic of Belarus, Minsk

Methods of information collection

a) Country reports and workshops

In order to collect country specific information and to analyse the needs of the potential new partners and their expectations regarding the future RI a one-day workshop with potential new data resources was organised during Summer 2008.

In advance of the workshop representatives of the invited data archives and data archive projects were asked to compile a country report which made reference to the following topics:

- National data landscape, data accessibility and data sharing culture;
- Funding for data access and infrastructure services;
- Organisational/institutional setting of the data archive/data archive project;
- Capacities and human resources of the data archive/data archive project;
- Services and facilities offered by of the data archive/data archive project.

Based on these reports the WP7 team compiled a comprehensive analysis of the current situation in all involved countries and presented the main findings to the workshop (Hausstein et al. 2009). The discussion during the meeting concentrated on the national data sharing culture in these countries, prospects for further development of the data archives/data archive projects, and their potential contribution to the planned upgraded CESSDA RI. Special emphasis was given to the fact that there is a large need for training and knowledge transfer (best practice, standards, and tools).

b) Onsite visits

In addition to the workshop and country reports a series of onsite visits were carried out. These visits aimed to meet with the associated partners of the CESSDA-PPP and also with relevant government-level funders in order to inform them about the project and raise awareness of the importance of cross-national research-led data infrastructures. By highlighting the ESFRI process during these meetings we raised awareness of the importance of governmental funding bodies prioritising long-term commitment to the support of social science data access, archiving and curation.

Beyond that the meetings were organised in order to gain a greater understanding of how the data archive and data archive projects operate, as well as how they may gain from, and what they could offer, a future European RI. The meetings included the organisation's director/head, representatives of any parent body (institute, faculty, university, academy), and national funding bodies. In addition, researchers who are producing and/or using empirical social science data in these countries were invited to offer a researcher's perspective/outline their requirements (Hausstein/Schürer 2008 and Hausstein/Schürer 2009.)

2. The role of national data sharing culture

The purpose of the analysis was to assess the current situation in the countries that are not yet part of the current CESSDA organisation, and to emphasise the country specific conditions which may affect future integration into the RI. The assessment was carried out on the basis of the delivered country reports and implies a country-by-country evaluation of the following issues:

- Conditions that affect the existing and future relations between data archives and data depositors;
- Relations that guarantee wide access to high quality data produced by the academic, public and commercial sector;
- Obstacles to data access that are deeply rooted in traditional data sharing cultures;
- Current provision of national infrastructure services, access conditions and ways to improve this;
- Level of demand for data for secondary analysis from the academic community and obstacles to an increased demand.

The analysis focuses in particular on efficient models and success stories that show the fundamental role of the archive as an agent of change with regard to national conditions of data access. Even if the current situation of the majority of the data archives and projects is not bright, there are feasible ways for its improvement. In this respect the relationship between the data archive and the main producers/providers of research data - with a special emphasis on continuing collaborative national topical projects (e.g. elections, crime, and general social surveys) - is of key importance.

The model of an integral data infrastructure is rooted in the fact that the research community often produces academic data with the intention to stimulate wide use. This implies a collaborative approach, where the future potential users are considered just as important as primary researchers. The relevant aspects of the rights to access the data for secondary use and certain level of quality requirements should be taken into consideration at the very beginning of a research project. So far there are no data producing projects facilitating this kind of access to secondary usage. However, there are opportunities for data archives to take a more active role, e.g. in the promotion of collaborative approaches in public research funding policy, where data producing and data distributing organisations are working hand–in-hand.

Nevertheless, the data archives and projects in all countries have a solid basis for their future activities in this field thanks to a long and firm tradition of existing empirical social sciences research. There are good relations and contacts with the major national long-term research projects or organisations. The idea and models of building a modern national data infrastructure were often inspired whilst collaborating in international comparative research projects such as the International Social Survey Programme (ISSP), or the European Social Survey (ESS) which aimed at providing data for a wide secondary use (as described in the paragraph on institutional settings).

The access to raw data produced by the governmental statistical offices is, in almost all countries, still an issue. This probably reflects a still existing state or governmental monopoly of possessing information. This prevents the research community from

using such sources by building access barriers that are practically insurmountable. A coordinated European approach to tackle the problem from a common perspective would help to change the situation.

Regarding data from the commercial sector, which showed a marked development after 1990, highly selective access can be observed. It seems that there is only a slim chance for future cooperation between commercial and academic projects.

One of the conditions for a well established academic data infrastructure is the development of good relationships and cooperation among research and service activities. So far there are no references to existing problems or any competition in this respect. It might be the case that potential problems arise once a particular data archive or project gains additional recognition and financial support for its infrastructural activities which could be seen as a threat to other substantive research activities.

Still, the provision of national infrastructure services and improvements to data access conditions are generally seen as problematic due to the lack of data sharing culture within the research community. There are psychological, research-tradition and copyright barriers to overcome, as well as the recurring tendency to privatise data and not sufficiently document research methodology. In these respects data archives need to act as agents of change in the data sharing culture while still taking care of their basic existence. There is a need for common strategy: a top-down approach. Drafting such a strategy would require negotiations with national official research policy representatives concerning the benefits of a publicly funded data archiving infrastructure for the national and international research community. Additionally, new models and practices are needed, including international declarations and petitions on free access to academic outputs (e.g. the OECD Declaration, Berlin Declaration on Open Access etc.).

Another lesson learnt from the analysis is that the demand for data, the dependency on research traditions, the level of statistical literacy, and the ease of obtaining funding for primary data gathering projects is not easy unless a data archive is sufficiently established and able to proactively promote the use of its data.

In conclusion, there is still an underdeveloped data sharing culture in the Eastern European countries. However, a data archive could serve as an agent of change in the data sharing culture and highlight the need for data. By working with data users and funders of data producing projects data archives have the chance to contribute to building up the research community. The data archives could show the benefits for both users and funders. In this respect data archives have to produce convincing arguments for both data producers and funders to invest in data archiving.

Recommendation 1
An upgraded RI should develop a common strategy on data sharing in order to support the young and emerging data archives in their role as agents of change in their national data sharing cultures. It should promote and support the implementation of national data sharing policies and initiate a European resolution on data sharing.

3. Capacities and needs of potential new data resources

Funding

Analysis shows that most of the organisations and projects are under-funded or have no financial resources at all (see figure 2). Even established data archives are not funded for their archival work. The Polish data archive (ADS) and SASD in Slovakia are supported within the framework of inter-institutional projects related to the comparative survey program ISSP. The few projects (LSZDA and the data archive project in Croatia) with dedicated funding sources are mainly supported by small grants provided by a ministry, university or academy. Often these are several small sources, which are not sufficient to expand the services of the data archive any further.

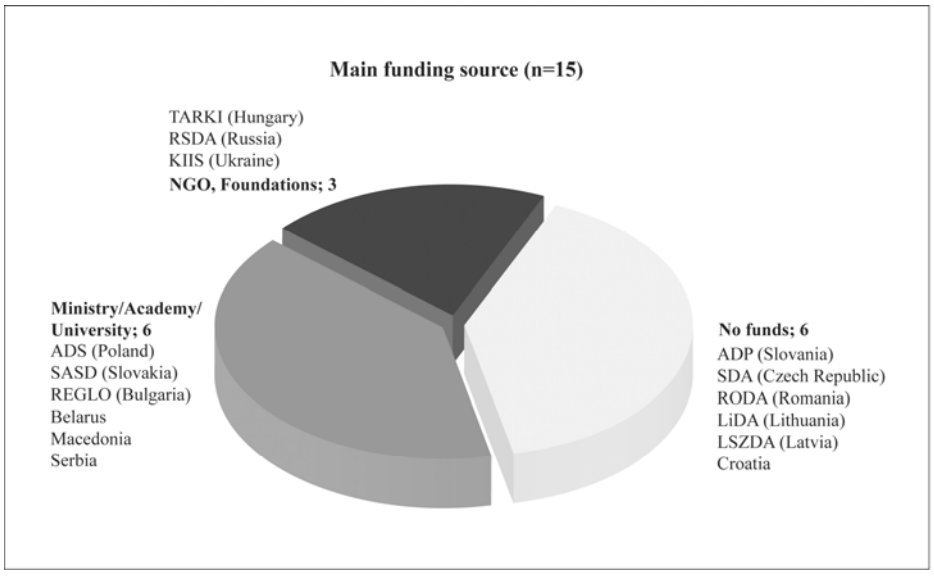

Figure 2: Main Funding Source of the Data Archives/Archive Projects (Source: Report on the CESSDA-PPP Workshop)

There are two international foreign foundations ('Ford', 'McArthur') that are supporting the work of the Russian sociological data archive (RSDA). The Lithuanian data archive (LiDA) received a short-term grant from the 7th EU Framework Programme (FP7) in 2006. The archive initiative at REGLO (Bulgaria) and NEDA (Serbia) received similar support via a UNESCO grant for 2002-2004 and 2006-2008 respectively. However, in general it turned out that foreign funding did not play an important role when it came to long-term funding. It is worth noting that EU support is marginal.

Institutionalisation and human resources

The funding situation is reflected in the organisational settings of the analysed organisations and projects. While there are a significant number of data archives that are already part of an organisational structure, the majority are organised as informal or project groups, with most of their members being voluntary and only paid for work in related research projects (see figure 3).

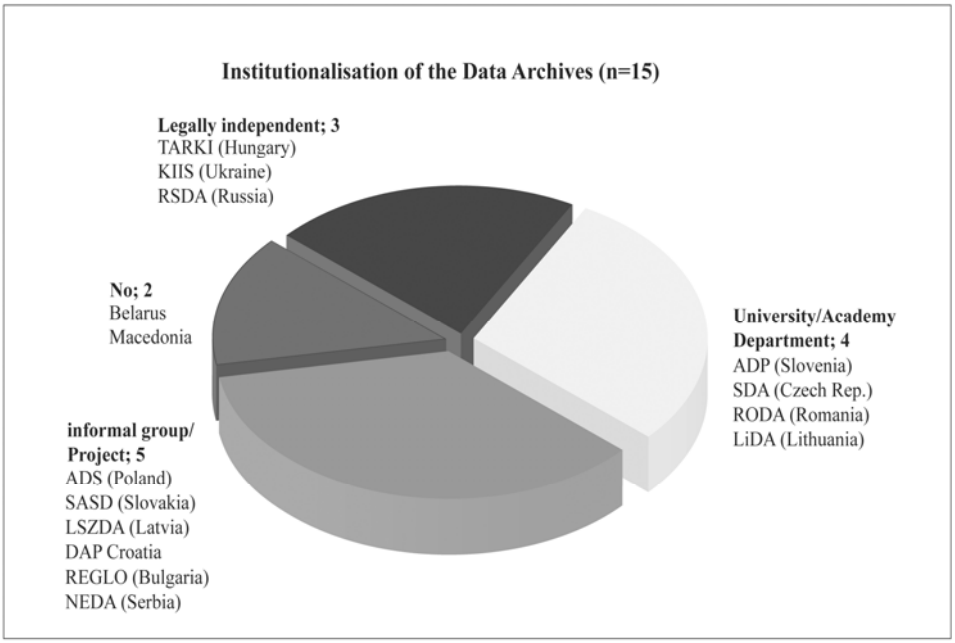

Figure 3: Institutionalisation of the data archives (Source: Report on the CESSDA-PPP Workshop)

The institutional affiliations of the data archives and projects vary from country to country. There is no standard model for the institutional setting of a successful data archive. Groups of affiliations are as follows:

- University (LiDA, Lithuania; Croatia);
- Academy of Science (NEDA, Serbia);
- Co-operation between a University and an Academy of Science (ADS, Poland; SASD, Slovakia; LSZDA, Latvia);
- Independent organisation/NGO (RSDA, Russia; KIIS, Ukraine).

Most of the data archives are at an early stage of their existence (founded after 2000): this is why they are relatively small in size. Apart from the archive initiatives in Bulgaria, Macedonia, Belarus, and Serbia, which do not have any personnel, data archives and projects have four staff members on average (see figure 4).

However there are also success stories. In Lithuania, a very young data archive was successful in getting strong support from their national Ministry of Science and Education and thus could develop a solid basis for a data service (2006-2008). The Russian data archive received stable funding for the coming years through the grants provided by the international foundations ‘Ford’ and ‘McArthur’ and, thanks to this support, will continue to provide data services at a high level.

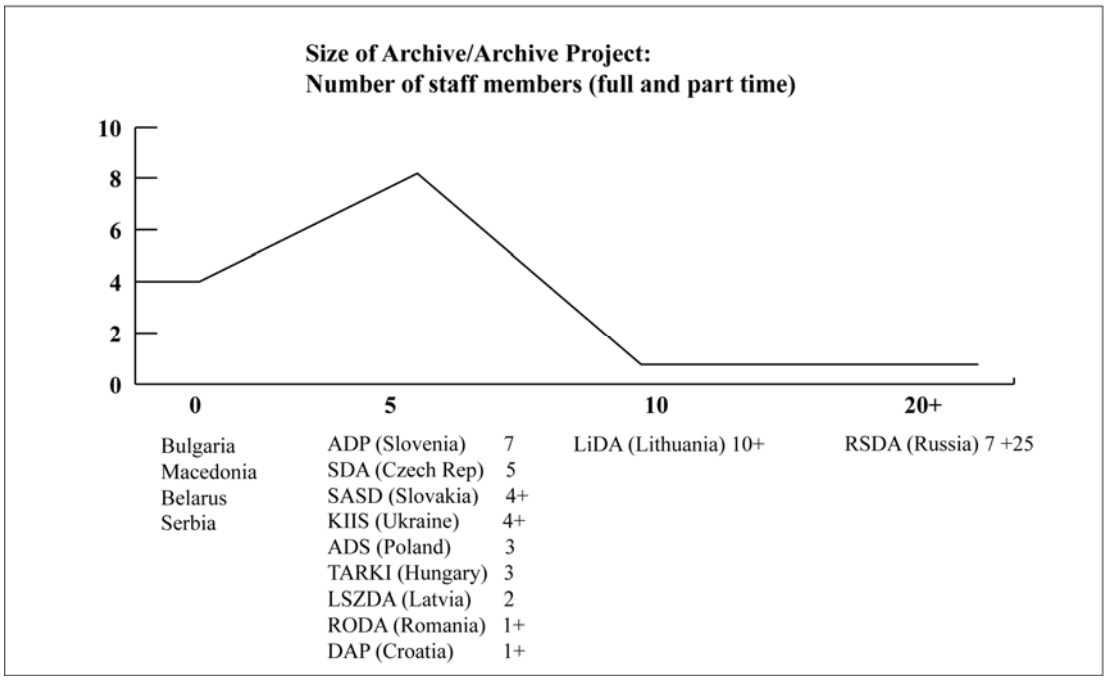

Figure 4: Size of Data Archive/Archive Project (Source: Report on the CESSDA-PPP Workshop)

On the whole, very small data archives and data archive projects with limited, short term national funding are most common in the majority of analysed countries. Although there are successful data archiving organisations in Lithuania (LiDA) and Russia (RSDA), most of the archives are working without any dedicated funding. They receive marginal support from national or comparative survey projects such as the ISSP, ESS, Polish General Social Survey (PGSS), or the European Values Survey (EVS). Furthermore, analysis showed a strong relationship between the amount of funding and a data archive's level of institutionalisation.

It will be necessary to mobilise national support to secure sufficient and long term funding for infrastructure services in the respective countries. However, the current funding policies in the countries of the region are not favourable for academic data archiving. Additionally the national funding bodies have little awareness of the role of infrastructure services for the social sciences. They have just begun to work out their national roadmaps for infrastructure and, unfortunately, social science research infrastructures are not a high priority.

Recommendation 2
An upgraded CESSDA should provide active support for potential new data resources by mobilising national support to secure sufficient and long term funding for infrastructure services.

Standards, tools and services

Beyond the basic requirements of running a data archive (the development of a national data sharing culture, sufficient funding and human resources) there are specific needs. These refer to the application of the existing archival standards and tools as well as to the use of service instruments. Standards are the result of the

professionalization of data archiving and the development of new digital preservation technologies. Currently there are quite few standards in the field of archiving digital social science data, e.g. the Data Documentation Initiative (DDI) and Open Archival Information System (OAIS). Even OAIS is described as "recommendations" that serve as a basis for specific standards. However, there are quasi-standards (best practices) and tools that are widely used in the professional communities, e.g. NESSTAR and the Dataverse Network (http://thedata.org/). There are many good reasons to apply standards in archival work. Specifically, standards support the interoperability of distributed data sources, thus ensuring pan-national access to data sources in organisations such as CESSDA.

There are big differences between the application of existing archival standards and best practices when comparing the data archives and projects. DDI is the most widespread standard among the archives. Five data archives are using this standard; only one archive operates a thesaurus (LiDA uses a HASSET based thesaurus). None of the organisations use the OAIS reference model, although some of them are thinking about introducing it.

There is a big demand for training seminars, guidelines and best practice on standards and related tools. Adaptations to the specific conditions of the archives, short versions of standards and a forum where experiences can be shared are needed to support young and emerging data archives in Eastern Europe.

Recommendation 3

An upgraded RI should develop a concise policy statement regarding the implementation of standards. Best Practice rules, minimum requirements of setting up and running a data archive, guidelines for quality assessment, and a quality label should be provided. This policy should also include a clear statement on the future requirements and development of the current CESSDA Portal.

Recommendation 4

An upgraded RI should provide training facilities that consider the specific circumstances of new and emerging data resources: training for data archivists, training on specific tools and standards, workshops for beginners, staff exchanges and summer schools.

4. Roadmap for widening the cessda-ERIC 2010-2020

Analysis has shown that most of the data archives and archive projects have to tackle a specific set of problems before they will be able to fulfil the planned full membership criteria of the upgraded RI. When planning future widening activities the new organisation should consider these circumstances and offer potential new partners different levels of, and therefore criteria for, membership. This would allow for inclusivity. Membership design should allow extension of the ERIC as a staged process.

Recommendation 5

We recommend implementing a roadmap for developing the membership of an upgraded RI. This roadmap should cover the period 2010-2020, following which the process should be accompanied by a biennial update.

In 2012 a dedicated workshop for potential new member countries and funders should kick-off a series of biennial meetings. These meetings should constitute the basis for a periodical update of the roadmap. The focus of the workshops should be: assessment of the development of data resources (including compliance with membership criteria), and the presentation of national data audits and data screening reports (identification of further data resources).

Recommendation 6

We recommend allocating a dedicated budget for the update of the roadmap, which should include monies for travel, the compilation of national data audits/data screening reports and meeting organisation. Besides funding we suggest capacities allocation for a Working Group on Widening. This working group should be led by a coordinator and assist the new organisation's Director in organising audits and developing future strategies on widening the upgraded RI.

Potential members of the suggested Working Group on Widening (WG Widening) could be drawn from GESIS, UKDA, ADP, and SDA.

Recommendation 7

We suggest putting the following countries and potential data resources on the Roadmap for Widening: Latvia, Lithuania, Estonia, Poland, Slovakia, Bulgaria, Croatia, Serbia, Macedonia, Russia, and Ukraine (see table 2).

Table 2: Roadmap for widening the proposed cessda-ERIC 2010 - 2020

Country	National Roadmap for RIs / commitment to upgraded CESSDA RI	Potential national service provider	Description	Size: Staff members / holdings	National funding	Membership criteria				
						Full		Associate		Affiliate
						designated	full	designated	associate	
Lithuania	2009/ **Yes**	LiDA, University of Kaunas	Data archive (set up 2006) providing a well advanced data service (website, NESSTAR online catalogue, thesaurus, DDI compliant meta data, training	10 / 66 national and international studies	Ministry of Education and Sciences		**2010**			
Latvia	Work in progress/ No	LSZDA, University of Latvia, Riga	Data archive initiative (set up in 2002) providing no data service (NESSTAR test server, DDI compliant metadata)	0 / national ISSP Studies since 2002	Small funding from the University	**2012**	**2015**			
Estonia	Work in progress/ No	ESSDA, University of Tartu	Data archive (set up in 1996) providing limited data service	1 / about 150 studies (originated before 1990)	No funding for data service	**2015**	**2020**			
Poland	Work in progress/ No	ADS, University of Warsaw and Polish Academy of Sciences	Data archive (set up in 2003) providing limited data service (web site and data catalogue, meta data based on DDI)	3 / 48 national and international studies	No funding for data service	**2012**	**2015**			
Slovakia	Work in progress/ No	SASD, Academy of Sciences of Slovakia, Bratislava	Data archive (set up in 2003) providing limited data service (web site, data catalogue, DDI compliant meta data)	4+ / 17 national and international studies	No funding for data service	**2012**	**2015**			
Bulgaria	Work in progress/ No	Bulgarian Academy of Sciences, Sofia	Data archive initiative at REGLO started in 2002, activities stopped because of lack of funding	0 / 20 national studies	No funding	**2015**	**2020**			

Country	National Roadmap for RIs / commitment to upgraded CESSDA RI	Potential national service provider	Description	Size: Staff members / holdings	National funding	Membership criteria				
						Full		Associate		Affiliate
						designated	full	designated	associate	
Croatia	Work in progress/ Unknown	University of Zagreb	Data archive project in its implementation phase (start 2008), cooperation with the library of the University	1 / -	Short term project funding			**2012**	**2015**	
Macedonia	-	CRPM, Center for Research and Policy Making, Skopje	Data archive initiative of a private research institute (since 2008), low support by national research community	- / -	No funding					**2020**
Serbia	-	NEDA, Institute of Social Sciences, Belgrade	Data archive initiative started in 2006, offline data base, DDI compliant meta data	- / about 200 national studies	No funding for data service			**2015**	**2020**	
Ukraine	-	KIIS – DB, Kiev International Institute of Sociology;	Data archive (set up in 2001) with limited data service (web site, KISS-DB online data bank, DDI compliant meta data)	4 / 140 national studies	Private funding (KIIS public opinion research institute)					**2012**
Belarus	Start of initial work	-	-	- / -	No funding					
Russia	-	RSDA, Independent Institute for Social Policy, Moscow	Data archive (set up in 2002) providing a well advanced data service, NESSTAR online catalogue, DDI compliant meta data, thesaurus, training courses for users	7+25 / about 600 studies	Private funding (FORD, Mc Arthur Foundation)					**2010**

5. Action Plan

Objectives and specific actions

Recommendation 8

In order to assist and support potential new national service providers in achieving ERIC the membership criteria set for the ERIC, we recommend implementing an action plan over the next three years (2010-2012).

The proposed action plan should be based on identified needs and focus on the following objectives:

(1) **Information** about the proposed ERIC;
(2) **Visibility** of data resources that are outside the new organisation;
(3) Assistance in mobilising **national support;**
(4) Provision of **Standards and Tools;**
(5) **Training** and knowledge transfer.

The action plan (table 3) indicates specific actions, responsibilities, time lines, and resources. Potential constrains and modes of evaluation are suggested. The proposed cessda-ERIC Widening Coordinator and WG on Widening should be responsible for the implementation and execution of the action plan.

Estimated resources needed for the Action plan (per year):

Director	1PM
Widening Coordinator	6PM
PR Manager	3PM
IT Manager	3PM
Training Coordinator	2PM
Working Group on Widening	10PM
Training groups	2PM
Expert groups (training, VCC)	60PM/first year 30PM/second year
External experts	5k Euro
Budget: PR, Travel, audits, surveys etc.	10k Euro
Capacity Building Fund (outreach)	10k Euro
Training facilities	5k Euro

Table 3: Action Plan 2010-2012

Objective	Activities	Responsibilities	Time line	Resources		Constraints	Evaluation
				available	needed		
1) Information on the proposed cessda-ERIC	Information on the new organisation and its membership criteria made available on the website	Director, central administration and services (PR)	set up until April 2010; ongoing	Website	• funding for PR activities • Strategy of promoting the new organisation, development of an information policy	no funding	number of page visits, annual online user survey
	Information brochures (in English, French, German, Russian) and dissemination	Central administration and services (PR)	2010		funding for PR activities and translation	no funding	dissemination, report
	Regular presentations at conferences on a European and national level (in the field of data archiving and services as well as SSH research)	CCentral administration and services (PR)	2010, ongoing		funding for PR activities, travelling	no funding	list of visited conferences, annual report
	Local information meetings with major national data providing organisations	CCentral administration and services (PR), WG Widening	2011, ongoing		funding for PR activities and travelling	no funding	list of visited organisations, annual report
	Newsletter	CCentral administration and services (PR)	first issue in April 2010; quarterly	Website	• funding for PR IT support (web master) • content provided from membership	no funding; missing content (membership)	number of page visits, annual online user survey
2) Visibility of data resources that are outside of the new infrastructure	Implementing and maintaining a News Page to introduce the work of emerging data service organisations outside of the infrastructure	CCentral administration and services (PR, IT), WG Widening	Set up in July 2010; content ongoing	Website	• funding for PR and IT (web master) • capacities of WG Widening	no funding; no content provided by new data resources	number of page visits, online user survey
	Annual Data Audits (reports to identify national resources) of countries which are not represented in the new	Director, WG Widening	First audit in 2011	CESSDA PPP WP7 country reports 2008	• funding to commission data audits (budget) • check list for audits • capacities of WG	no funding	country reports, analysis of main findings

Objective	Activities	Responsibilities	Time line	Resources		Constraints	Evaluation
				available	needed		
	organisation; organising workshops every two years				Widening		
	Meetings of data resources on a regional level to mobilise and facilitate collaboration within that area (ESFRI meta regions)	WG Widening	2011, ongoing	CESSDA PPP WP7 results	• fundraising for travelling and hosting meetings • capacities of WG Widening	no funding	reports
	Implementing national cooperative bodies which facilitate access to the new organisation's services for their countries (should become affiliate members)	Director, WG Widening	2011, ongoing	CESSDA PPP WP7 results	capacities of WG Widening	selection of country representatives problematic	report, number of users, online user service
3) Assistance in mobilising national support	Meetings with national funding bodies, national stakeholders and emerging data resources	Director, WG Widening	2010, 2011	-	• funding for travelling and hosting meetings (budget) • capacities of WG Widening	no funding	reports
	Assisting and advising national funding bodies in the selection of potential new member data resources (general information, special meetings, provision of the results from the proposed Annual Data Audits)	Director, WG Widening	2010, 2011, 2012	-	• funding for travelling and hosting meetings (cessda-ERIC budget) • capacities of WG Widening	no funding	reports
	Facilitate the involvement in international data collection programmes, support fundraising	WG Widening	2010 ongoing	Data collections (ISSP, EVS, ESS, etc.)	capacities of WG Widening		reports
	Initiate international data producing and archiving projects	WG Widening	2011, ongoing		• fundraising, cooperation with researchers, • capacities of WG Widening	no projects, no capacities, no funding	reports

Objective	Activities	Responsibilities	Time line	Resources		Constraints	Evaluation
				available	needed		
4) Standards and Tools	Provision of a core set of interoperable standards in cooperation with stakeholders in other fields (from minimum to best practice)	WG Standards and Tools, WG Widening	2010, ongoing	Website	• capacities of WG Standards and Tools and WG Widening • agreement on standards and best practice	no agreement on standards and best practice, no WG Standards and Tools, WG Widening	reports, web page
	Guidelines on fulfilling the requirements of the certification scheme	WG Standards and Tools, WG Widening	2010, ongoing	-	• capacities of WG Standards and Tools and WG Widening • agreement on elements of certification	no agreement on elements of certification; no WG Standards and Tools, WG Widening	Certified scheme, number of certifications
	Free access to a Basic Toolkit (tools for start up)	WG Standards and Tools, WG Widening	July 2010, ongoing	Website	• capacities of WG Standards and Tools and WG Widening, • toolkit	no toolkit; WG Standards and Tools, WG Widening	access numbers
	Guidance on open access/data sharing culture and dissemination	WG Widening, WG Standards	Autumn 2010	Website	• Capacities of WG Standards and Tools and WG Widening	no WG Standards and Tools, WG Widening	Guidelines
	Involvement in the application and testing phase of the proposed cessda-ERIC's R&D projects	WG Widening	2011, ongoing		fundraising for R&D Projects, coordination with ongoing projects	no projects, no funding	number of projects and involvement of non-members
5) Training and knowledge transfer	Outreach program for data archivists/scientists (placements made on the basis of an application and selection process)	WG Widening, WG Training, Training Coordinator	2011, ongoing one visit per year and proposed cessda—		• Training Plan • support from member data resources • applications for external funding	Lack of support from member data resources, Missing external funds	number of access days per member

Objective	Activities	Responsibilities	Time line	Resources		Constraints	Evaluation
				available	needed		
			ERIC service provider				
	Training on different levels with certification by the new organisation; Training methods (according to an agreed Training Programme); Summer School for Data Scientists; Visiting/Staff Exchange Programmes; Workshops for Beginners; Annual Expert Seminars.	WG Widening, WG Training, Training Coordinator	2010 onwards	Established annual CESSDA Expert Seminar	• Training Plan • support from data resources; • applications for external funding • cooperation with ICPSR and IASSIST	missing support from data resources	number of training activities, records of attendance, number of certificates awarded
	Virtual Centre of Competence (VCC): including a Help Desk (skills and expertise page), moderated expert email lists, etc.	WG Widening, Central administration and services (PR, IT)	Set up by April 2010	Website	• funding for PR and IT (web master) • Help Desk Moderator and support by specialists from the new organisation's membership	no funding for PR and IT, lack of support from data resources, no moderator	number of contributors and users/ subscribers, content analysis

Training and exchange programmes for an upgraded RI personnel

Training of new personnel will be an essential part of the action plan for widening the proposed ERIC in its first two years. In WP6 the establishment of a Training Programme is recommended. WP6 suggests organising the programme in a structured form in order to address specific needs of different target groups and different types of knowledge transfer with the most appropriate training methods (Krejci/Hausstein 2008). This approach allows for including potential new data resources in the recommended training programme. Therefore the training activities for new members' personnel should be planned as part of the Training Programme.

As far as the new organisation's personnel are concerned, training at different levels will be required and should include:

a) Basic archiving skills for small data archive units;
b) Advanced archiving skills;
c) Skills in data archive management;
d) Skills related to CESSDA specific services and tools;
e) Skills related to technology development;
f) The opportunity to gain hands-on experience of data archiving tasks.

Recommendation 9

Based on the recommendations of the Training Programme we suggest including potential new data resources in the proposed Summer School for Data Scientists, Workshops for Beginners, Visiting/Staff Exchange Programmes, and Expert Seminars.

Special emphasis should be put on the Visiting and Staff Exchange Programme. Gaining personal experience in dealing with data archiving tasks is very important, especially for new and underdeveloped archives, and for launching new activities. It will allow less experienced partners and potential new members to fully take advantage of existing experience within the new organisation.

- **Visiting Programme (VP):** Short-term stay(s) of employee(s) at established archives to improve expert knowledge and/or facilitate know-how on archival practices.

  Individual visits have already been successfully used within CESSDA and EDAN activities (Hausstein 2002). An additional argument for considering these methods is possible support from specialised EU programmes, e.g. Marie Curie Fellowships.

- **Staff Exchange Programme (SEP):** Establishment of bilateral partnerships and staff exchange between experienced members and less developed archives.

  The SEP should be an essential part of the Training Programme and all future widening tasks. It is a way to leverage the skills of staff within member organisations and foster knowledge exchanges with potential partners in the field.

The concept of exchanging staff with partners (at various levels of the membership, but particularly at a designated level) is a way of sharing knowledge and building new perspectives to help spur on the development of new initiatives.

The objectives of the VP and SEP are to:

- Increase the new organisation, and its partners', capacity to foster development;
- Strengthen relationships amongst participating organisations;
- Enhance the skills and knowledge of program participants;
- Foster cultural diversity and awareness for both the new organisation and its partners.

Every year there should be an announcement offering staff exchange opportunities. The duration of stays/exchanges should be between one week and a month. The provision of SEP's will depend on securing available resources and offering interesting topics. The SEP should be organised by the Training Programme Coordinating Team (proposed by WP6) and supported by the WG Widening.

All designated and affiliated members of the new organisation, and potential new members, are eligible to participate in the SEP. The programme should provide a placement in one of the proposed cessda-ERIC's Full member organisations.

Recommendation 10

The establishment of a Staff Exchange and Visiting Programme should be undertaken in a considered manner. The programmes should be structured, and backed by financial and organisational conditions favourable for both guests and hosts.

Financial resources and capacities

The training of potential/new members as part of the proposed Training Programme will require appropriate institutional resources including requisite levels of knowledge, training facilities and equipment. The proposed Training Programme indicates four groups of funding sources for the planned activities (Krejci 2008):

1) Central budget of the -ERIC;
2) Participation fees (according to level of membership);
3) External support (fundraising);
4) Fundamental contributions from one or several of the ERIC's members.

Recommendation 11

In addition to the proposed financial sources for the Training Programme we recommend establishing a dedicated **Capacity Building Fund** within the central budget in order to facilitate full participation of new members in the proposed Training Programme.

The ERIC's data resources should be the primary source of expertise. Most CESSDA archives have experience of organising training activities. Many of them are able to contribute to the programme through the provision of access to training facilities and expert knowledge (Fabian 2009a). Additional expertise should be generated through co-operation with organisations with similar objectives (e.g. ICPSR, IASSIST, DDI Alliance, Tools Alliance etc.).

Recommendation 12

In accordance with the proposed Training Programme we recommend the setting up of **two training teams and five expert groups**, which consist of experts who should contribute to the proposed Training Programme.

The two **Training Groups** should focus on:

- General digital preservation skills (courses in long-term digital data preservation, digital repositories standards, assessment methods, data digitalisation, etc.);

- Specific social science data archiving skills (specialised courses of metadata production, DDI, NESSTAR, specialised controlled vocabularies and thesauri, data harmonisation, etc.).

The proposed **Expert Groups** should focus on:

- Standards;
- Information Technology and Network Infrastructure;
- Data Harmonisation;
- Controlled Vocabularies;
- Data Protection and Open Access (legal issues).

The CESSDA-PPP Training Programme (Krejci/Hausstein 2008) has suggested that expert groups should support the proposed training activities. Ideally, expert groups should have four to eight permanent members, with between 1-3 external advisors (where required). Regarding the funding of expert groups, we suggest that the new organisation's service providers support the expert groups from their local budgets. This should be taken into account when outlining the membership criteria and negotiating the service level agreement (SLA). The costs of external experts should be covered by the central budget.

Description of Expert Groups

Expert Group on Standards: Members of this group have expertise in recognised international standards for data archiving and dissemination with a special focus on the social sciences (OAIS[1], DDI, other including METS). Their expertise covers the

[1] Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-B-1, Blue Book, January 2002, http://public.ccsds.org/publications/archive/650x0b1.pdf (accessed June 2009)

theoretical grounds and practical implementation of the standards, including tools that are used for data management and data description.

Expert Group on Information Technology and Network Infrastructure: Experts who have knowledge and experience of Web Content Management Systems (WCMSs) in general and WCMSs that were specifically developed for data archives / digital repositories. These software tools enable data curators and data managers to create metadata descriptions which are compliant with specific standards, manage digital objects, define different access level and grant access rights to users and groups of users, etc. There are widely known applications like NESSTAR and Dataverse that can be used as examples of such tools.

Expert Group on Data Harmonisation: This group comprises of experts that have specific knowledge and experience of making key variables comparable across countries and time. Examples of routine data harmonisation tasks are: converting national occupation codes to international standard codes (e.g. ISCO) or to standard occupational prestige scores, harmonising income and educational variables, and attitude scales.

Expert Group on Controlled Vocabularies: Experts in this group are specialised in producing controlled vocabularies to organise knowledge for subsequent retrieval. These are used for indexing subject keywords ("tags"), creating and maintaining thesauri as well as taxonomies, by which an end-user of a database can find information more easily. The ERIC should have a common topics classification and a multilingual thesaurus. This expert group should provide training and guidelines in these fields. Beyond that the expert group should cover the following topics:

- Use of the thesaurus in documentation;
- Multilingual documentation and the role of their controlled vocabularies;
- Functionality of the thesaurus for searching;
- Translation of the thesaurus (concepts should be mapped and not translated literally);
- Maintenance of the thesaurus.

Expert group on data protection and legal issues: Data archives should provide access to research data, whilst taking into account ruling legislation with regard to personal information and intellectual property of the data. The experts of this group are familiar with the national legal frameworks and the European level regulation for protection of personal (sensitive) data and its implementation. Some collections require data archives to anonymise the data collections at ingest and they should provide contractual conditions for end users in which they set the rules for appropriate, adequate use of data collections. This expert group should also have expertise on open access issues and initiatives like the Science Commons project[2].

[2] See at: http://sciencecommons.org

Planning the Virtual Centre of Competence (VCC)

In order to set up new data archives and improve the performance of already existing data resources, there is a need to find information that is both reliable and up-to-date.

Due to financial and staffing restrictions, many of the emergent data archives have the common problem of knowledge ingestion. Operating a fully functional data archive is based on a set of standardised activities that require different staff expertise, and specific material and financial resources. There is no single way to remedy issues arising from this problem, and even if there were a "manual", there would still be difficulties in understanding everything without the help of the more experienced data archives.

What emerging data archives need is to first have free access to a common set of tools, standards and operational procedures, and second to have the possibility to interact with:

- Institutions that are confronted with similar challenges, in order to learn from each others' experience;
- Guidance on unsolved problems from the RI's members.

Data archiving is a field where dramatic changes are taking place, with new standards replacing old ones and new tools being developed in order to meet the new challenges triggered by the recent developments in global communications. This context, in addition to the developments in communication devices (newer generations of hardware, more intelligent connectivity etc.) created a revolution in the data archiving field without precedent in the past 50 years, not only transforming the European landscape but the global one as well.

New standards and tools are being developed and it is difficult for any individual institution to operate in isolation from other parts of the network; implementing a standard that was perfectly operational several years ago will not solve the current communication needs between members of a large network.

All these efforts must be inter-connected and informed by a global perspective, and this can be done only by creating a virtual meeting place where specialists can help each other and create new common projects. It is therefore essential that the proposed ERIC enables both newcomers to find valuable information and specialists to work together more efficiently and share information via a single web point.

In the CESSDA-PPP's work package 6 (WP6) several training activities dedicated to strengthening both the existing (but less well resourced) data archives and emerging ones are recommended. This is one method to provide expert knowledge to those institutions that need it, but using the VCC is another very effective way of disseminating information and solving potential implementation problems.

Recommendation 13

The VCC should become an integral part of the planned cessda-ERIC training programme in order to provide a dedicated information and management platform for future cessda-ERIC service providers.

Aims of the VCC

The general aim of the VCC is to facilitate knowledge dissemination throughout the European landscape and beyond, in all types of data archiving institutions ranging from very mature to incipient ones, thus serving as a platform of exchange and collaboration.

Defined as a web based virtual information and training platform, the creation of the VCC aims specifically to:

- Facilitate knowledge sharing and interactions with experts in the field;
- Support professional training activities;
- Provide continual support and training opportunities by making use of available knowledge and skills within the membership;
- Help integrate knowledge and its applications;
- Strengthen the performance of the ERIC service providers, improve the quality of their services and ensure the means to reach a minimum level of performance;
- Serve as a learning centre for both newcomers and new staff at service provider organisations;
- Generate new knowledge in the field.

All of this depends on the information presented in the VCC, and it is crucial for the ERIC membership to contribute information to the portal.

Recommendation 14

The management board of the upgraded RI should ensure that members/service providers will contribute relevant information to the VCC. We recommend including this as a requirement in the service level agreement.

The structure of the VCC

The VCC should be defined as a web point, administered by the upgraded CESSDA RI, where all information of interest can be searched and various discussion forums can be set up for specific topics.

As many emerging data archives are likely to have similar questions or difficulties, discussion threads are to be archived for future reference in a repository of expertise. Using a search engine similar to Google (it is possible to employ the Google engine

itself for this purpose) newcomers can search for similar questions and benefit from a database of solutions.

The VCC should provide different information and training services to enable its users to:

- Read news and other on-line materials;
- Browse information of interest;
- Download tools, patches and generally relevant software;
- Watch live demonstrations by experts in the field;
- Discuss common problems and propose solutions;
- Participate actively in the community.

The VCC will be similar to a building with many doors, each one representing a discussion forum for a certain topic of interest. Unlike a real building, these doors can be merged if needed and new doors can be created where the latest developments demand action plans.

A first version of the VCC structure should include the following areas:

1. **Standards** section, with two sub forums:
   a) Standards under development, and
   b) Current operating standards;
2. **Tools** section, with two sub forums:
   a) Tools under development, and
   b) Current operating tools;
3. A "**How To**" section, with step-by-step examples in using/implementing tools and/or standards;
4. **Relevant documents**: reference guides, tutorials, green papers, white papers, manuals and general bibliography;
5. **Assistance** in mobilising national support;
6. **News** and upcoming events: meetings, workshops, and conferences;
7. **Help Desk** with audio/video conferencing rooms, for direct discussions with experts in other countries;
8. A "**What's new**" section, where modifications in other sections are automatically flagged for those that are interested (similar to RSS feeds).

WP6 developed a CESSDA toolkit (Fabian 2009b) and recommended documents for the "relevant documents"- section of the VCC. The suggested documents are:

<u>Legal documents, policy recommendations</u>

- Sample contracts for dataset deposition and distribution;
- Sample end-user license;
- Guide to EU legislation / directives on research data;
- Best practice of sharing public use micro-data from official statistics;
- Sample policy documents.

<u>Technical tools and best practice guides</u>

- DDI best practice guides;
- cessda-ERIC archival standards guide;

- DDI software tools;
- Guide to open source web content management tools and data publishing tools;
- Step by step guide to sharing your metadata with members;
- ELSST thesaurus localisation and maintenance guide.

Annex 1 lists the currently available sources. The estimated work effort for preparation of the documents is indicated in table 4:

Table 4: Estimated work effort for preparation of the documents for the VCC:

TASK	**ESTIMATED NUMBER OF PERSON-MONTHS**
Sample contracts for data deposition and distribution	4
Sample end-user licenses	2
Guide to EU legislation / directives on research data	4
Guide to best practices of sharing public use micro-data from official statistics	4
Sample policy documents	2
DDI best practice guides	4
Archival standards guide	5
Data harmonisation guide	4
DDI software tools: a catalogue and a step-by-step guide	5
Guide to open source web content management tools and data publishing tools	5
Step by step guide to sharing your metadata with the cessda-ERIC network	4
ELSST thesaurus localisation and maintenance guide	4
Total	**47**

Some areas of the VCC (for example current operational tools) can be placed in a wiki-like structure, where members can freely add or modify information where improvements are needed. There are multiple possibilities for granting modification rights, ranging from the open to everyone (as per Wikipedia) to some information being restricted to a core of specialists. This system is similar to the current data access system.

Where technical details have been improved, it should be easy to amend the information from the portal, and an alert to be automatically published in the “What’s New?” section.

Recommendation 15

The VCC should support all levels of membership and be dedicated to assisting both full members and designate members in the process of fulfilling specific membership criteria, as well as supporting associates and non-members (by providing general information).

Technical implementation

The VCC is more than a simple information gateway: it is also a platform that should include communication and interlinked services.

The technical and technological setting of the VCC should enable a modular configuration with the aim of integrating additional services easily and modifications according to user needs and further extensions.

In terms of infrastructure, the VCC should be based on a content management system (CMS), where users are able to login and create/modify documents, with the possibility of tracking changes for historical and practical reasons.

Recommendation #16:

The VCC should be accessed via the Portal and incorporated into the current CMS.

Operating this VCC will be one task of the upgraded RI. It is therefore important to plan for its implementation and maintenance, e.g. IT, VCC administrator, content provider(s) etc. Individuals or individual organisations should be able to use a seamless login system, similar to that of the standard system.

Recommendation 17

The login system should be based on the same middleware used in the upgraded RI, and details regarding operating system, web server, CMS and database system are to be agreed upon construction.

Based on an audit of expertise (Fabian/Hegedus 2008) WP6 have recommended that expert groups are set up, these should facilitate a network of expertise that feeds the VCC with expert information and support. The main responsibility of the expert groups is to populate the forum with information, but also to validate/enhance the newly introduced information. Even though some areas of the forum are open, there should still be a professional filter/moderator in order to ensure that correct information is published on the VCC.

Recommendation 18

The ERIC should set up a working group on the VCC with duties ranging from updating web pages with fresh information to answering specific questions in the dedicated forums. This working group should be supported by the proposed expert groups.

It is expected that the VCC will provide a wealth of information for both experts and newcomers, in various levels of complexity. For this reason, in order to find information of interest, it is sometimes better to use a search engine. There are many forums that use such engines, and an example of one such possible solution is vBulletin[3]. Built on PHP and MySQL, it supports thousands of simultaneous users and, in a subsequent version, also has a CMS tightly integrated with the search engine. The search is executed across forum posts, blog entries and comments, in a system that uses semantic mark-ups and CSS styling with customizable layouts. Even though this is a commercial product, there are similar open-source projects available like Sphinxsearch[4] - the purpose of this presentation is to sketch the potential of the search engine. An alternative would be to simply use Google specially adapted to the VCC web pages. Searching the Internet is based on automatic indexes in all possible web pages, so there would be the possibility to use a Google search box embedded in the VCC, with an automatic tag added to the search e.g. “site:cessda.org”

Recommendation 19

The Forum should employ a user-friendly search engine in order to help users find information of interest.

The forum is not only a simple static or dynamic webpage, but should also be used interactively e.g. use teleconference systems to organise eMeetings and eWorkshops. Participants can subscribe electronically and since the meetings are designed as virtual there would be no need to limit the number of participants. There are already such systems in place, for example Marratech (now unavailable)[5] and Smart Board Bridgit[6] conferencing software. Both are mature solutions and can be used effectively for the purpose of interactive eMeetings. However, the Smart Board solution seems to be more flexible, particularly as it uses interactive displays to create sketches and various interactive documents.

Recommendation 20

The VCC should employ a teleconferencing system in addition to the forum web pages.

[3] http://www.vbulletin.com/
[4] http://www.sphinxsearch.
[5] http://www.marratech.com
[6] http://www2.smarttech.com/st/en-US/Products/Bridgit/

Programme for inclusion of new national service providers, based on minimum requirements

For decades, procedures in the data archiving world were developed more or less as a process of parallel, individual efforts. These procedures have slowly become more and more similar, either through the archive-to-archive exchange of information or through formal discussions in expert seminars. The status of data archiving is now developing from serving the researcher towards a true profession, with clear and measurable outputs, standardised operating criteria, and a unified view of how things should be done. In this respect, adopting the philosophy of the OAIS is a perfect example of a worldwide community transformation.

However, there is a clear separation between at least two groups of institutions: experienced and well established data archives with decades of history and experience behind them, and who lead the advancements in data archiving, and those that are less experienced or recently established institutions, with little capacity to adopt the new standards and an even smaller chance of contributing substantially to the development of new standards, technologies, and tools.

Apart from these two groups, there is a potential third group of data resource providers that do not formally operate as data archives but nevertheless have accumulated a wealth of data that are of general interest (see also Part II of this report). This is also the group of potential service providers in countries where no formal data archiving activities currently exist.

This differentiation will be taken into consideration in the Statutes for the ERIC. In order to operate effective service provision, each member should ensure a certain level of compatibility with a common, standardised system of data archiving, communication, and data exchange.

It is a shared responsibility: on the one hand, the new organisation needs to provide a complete set of tools and guidance in order to assist all members in meeting the full membership requirements, but on the other hand it is the responsibility of each organisation to create and secure the necessary environment for professional data archiving activity.

There are many reference models for the digital preservation of data, many of them compared by Štebe and Dusa (2009) and generically defined as DARM (Data Activities Reference Model). They are all concerned with the long term preservation of digital or even physical data for the benefit of a certain community.

One family of DARMs is based on the OAIS standard (Fábián, 2009b). However, the OAIS itself was not designed specifically for social science data archiving, therefore the standard serves as a model, rather than a strict operating procedure. Krejcí (2009), shows that the Data Seal of Approval (DSA)[7] seems to be a better framework for defining a minimum set of criteria for data archiving.

[7] Data Seal of Approval. Quality guidelines for digital research data, v1.4. http://www.datasealofapproval.org (accessed June, 2009).

Minimum requirements

According to the DSA, digital research data must meet five quality criteria:

- The research data can be found on the Internet;
- The research data are accessible, while taking into account relevant legislation with regard to personal information and intellectual property of the data;
- The research data are available in a usable format;
- The research data are reliable;
- The research data can be referred to.

For a newcomer to join the data archiving effort, it should be an absolute minimum to publish the data (or at least the information about the data in a structured catalogue) over the Internet.

With the advent of Internet usage, new members and their service providers should at least demonstrate connectivity with the other members of the network, especially because the data exchanged take an electronic form.

Requirement 1

Potential new member service providers should make their data discoverable over the Internet. They should ensure network connectivity through appropriate national, regional and local area networks.

The fact that an organisation is linked to the Internet and that it advertises the availability of some data is a good start, but users need to know more about available datasets. Some minimum information has to be structured in a formal catalogue of data, in order to be easily browsed, and even searched, if the organisation provides this service.

Usually, a data archive creates a special webpage, initially a static one (a simple HTML table that is manually modified with every new dataset published) and over time redesigns the webpage so that it automatically changes once the internal database has been modified by the publishing scripts.

At the most advanced stage, the server allows for other servers to automatically search for new data (a process called 'harvesting'), so that each time a new dataset is published, the harvesting servers automatically retrieve the information from the host server. When users from other countries need some of the published data, the harvesters should be able to download those data on a principle of reciprocal services.

Requirement 2

The data holdings of ERIC service providers should be published in a structured catalogue, available for harvesting by the Portal, and data should be downloadable through common data gateways.

Along with the data, users should be provided with at least some information about the dataset, the procedures employed when the data were produced, the data producing agency and other basic information, e.g. time coverage, universe of respondents etc.

All these items are covered extensively by the Data Documentation Initiative (DDI), which has become CESSDA's de facto metadata standard. In all its versions, DDI has a set of mandatory information plus a much more extensive set of items that are needed to completely document a dataset, especially for cross-national research or data that are collected in multiple waves.

The DDI metadata files have a pre-defined structure, currently in XML format, with parent and child nodes determining categories and subcategories of information.

Requirement 3

Potential service providers should comply with agreed interoperable metadata standards. Specifically, they should document their data with at least the mandatory items in the DDI standard and publish their metadata files in the XML format.

Having the data prepared in the local language is of great benefit for the local community of users, but for international users this is of little use. At least for the most important datasets (if not for all), translating the data and the documentation into a common, international language is an essential requirement for co-operating with other members of the data archiving network.

This is not simply an issue of regular translation, because most of the documentation (especially that referring to the keywords and concepts associated with the variables in the dataset), needs to be comparable with similar documentation in other data archives. For example, when users from one country search the network for a particular keyword, it has to have the very same meaning in all the languages in order to yield comparable search results.

For this reason, CESSDA's multi-lingual thesaurus (ELSST) should be employed, translated and used by local organisations.

Requirement 4

Potential service providers should maintain their local language within the multi-lingual thesaurus. Index terms from ELSST should be utilised to enhance resource discovery.

As many data are in danger of being lost after the analysis phase, various communities have initiated procedures to preserve the data long enough in order to cope with technological changes over an indefinite period of time.

Software versions are likely to change, and this is also true for operating systems. Therefore it is almost impossible to rely on a certain technology for long-term preservation of data. Most of the data archives in the current CESSDA network (Dusa 2009b) store their data in two separate formats:

1) Temporary dissemination formats: current versions of different statistical packages (like SPSS, SAS, Stata, R etc.);
2) Long-term archival formats using the ASCII format, which is platform and software independent.

Using the ASCII format has the great advantage of eliminating the need for data migration. Large data archives like GESIS or UKDA now face a tremendous challenge of having to migrate many of their data (hundreds, sometimes thousands of datasets) from old software versions like OSIRIS into newer formats employed by the academic community. This task is not easy, as different software is not always perfectly compatible in terms of storage structures.

In order to avoid such problems in the future, data are also stored in the ASCII format, most often using a tab-delimited structure to separate the columns. Whenever a new software version appears to be used by the academic community, the data archive simply changes the export script instead of migrating each and every dataset in the catalogue.

Requirement 5

Potential cessda-ERIC data resource providers should employ the ASCII format for long-term preservation of data files.

There are also other possible requirements for future cessda-ERIC data resource providers, such as the adoption of SSO (Single Sign-On user authentication system), which also implies operating a user registration system compatible with the central server, which is based on the Shibboleth system. However, we do not consider these as minimum requirements.

There are different layers of requirements for different types of membership; our recommendations deal with the absolute minimal requirements for newcomers to start preparing their membership application.

Recommendation 21

Minimal requirements for new service providers should be:

(1) Making their data discoverable over the Internet. They should ensure network connectivity through appropriate national, regional and local area networks;
(2) Publishing their data holdings in a structured catalogue, available for harvesting by the Portal. Data should also be downloadable through common data gateways;
(3) Complying with agreed interoperable metadata standards. Specifically, to document their data with at least the mandatory items in the DDI standard and publish the metadata files in the XML format;
(4) Maintaining the 'local' language within the multi-lingual thesaurus. Index terms from ELSST should be utilised to enhance resource discovery;
(5) Employing the ASCII format for the long-term preservation of data files.

Recommended and alternative hardware and software

WP6 dealt with the software and hardware dimensions of data archiving activity. Over the years each and every data archive has developed local software tools using hardware fitted for particular needs in particular moments in time.

For this reason, the hardware and software employed is as diverse as the number of institutions members in the current CESSDA network. It may be close to impossible to say which hardware and which software is "best" or "recommended" for new data archives, therefore the following text is to be treated as a proposal only, and not a formal recommendation.

Hardware minimum requirements

First of all, exact hardware specifications are of little practical importance. As a general rule of thumb, hardware should be powerful enough to serve all local community requests in less than a few seconds.

For a data archive operating with a single server, the following configuration is probably sufficient (at June 2009):

Processor:	3 GHz
Memory:	at least 2GB, preferably more (perhaps with a 64bit architecture)
HDD:	at least 100GB

More important than hardware specifications for a single machine is the number of machines that a data archive should operate in order to be fully compliant with the OAIS standard (Dusa 2009b).

In this instance, the number of servers needed is four at a minimum, mostly to ensure the security of the archived data, which need to be protected from malicious attacks over the Internet, but also because tasks are better split across machines in order to ensure continuity for other activities if one server breaks down.

The first and the most exposed server is the one holding the public webpage, the very front-end where anyone from the Internet can try to force an entry. For this reason a different, separate server is needed to store the actual archived data (the data storage server). This is the server where temporary versions of the data are stored, in the statistical package formats that are presently used by the community. The reason why temporary versions of the data are stored on these servers is related to decreasing the response time for various user requests.

There are two possible scenarios: one is to keep all data in the ASCII format and generate the version needed by the user for each user request; the other is to generate all possible versions once and then store the result. In the latter case, the response time is substantially lowered because the data are pre-generated and the user is presented with the download options instantly.

For the very long-term preservation of the data, using the archival formats agreed as standard, a different server is needed. While the data storage server should

demonstrate speed, the preservation server should demonstrate reliability and stability over time. Under no conditions are data to be altered - original data are preserved indefinitely.

Last but not least, a fourth server generates preliminary statistical analysis with the data before users download it. In some cases (for example in those data archives holding official statistical data), the users are not allowed at all to use the data outside the data archive. Statistical analyses are only performed using a secure connection with the processing server, in which case this server should be both powerful and separated from the other servers for both decent performance and minimum response time. See the WP10 report on secure data services for further details.

Software minimum requirements

If hardware is more or less common across institutions (the only real difference lies in power), it is even more difficult to make software minimum requirement recommendations.

In WP6 a special survey was carried out across more than ten data archives in Europe (and ICPSR in the US) in order to assess the commonality between a vast universe of diverse tools. (Dusa 2009b)

The survey was based on a series of interview with technical personnel from these data archives: one of the questions related to the kind of software they would use if they were newcomers. Although a considerable amount of expertise exists in relation to the commercial software employed, almost unanimously respondees recommended the use of open-source software for newcomers. In some cases (for example operating systems), open-source software is considered to be even better than commercial software for servers and web publishing. This is the classic case of Linux versus Windows, both of which have pluses and minuses. For servers, general opinion recommends the use of Linux, and for workstations people seem to prefer Windows, which is more user friendly.

For data archiving purposes, there is a suite of open-source software specially designed for general web publishing: Linux, Apache, MySQL and PHP (the so-called LAMP). Equivalent commercial software is Windows, IIS, Microsoft SQL Server, etc.

Recommendation 22

Where cost is an issue, new data archives should turn to open-source software. Where ease-of-use is an issue they should turn to commercial equivalent software.

Support from the cessda-ERIC

Recommendation 23

In order to support potential new service providers in fulfilling the minimum requirements for membership of the upgraded CESSDA RI the new organisation should provide guidelines, best practice advice, a toolkit and specific training.

By using the proposed VCC and Training Programme we recommend launching the following activities:

- Provision of guidelines on minimum requirements;
- Direct help offered via the Help Desk of the VCC;
- A one-day workshop for beginners: "How to set up a social science data archive – what are the minimum requirements set by the cessda-ERIC?"
- A Staff Exchange Programme;
- Provision of a Toolkit and common software solution.

Part II: Comprehensiveness: Co-operating with service providers outside of the current CESSDA network

1. Overall Aim

There are a number of important social science data collections currently residing in organisations and repositories outside of the existing CESSDA network. In building an integrated ERIC it is necessary to develop a strategy whereby data holdings from within member organisations can be simultaneously accessed and integrated with data held externally.

The overall aims of the proposed strategy are to:

- Ensure that the European social science research community has access to the relevant data resources they require, irrespectively of researcher location;
- Advance the critical role of the RI as a key infrastructure for the social sciences;
- Increase the value of the collections held by members of the ERIC;
- Provide a high quality, comprehensive and responsive service for the users;
- Support the implementation of best practice with respect to data management in general and promotion of archival standards among data providers in particular.

2. Identification of data resources

Scope of CESSDA holdings

According to a survey among 18 CESSDA members conducted in April 2008 the scope of their holdings cover a broad range of topics.[8] However, the majority of their data collections cover topics such as politics, social stratification, labour/employment, society/culture, and demography (see figure 4 below).

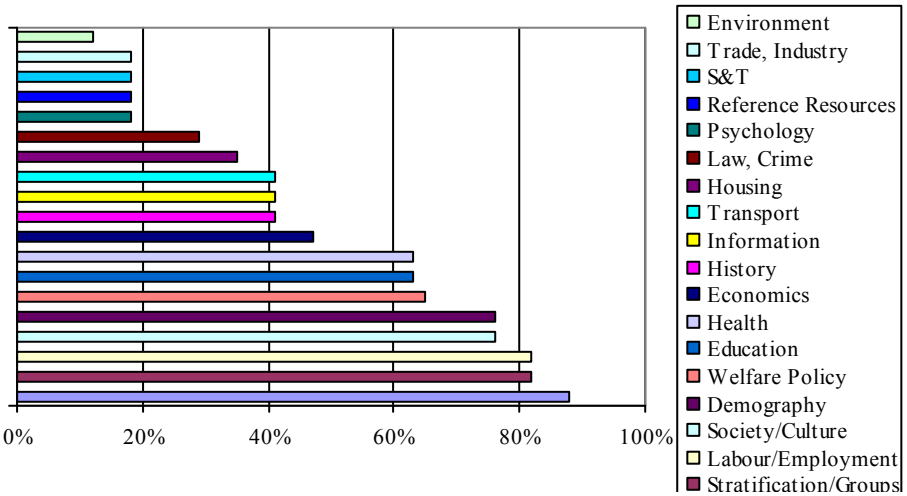

Figure 4: Scope of CESSDA holdings (April 2008)

[8] Question 33: Which of the categories below best describes your collection(s)?
(The list is extracted from the Topic Classification Catalogue for the CESSDA Catalogue), n=17.

From a national point of view the surveyed CESSDA members indicated a number of areas as relevant for the potential broadening of their collections[9], including administrative and business data, data from the financial and retail sector, health and medical sector, as well as public opinion data (see table below).

	Areas relevant for future acquisition
Sweden	*Statistics*
Germany	*Statistics*
France	*Eurostat data,* ***geo-referenced data****, historical data, psychological data*
Greece	**Administrative and business data**
Romania	*Statistics*
Slovenia	Education, economy, psychology, **health**
Hungary	*Statistics*
UK	**Opinion polls**, banking, **financial and retail sector, health and medical sector**
Denmark	Economic time series, natural science data
Finland	*Statistics*, education, **health**
Norway	**Opinion polls**
Switzerland	*Statistics*

Table 5: Areas for future national acquisition policy

Beyond that there are further important datasets hosted by data producers and research groups that are of great interest to the research community. Beside the demand for data at a national level there are also the specific interests of European and international level research to consider. There is a need to understand how economic and social integration operates across the EU and to identify both strengths and weaknesses in policy implementation. This need points to the availability of large comparable micro level datasets (cross-sectional and longitudinal), macro data banks (providing indicators of trends and yielding information on countries and regional differences across Europe) as well as administrative data and data from the commercial sector. (See Elias, 2008)

The majority of the data (67 percent) held in CESSDA archives are microdata. The provision of access to data from official statistics is rare. Therefore most of the archives consider statistical data as a priority for their future acquisition activities (see table above). Workpackage 10 (WP10) of the CESSDA-PPP is dedicated to the development of strategies to facilitate access to statistical data via the upgraded RI – we are therefore not elaborating further on these issues here.

A list of resources and criteria for selection

In order to prepare a list of relevant sources we conducted a comprehensive analysis of available resources on the Internet. Several data portals that list valuable data resources for social sciences were checked (see annex 2). As a result a preliminary overview of a few hundred organisational web pages of potentially relevant organisations was compiled.

Based on this list resources were extracted that fulfil most of the following selection criteria:

- Social science topics which do not overlap with current member provision;

[9] Open question: 36. Please list the types of data that you would like to include in your collection but which are not currently deposited; n=12.

- Data of general interest (including non-social science sources);
- High data quality;
- Well established data access service and data management.

The complete list of resources is provided in Annex 2[10]. In order to develop a specific interoperating programme we suggest grouping the available data sources in the following way:

Group A: **International and regional data producing and providing organisations** A-1 International and regional data providing organisations (ILO, OECD, World Bank, IMF etc.); mainly macrodata; A-2 International operating public opinion poll agencies that offer access to their data on a commercial basis (Gallup, Roper, etc).
Group B: **International and regional project consortia with an emphasis on designing and conducting data production activities:** B-1 Academic oriented (ISSP, EVS, CSES, EES etc.); B-2 (Inter)governmental (Eurobarometer, Afrobarometer, Latinobarometer, Household Surveys etc.).
Group C: **Well established organisations with a mission to provide data access** C-1 Holdings similar to CESSDA (like ICPSR, IPUMS etc.); C-2 Holdings from other similar infrastructures (e.g. SHARE, ESS, etc.) and other fields (economic data, transaction data, administrative data).
Group D: **Special data collections from small to medium size countries** D-1 with a functioning data infrastructure; D-2 without a national data infrastructure.

Barriers to the wide use and sharing of these data resources

Currently there are a number of barriers and constraints to data availability. In developing a strategy of interoperating with various data providers the following constraints were considered:

1. **National access conditions**
   - The multitude of data access policies and regulations implemented by national governments;
   - Funding systems;
   - Language barriers;
   - Data access restricted to nationals.

2. **Constraints related to data providers**
   - Lack of capacity (including financial) and knowledge in the field of data archiving (long term preservation);

[10] Governmental (e.g. statistical offices) and intergovernmental organisations that distribute data as well as international data collection research projects financed by EU Commission are not included in this list.

- The structure of websites making downloading data impractical;
- The data produced are not for wider use (difference between data provision and data sharing);
- Insufficient disaggregation levels of provided data;
- Difficulties in obtaining access for collaborators;
- Data are not of sufficient quality.

3. **Constraints related to researchers / potential users**
   - Insufficient knowledge/ support to find data;
   - Insufficient statistical knowledge/ support to use data;
   - Insufficient international contacts.

4. **General constraints**
   - Costs of data;
   - Copyright issues;
   - Legal issues;
   - Data protection, confidentiality.

3. Strategic aims and proposed actions

Recommendation 24

Based on the overall aim of interoperating with external data sources we suggest implementing a set of actions in order to address the existing barriers (see table 5).

Table 5: Strategic aims and proposed actions

Strategic Aims	Actions
1. Broaden the scope of individual member's data collections according to identified areas.	Monitor the development of the data holdings of the cessda-ERIC: National data audits on a regular basis, review of the national acquisition concepts of member service providers.
2. Extend data access for European researchers according to their needs and available data sets.	Develop analytical capacity to identify needs (user survey) and available resources (yearly screening reports); Establish contact with national and international funders to identify datasets of interest to users arising from funded research both at the grant application stage and finished projects
3. Provide one gateway to relevant data sources; data resources can be easily and seamlessly located and accessed.	Technical requirement concerning cessda-ERIC Gateway for data location; Implementation of tools for incorporating resources from outside the network, including Single-Sign On (authentication system); Set up a European Access Management Federation (however, see recommendations in D5.5.
4. Provide services to macrodata users.	Preparation of user guides for macro datasets, training courses, use of the Virtual Centre of Competence.
5. Provide service for data providers by making their data and/or metadata widely accessible	Development of tools that allow for easy technical interoperability; Implementation of a web data server (beyond 20/20).
6. Provide active assistance in the implementation of best practice and standards among data providers	Offer training facilities (Training Programme, use of the Virtual Centre of Competence); Involvement of the ERIC in data gathering projects (national and European) at an early stage thus allowing for interoperability at a content level.

4. Programme of interoperating

a) Principles and forms of interoperating

General principles of interoperating with data sources outside the ERIC should be:

- In accordance with ERIC priorities (services);
- Compliant with legislation and guidelines (open access);
- Considerate of the legal rights/obligations of partners;
- In the spirit of partnership and fairness;
- Cost-efficient;
- Offer mutual benefits;
- Professional and innovative;
- Effective;
- Committed and reliable.

b) Forms/mode of interoperating (different models depending on data):

There are different forms of cooperation with external resources. In our analysis we found various procedures which are in place (e.g. in ICPSR, ESDS, RELU). The following list gives an overview of the modes of operation relevant to the ERIC:

- Option of (affiliated) membership;
- Applying for membership in data providing organisations (e.g. ICPSR);
- Culture of reciprocity: Access to data for CESSDA members – provision of service (dissemination of data, exclusive dissemination of data, dissemination on metadata level, registration service for users etc.);
- Data redistribution licence agreements on a contractual basis;
- Purchasing of licenses (national, European);
- Remote access;
- Public use files;
- Exchange of metadata;
- Offering facilities for self archiving;
- Cross referencing;
- Special co-operation with national research councils;
- Support for the setting up of national co-operative bodies with the explicit purpose of ensuring all national data providers are associated with the ERIC;
- Person-to-person contacts.

The most appropriate mode of interoperating depends on the institutional status of the data sources, the kind of data and metadata available, the specific conditions of access, and the ERIC service priorities (which will be based on the needs of both members and users).

Recommendation 25

In order to implement the strategy of interoperating we recommend a specific programme of interoperating (see table 6).

Table 6: Programme of Interoperating

Data Resource		Recommended Mode of Cooperation
Group A: International and regional data producing and providing organisations	**Group A-1:** International and regional data providing organisations	1. Access to metadata or aggregate data (offering user guides to macro databanks); 2. Licence agreements; 3. Affiliate membership option.
	Group A-2: International operating public opinion poll agencies	1. Data redistribution license agreement negotiations: Contracts, Purchasing of licences.
Group B: International and regional project consortia with an emphasis on data production activities	**Group B-1:** Academic oriented	1. Affiliate membership option; 2. Provision of service: • One ERIC service provider takes over the responsibility for data management, preservation and service access (depositing the data or providing training and support to enable the project team to provide access to data); • Cooperation on the basis of reciprocity: access to data for ERIC – provision of services.
	Group B-2: (Inter)governmental	1. Affiliate membership option; 2. Special services available: dissemination of data, registration service for users; 3. Data redistribution license agreement negotiations: Contracts, Purchasing of licences.
Group C: Well established organisations with a mission to provide data access	**Group C-1:** Organisations with holdings similar to the cessda-ERIC and other ERICs	1. Affiliate membership option; 2. Application for membership to these organisations.

Data Resource		Recommended Mode of Cooperation
	Group C-2: Holdings from other fields	Cross references, metadata exchange based on OAI PMH; Additionally: Networking and collaboration on common issues and policies, eg. data sharing, open access, professional training, collaboration on development of interoperable standards, technologies, and organising conferences etc.
Group D: Special data collections of small to medium size in countries	**Group D-1:** Those with a functioning data infrastructure	1. Model of self-archiving (European-wide self-archiving service); 2. Acquisition by national partners, building national/regional consortia.
	Group D-2: Those without a national data infrastructure	1. Model of self-archiving (European-wide self-archiving); 2. Special cooperation with national research councils; 3. Support the setting up of national cooperation bodies for the purpose of associating all national data providers with the ERIC.

5. Business model, capacities and timetable

Recommendation 26

We recommend including a dedicated budget for interoperating with external data sources.

This budget should include financial resources for:

- Membership fees;
- The purchase of licences;
- Annual training for data providers (archival standards, metadata production, data protection etc.): two days course with hands on sessions, 8-10 participants;
- Annual user training: full day courses with a hands-on sessions, 10-12 participants;
- Technology development (external IT expert: development and implementation of special tools).

Recommendation 27

Besides funding we suggest the allocation of dedicated resources for a number of specific coordinating tasks. These tasks should be carried out within the framework of a permanent cessda-ERIC Working Group on Interoperating.

The co-ordination tasks should comprise of:

- Co-ordination of the relationship to resources and dealing with legal issues (3 person months (PM) per year);

- Implementation of an ERIC Gateway to external data including a web data server, an authentication system, self archiving facilities (in the construction phase: IT specialist 6PM in the first year , 3 PM/year for ongoing support);

- Maintaining and development of the service (web pages, ongoing maintenance/updating, help desk, mailing list, software guides) (6PM per year);

- Preparation of user guides (in construction phase: 3 PM per year);

- Preparation of annual screening reports, user surveys, reviews of national resources (in operational phase: 3PM per year);

- Support for training courses and preparation of the VCC tutorial (in operational phase: 3PM per year).

Potential members of this working group could include **UKDA, NSD, GESIS and ADP**.

Estimated resources needed for interoperating:

Director	0.5 PM
Coordinator for interoperating	12PM
IT manager	6PM (first year) 3PM
Working Group on Interoperating	6 PM
Budget for interoperating	10k Euro
Training facilities	5k Euro

Recommendation 28

The proposed strategy and actions should be developed for a medium sized term of five years (2010-2014) following which we recommend a construction phase (2010-2011) and an operational phase (2012-2014) for implementing the strategic programme.

The following tasks should be carried out:

I) Construction phase (2010-2011)

1. Adoption of the strategic plan and interoperating programme;
2. Selection of data resources;
3. Negotiation on modes of interoperating with selected resource providers;
4. Technical implementation of the ERIC Gateway to external data;
5. Preparation of user guidelines and a Help Desk (VCC);
6. Evaluation of the programme.

II) Operational phase (2012-2014)

7. Maintenance and development of the service;
8. Annual screening reports, user surveys, review of national resources;
9. Annual training for users and data providers;
10. Evaluation of the programme.

Recommendation 29

The strategy and proposed actions should be reviewed and updated periodically. We recommend an annual update/review.

6. An Acquisition Policy for the ERIC (External Sources)

Recommendation 30

We recommend implementing an Acquisition Policy in relation to external data sources.

Note the suggested policy (below) has neither been discussed nor agreed by the Management Board of the PPP.

.

cessda-ERIC Acquisition Policy - External Sources
(excluding official statistical data)

Purpose

This document is designed to articulate the acquisition policy of the cessda-ERIC with respect to data sources resident in organisations and repositories outside of the network (excluding official statistical data).

Mission

The cessda-ERIC is working to become the underlying operational infrastructure behind an integrated network of data and data resources across the European Research Area. The cessda-ERIC develops and co-ordinates standards, protocols and professional best practise, in order to facilitate and support research, teaching and learning of the highest quality throughout the social sciences.

Aim

Providing access to important national and international data collections outside the current network in order to ensure that the European social science research community has access to the relevant data resources they require, irrespective of location.

Scope

Fields: Social sciences and humanities, economics, health and medicine, administrative and business data, transaction data.

Types of data:

Macrodata - aggregate data produced by intergovernmental organisations such as the World Bank, International Monetary Fund or United Nations.
Microdata - survey or questionnaire datasets collected from international and regional project consortia with an emphasis on designing and conducting data production activities, as well as specific national data collections.
Geographical Focus - International.

Acquisition criteria

Based on the annual identified needs of the user community (national data audits and user surveys) resources should be selected in accordance with the following criteria:

- Relevance and content of data source (non-overlapping with members' provision);
- High quality data and metadata;
- Well established data access services and data management;
- English language documentation of data.

Methods

Interoperating with external sources is carried out by:

- Offering (affiliated) cessda-ERIC membership;
- Applying for membership of data providing organisations (e.g. ICPSR);
- Reciprocity: Access to data for the ERIC – provision of services, (dissemination of data, exclusive dissemination of data, dissemination on a metadata level, registration service for users etc.);
- Data redistribution licence agreements on a contractual basis;
- Purchasing of licenses (European);
- Remote access;
- Exchange of metadata;
- Cross referencing;
- Offering facilities for self archiving;
- Special cooperation with national research councils;
- Supporting the establishment of national cooperation bodies for the purpose of associating all national data providers with the cessda-ERIC;
- Person-to-person contact.

The appropriate mode of interoperating depends on the institutional status of the data sources, types of offered data and metadata, the specific conditions of access, and cessda-ERIC service priorities (which are based on the needs of members and users).

Roles and responsibilities

cessda-ERIC Board

The cessda-ERIC Board approves the acquisition policy and strategies. It is responsible for allocating the required financial resources for its operation.

cessda-ERIC Directorate

The cessda-ERIC Directorate will develop an acquisition strategy and plan to carry out the acquisition policy. It will select and negotiate with service providers. It will set up a dedicated working group on interoperability which will be responsible for specified coordinating tasks.

cessda-ERIC Members

Representatives of the cessda-ERIC membership and ERIC service providers will contribute to the identification of resources (data audits) and support interoperating with single national resource providers. Collectively and collaboratively they will carry out the following tasks:

- Co-ordination of the relationship to resource providers and dealing with legal issues;
- Implementation of a cessda-ERIC Gateway to external data;
- Maintenance and development of the service;

- Support training courses and use of the Virtual Centre of Competence.

Availability of acquisition policy

Copies of this acquisition policy will be made available on the the ERIC webpage.

Review

This policy will be reviewed every two years. Prior to this review comments or concerns should be directed to the cessda-ERIC Directorate.

Approval

This acquisition policy has been recommended by the cessda-ERIC Directorate and approved by the cessda-ERIC Board on <date>.

References

Dusa, Adrian (2009a) Planning a specific Virtual Centre of Competence (VCC). Draft version, discussion document for CESSDA-PPP. (02/24/2009)

Dusa, Adrian (2009b) Task 6.10 interim report: Writing the list of minimum hardware requirements necessary to set up a data archive. CESSDA-PPP, workpackage 6.

Dusa, Adrian (2009c) Prepare a programme for inclusion of new members to the CESSDA RI based on the organisation meeting the minimum requirements. Draft version, discussion document for CESSDA PPP. (06/20/2009)

Elias, Peter (2008c) Providing data on the European level. Working paper of the German Council for Social and Economic Data (RatWSD) No. 46, Berlin.

Fabian, Zoltan (2009a) Task 6.2 report: Self-assessment procedures for the CESSDA RI infrastructures. CESSDA-PPP, work package 6. (01/19/09)

Fabian, Zoltan (2009b) Develop a strategic framework for a "CESSDA toolkit" Interim report from Task 6.6. (09/06/01)

Fabian, Zoltan and Peter Hegedus (2009) Selection of working groups and training teams from existing CESSDA RI (experts in the field and necessary new staff). Draft version, discussion document for CESSDA-PPP. (09/06/09)

Fabian, Zoltan and Peter Hegedus (2008) Audit of expertise within the CESSDA network of excellence. Interim report from Task 6.3 (10/06/08)

Hegedus, Peter (2009) Development of recommendations for training and exchange programmes for new CESSDA RI personnel. Draft version, discussion document for CESSDA-PPP. (09/06/09)

Hausstein, Brigitte; Janez Stebe, Adrian Dusa, and Peter Hegedus (2009) Report on the WP7 Workshop on Extending the CESSDA Research Infrastructure: Potentials and Possibilities, Bucharest (Romania), 17 July 2008. CESSDA-PPP Report of WP7. (01/06/09).

Hausstein, Brigitte and Kevin Schürer (2008) The CESSDA-PPP: Fact finding trip to Slovakia, Bulgaria, Macedonia, Serbia, and Croatia, 30 October to 6 November 2008. Joint activity of WP3 and WP7. CESSDA-PPP Report of WP7. (11/28/09)

Hausstein, Brigitte (2008) Action Plan: Widening ERIC Membership. Draft version, discussion document for CESSDA-PPP. (12/15/2008)

Hausstein, Brigitte and Kevin Schürer (2009) The CESSDA-PPP: Fact finding trip to Poland, Ukraine, Belarus, Russia, Lithuania, Latvia and Estonia 2 to 13 March 2009. Joint activity of WP3 and WP7. CESSDA-PPP Report of WP7. (03/23/09)

Hausstein, Brigitte and Janez Stebe (2009) Strategic and Business Plan for Interoperating with Data Providers/Resources outside *cessda-ERIC*. Draft version, discussion document for CESSDA-PPP. (06/12/09)

Hausstein, Brigitte, Paul de Guchteneire (Eds.) (2002) Social Science Data Archives in Eastern Europe. Results, Potentials and Prospects of the Archival Development. Berlin, Cologne, Paris: Ferger Verlag.

Krejci, Jindrich (2009) Task 6.8 Interim Report: Data archiving standards for the *cessda-ERIC*. CESSDA-PPP, work package 6. (06/17/09)

Krejci, Jindrich and Brigitte Hausstein (2008) Establishing CESSDA Training Programme: Challenges and possible strategies. Draft version, discussion document for the CESSDA-PPP. (01/12/2008)

Schürer, Kevin (2008) Background document on Legal, Governance and Funding issues relating to the CESSDA-PPP for Funders Meeting, London, 27 October 2008. URL (26/06/2009):
http://www.cessda.org/ppp/wp03/CESSDA-ERI-FundersMeeting.pdf

Schürer, Kevin (2008): CESSDA Funders' Meeting. Presentation at the British Academy, London 27 October 2008
URL (26/06/2009): http://www.cessda.org/project/doc/fm_20081027_ks.pdf.

Stebe, Janez and Adrian Dusa (2009) Task 6.4 Report: Recommendations concerning best practices. CESSDA-PPP Workpackage 6. (07/04/09)

Annex

Annex 1: List of available resources for expert groups, VCC contents

Sample contracts for data deposition and distribution

Finnish Data Archive (FSD) Data Deposition agreement:
http://www.fsd.uta.fi/english/forms/deposit.rtf
An example of a depositor agreement (for the Edinburgh DataShare repository):
http://datalib.ed.ac.uk/DataShare/Depositor-Agreement.pdf

Sample end-user licence

UKDA Special Licence - Approved Researcher access to more detailed social survey microdata: http://www.data-archive.ac.uk/orderingData/specialLicence.asp
CESSDA description of end user license:
http://www.cessda.org/sharing/dissemination/3/
FSD End user license: http://www.fsd.uta.fi/english/forms/conditions.pdf
UKDA End User license: http://www.data-archive.ac.uk/aandp/access/licence.asp

Guide to EU legislation / directives on research data

NESSIE "Report on NESSIE Roundtable Three":
http://www.nessie-essex.co.uk/roundtable_3Report.pdf
Description of EU legal framework:
http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/statistical_confidentiality/legal_framework

Best practice of sharing public use micro-data from official statistics

Hundepool, Anco et al. 2009. "Handbook on Statistical Disclosure Control." V.1.1:
http://neon.vb.cbs.nl/casc/.%5CSDC_Handbook.pdf
Combined homepage of several European projects working in the field of Statistical Disclosure Control: http://neon.vb.cbs.nl/casc/..%5Ccasc%5Cindex.htm
The Economic and Social Data Service (ESDS) Government statistics, United Kingdom: http://www.esds.ac.uk/Government/

Sample policy documents

Woollard, Matthew. 2008. " UK Data Archive Preservation Policy": http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0308.pdf
Economic and Social Research Council (ESRC). 2000. (April) "Data policy":
http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/DataPolicy2000_tcm6-12051.pdf
The Rural Economy and Land Use (RELU) Programme's data management policy:
http://www.data-archive.ac.uk/relu/RELU%20Data%20Policy.pdf

ICPSR Policies

Collection Development Policy:
J. Esanu, J. Davidson and S. Ross, and W. Anderson. "Selection, Appraisal, and Retention of Digital Scientific Data: Highlights of an ERPANET/CODATA Workshop " http://www.jstage.jst.go.jp/article/dsj/3/0/227/_pdf
ICPSR Collection Development:
Policyhttp://www.icpsr.umich.edu/ICPSR/org/policies/colldev.html "This document describes the principles that underlies ICPSR's collection development activities." (Quoted from the cited document.)

Data Enclave Policy and Procedures
http://www.icpsr.umich.edu/ICPSR/org/policies/enclave.pdf "The ICPSR Data Enclave is located at the ICPSR offices and provides a secure environment in which researchers can access and analyze restricted data."
http://www.icpsr.umich.edu/ICPSR/org/policies/index.html (Quoted from the cited document.)
ICPSR's Preservation with Delayed Dissemination Policy
http://www.icpsr.umich.edu/ICPSR/org/policies/delayed.html
Redistribution Policy - http://www.icpsr.umich.edu/ICPSR/org/policies/index.html "The following statement, from Article 1.2.B of the Bylaws, is agreed to by each campus Official Representative upon their institution's becoming a member of ICPSR: Members will not distribute data or other materials supplied by ICPSR to other members, organizations, or individuals at other institutions, without the written agreement of ICPSR." (Quoted from the cited document.)
Requests for Permission to Redistribute ICPSR Data
http://www.icpsr.umich.edu/ICPSR/org/policies/redistribute.html (Quoted from the cited document: "ICPSR's policy regarding the redistribution of ICPSR data is explained in the document.")
Web Privacy Policy http://www.icpsr.umich.edu/ICPSR/org/policies/privacy.html ICPSR's privacy policy explains what information they collect from visitors to their website.

Technical tools and best practice guides
DDI best practice guides
Anne Green. 2008. "Data Documentation Initiative - DDI ." Presented at the DataShare project meeting University of Edinburgh, Feb 5-6, 2008. http://www.disc-uk.org/docs/DDI_Green.pdf
DDI 3. Documentation. http://www.ddialliance.org/ddi3/index.html
CESSDA 2001 Recommended DDI 2.0 elements.
http://www.ddialliance.org/DDI/related/cessda-rec.pdf
DDI Best Practices Across the Data Life Cycle.
http://staging.icpsr.umich.edu/DDI/bp/
This site contains best practice recommendations for DDI metadata description production.
Edwards Michelle, Jane Fry, Peter Granda, Kirstine Kolsrud, Stefan Kramer, Ken Miller, Ron Nakao, Mary Vardigan . 2009. "Implementation and Governance."
http://staging.icpsr.umich.edu/DDI/bp/DDIBestPractices_ImplementationAndGovernance.doc.pdf
Abstract: "This best practice covers the process of developing agreements and guidelines for implementation and governance of DDI as a metadata scheme for a community. The best practice recognizes that different communities will, by necessity, have different processes for this, and offers a checklist of issues for communities to consider before implementing DDI. "
Dinkelmann, Karl, Michelle Edwards, Jane Fry, Chuck Humphrey, Ron Nakao, Wendy Thomas. 2009. "Work flows - Data Discovery and Dissemination: User Perspective." http://staging.icpsr.umich.edu/DDI/bp/DDIBestPractices_Workflows-DiscoveryAndDissemination.doc.PDF Abstract: "Describes the best practices for metadata producers to provide end users with the resources for data discovery and dissemination."

Dinkelmann, Karl, Michelle Edwards, Jane Fry, Chuck Humphrey, Kirstine Kolsrud, Stefan Kramer, Jenny Linnerud, Hans Jørgen Marker, Meinhard Moschner, Ron Nakao, Wendy Thomas, Achim Wackerow, Wolfgang Zenk- Möltgen . 2009. "Work Flows - Archival Ingest and Metadata Enhancement."
http://staging.icpsr.umich.edu/DDI/bp/DDIBestPractices_Workflows-ArchivalIngestAndMetadataEnhancement.doc.pdf Abstract: "This Best Practice discusses workflows for DDI usage in the context of archival ingest and metadata enhancement, beginning at the point of the handoff between the data provider and the archive."
Marker, Hans Jørgen, Wolfgang Zenk-Möltgen, Wendy Thomas, Achim Wackerow. 2009. "Workflows for Metadata Creation Regarding Recoding, Aggregation and Other Data Processing Activities."
http://staging.icpsr.umich.edu/DDI/bp/DDIBestPractices_Workflows-MetadataCreationRecodingAggregation.doc.pdf Abstract: "This best practice discusses the capturing, in DDI metadata, of the processes of data aggregation, recoding and data processing."
Granda, Peter, Stefan Kramer, Jenny Linnerud, Hans Jørgen Marker, Ken Miller, Mary Vardigan . 2009. "Controlled Vocabularies."
http://staging.icpsr.umich.edu/DDI/bp/DDIBestPractices_ControlledVocabularies.doc.pdf Abstract: "The benefits of using controlled vocabularies within DDI metadata creation are the primary underlying principles behind this Best Practice document. Use of a controlled vocabulary enhances consistency and efficiency in the production of DDI metadata, gives precision in searching the DDI metadata, and allows semantic and technical interoperability between organisations creating DDI instances."
Ionescu, Sandra. 2009. "Creating a DDI Profile" (2009-02-15)
http://staging.icpsr.umich.edu/DDI/bp/DDIBestPractices_CreatingAProfile.doc.PDF
Abstract: "This document outlines recommended best practices for creating a local DDI 3.0 Profile, which is a subset of DDI 3.0 fields to be used by an organization or shared by a community of users. The DDI 3.0 specification is extensive and designed to cover a multiplicity of use cases. However, not all of DDI's possible applications will be relevant at the level of specific organizations or user communities. By creating and implementing user-specific profiles, organizations will: Ensure that DDI documentation is suited to their particular requirements; Expedite and simplify DDI production and processing; Optimize interoperability and facilitate document sharing with other DDI users. The guidelines included in this document are intended to assist potential users in building a technically accurate and complete profile that will serve as an effective tool in managing DDI instances as well as data sharing operations."

A cessda-ERIC archival standards guide

ICPSR guide on data deposition "How to deposit data":
http://www.icpsr.org/ICPSR/access/deposit/
ICPSR Guidelines for Depositing Data:
http://www.icpsr.umich.edu/ICPSR/access/deposit/guidelines.html
Dataset description for the Finnish Data Archive:
http://www.fsd.uta.fi/english/forms/description.pdf
Dusa, Adrian and Janez Stebe. 2009. "Task 6.4 Report: Recommendations concerning best practices". Draft v0.4 http://www.cessda.org/ppp/wp06/WP6_T6.4.pdf
ICPSR. "Confidentiality Review." Ann Arbor: ICPSR. Available from:
http://www.icpsr.umich.edu/access/deposit/conf-review.html

Green, Ann, Stuart Macdonald and Robin Rice. 2009. *Policy-making for Research Data in Repositories: A Guide.* DISC. DataShare Project: http://www.disc-uk.org/docs/guide.pdf
Interim T6.8 report on task standards in data archiving, CESSDA-PPP.

Data harmonisation

Work Package 9 of the 2008-2009 CESSDA PPP under the 7th Framework Programme. "Building an Infrastructure for Content Harmonisation and Conversion": http://www.cessda.org/project/doc/wp09_descr2.pdf
Juergen H.P. Hoffmeyer-Zlotnik. 2004. "Data Harmonisation." Roundtable 4. 26-27 November 2004 Luxembourg . NESSIE Network of Economic & Social Science Infrastructures in Europe.

DDI software tools

DDI tools Web page: http://tools.ddialliance.org/
NESSTAR: www.nesstar.com
Dataverse: www.thedata.org
IHSN Microdata Management Toolkit: http://www.surveynetwork.org/home/?lvl1=tools&lvl2=documentation&lvl3=toolkit
Interim T6.9 report on Recommended and alternative software for data archiving, CESSDA-PPP.

Guide to open source web content management tools and data publishing tools

Data publishing tools / software for digital repositories:
NESSTAR: www.nesstar.com / Dataverse: www.thedata.org
Dspace: http://www.dspace.org/
FEDORA Commons: http://www.fedora.info/
Major open source web content management systems (WCMS):
Drupal: www.drupal.org (Linux, Apache, MySQL, PHP)
Joomla!: www.joomla.org (Linux, Apache, MySQL, PHP)*
Plone: www.plone.org (Windows/Linux, Zope, Python)
Note: there are commercial modules as well for Joomla!
Step-by-step guide to share your metadata with the proposed cessda-ERIC network:
Interim T6.11 report on Step-by-step guide to publishing a dataset to the CESSDA Portal, CESSDA-PPP;
Interim 6.10 report on Writing the list of minimum hardware requirements necessary to set-up a data archive, CESSDA-PPP.

ELSST thesaurus localisation and maintenance guide

Workpackage description. 2008. "Enhancing the multi-lingual Research Infrastructure (RI). Workpackage 4 of the 2008-2009 CESSDA-PPP under the EC 7th Framework Programme." http://www.cessda.org/project/doc/wp04synopsis_sept08.pdf *
Balkan, Lorna, Ken Miller, Birgit Austin, Anne Etheridge, Myriam Garcia Bernabé and Pam Miller. 2008: ELSST: a broad-based Multilingual Thesaurus for the Social Sciences: http://gandalf.aksis.uib.no/lrec2002/pdf/3.pdf.

Annex 2: List of relevant data service providers outside of CESSDA RI

The following table lists the identified data resource providers that currently reside outside of CESSDA. They are listed according to four groups (as described in text):

Group A: International and regional data producing and providing organisations.
Group B: International and regional project consortia with an emphasis on designing and conducting data production activities.
Group C: Well established organisations with a mission in providing data access.
Group D: Special data collections of small to medium size in countries.

Governmental (e.g. statistical offices) and intergovernmental organisations that distribute data as well as international data collection research projects financed by EU Commission are not included in the list.

Legend:
Quality (Data + Metadata): (1) High
Topic: (1) Disciplinary SS
Relevance/General interest: (1) Yes
Size: (1) Large, more than 100 datasets

Group	Data Source	Institution	Country	Internet Address	Quality	Topic	Relevance	Size	Notes (topic)	Notes (collection)	Access
A	United Nations Industrial Development Organization (UNIDO)	UN	Austria	http://www.unido.org/	1	1	1	1			
A	OECD	Intergovernmental organisation	France	http://www.oecd.org/	1	1	1	1			
A	World Health Organisation	WHO	Switzerland	http://www.who.int/	1	1	1	1	country statistics		
A	UN Statistical Databases	United Nations	USA	http://unstats.un.org/unsd/databases.htm	1	1	1	1	country statistics		Most datasets are available as special orders or as publications for sale.
A	World Bank	World Bank	USA	http://www.worldbank.org/	1	1	1	1	country statistics		Freely available
A	National Opinion Research Center	University of Chicago	USA	http://www.norc.org	1	1	1	0			
A	Energy Statistics Division	IEA	France	http://www.iea.org/about/esd.htm	1	1	0	1	energy consumption		
A	Gallup Brain	Gallup Organization	USA	http://brain.gallup.com/	1	1	0	1	social indicators		
A	Pew Global Attitudes Project data archive	Pew Research Center ,	USA	http://pewglobal.org/	1	1	0	1	opinion surveys		
A	SensorBase	Center for Embedded Networked Sensing	USA	http://sensorbase.org/	1	1	0	1	sensor data		

Group	Data Source	Institution	Country	Internet Address	Quality	Topic	Relevance	Size	Notes (topic)	Notes (collection)	Access
A	Bureau of Labour Statistics	Government	USA	http://www.bls.gov/data/	1	1	0	1	labour statistics	tables online	
A	EconData	Inforum	USA	http://www.inforum.umd.edu/econdata/econdata.html	1	1	0	1	economic time series		
A	Roper Centre	Roper Centre	USA	http://www.ropercenter.uconn.edu/	1	1	0	1	public opinion		Fees for non-members
A	The International Labour Organisation			http://www.ilo.org/	1	1	0	1	labour		
A	United Nations Surveys on Crime Trends and the Operations of Criminal Justice Systems	United Nations Crime and Justice Information Network	Austria	http://www.uncjin.org/Statistics/WCTS/wcts.html	1	1	0	0	crime		
A	Division of Science Resources Statistics	National Science Foundation	Belgium	http://www.nsf.gov/statistics/database.cfm	1	1	0	0	data and indicators about science	surveys on education, list of scientists / national scientific boards	Need for a licence agreement
A	TransMONEE UNICEF Innocenti Research Centre's	UNICEF	Italy	http://www.unicef-irc.org/databases/transmonee/	1	1	0	0	social indicators of children		
A	UNECE Population Unit	United Nations Economic Commission	Switzerland	http://www.unece.org/pau/Welcome.html	1	1	0	0	family, ageing		

Group	Data Source	Institution	Country	Internet Address	Quality	Topic	Relevance	Size	Notes (topic)	Notes (collection)	Access
		for Europe									
A	LABORSTA	ILO Bureau of Statistics	USA	http://laborsta.ilo.org/	1	1	0	0	labour	aggregate statistics	
A	Nutrition and food safety resource	International Food Information Council (IFIC)	USA	http://www.ific.org/research/foodandhealthsurvey.cfm	1	1	0	0	nutrition / nutrition practices	limited to topic /2006-2008	Data tables (pdf)
A	SeaDataNet	Network	EU	http://www.seadatanet.org/	1	0	0	1	oceanographic		
A	Earth System Science Data	Copernicus GmbH	Germany	www.earth-system-science-data.net	1	0	0	1	environment	data publishing journal	
A	International transport forum	OECD	France	http://www.internationaltransportforum.org/irtad/index.html	1	0	0	0	transport	excel sheets available / membership access to data bases	Full access dependent on membership
B	Latinobarómetro	Corporación Latinobaróme tro	Chile	http://www.latinobarometro.org	1	1	1	0	public opinion		100 EUR charge apply
B	Mannheim Centre for European Social Research	Universität Mannheim	Germany	http://www.mzes.uni-mannheim.de/fs_daten_e.html	1	1	1	0	elections, welfare states	100 (shipping costs)	Fees of up to 500 EUR for some of the products
B	Luxembourg Income Study	Luxembourg Income Study	Luxemburg	http://www.lisproject.org/	1	1	1	0	wealth, labour	small collection, two major topics	End user registration

Group	Data Source	Institution	Country	Internet Address	Quality	Topic	Relevance	Size	Notes (topic)	Notes (collection)	Access
B	Russia Longitudinal Monitoring Survey (RLMS)	University of North Carolina	Russia	http://www.cpc.unc.edu/projects/rlms	1	1	1	0	health, welfare	16 waves since 1992	Data on a cost-recovery base
B	Asian Barometer	National Taiwan University	Taiwan	http://www.asianbarometer.org	1	1	1	0	Social surveys		Access application
B	Latin American Public Opinion Project	Vanderbilt University	USA	http://www.vanderbilt.edu/lapop/	1	1	1	0	public opinion		Subscription only
B	Multinational Time Use Study (MTUS)	Centre for Time Use Research	United Kingdom	http://www.timeuse.org/mtus/	1	1	0	1	time use		
B	Center for International Development (CID)	Harvard University	USA	http://www.cid.harvard.edu/ciddata/	1	1	0	1	global poverty		
B	Dafne - DAta Food Networking	School of Medicine	Greece	http://www.nut.uoa.gr/english/index.asp?page=202	1	1	0	0	food, dietary related data		
B	European Crime and Safety Survey (EU ICS)	UNICRI	Italy	http://www.unicri.it/wwd/analysis/icvs/index.php	1	1	0	0	victimisation		
B	World Database of Happiness	Erasmus University Rotterdam	Netherlands	http://worlddatabaseofhappiness.eur.nl/	1	1	0	0	welfare indicators		
B	The International Institute for Democracy and Electoral	Intergovernmental organisation	Sweden	http://www.idea.int/resources/databases.cfm	1	1	0	0	democracy		

Group	Data Source	Institution	Country	Internet Address	Quality	Topic	Relevance	Size	Notes (topic)	Notes (collection)	Access
	Assistance										
B	Lijphart Elections Archive	UC San Diego	USA	http://dodgson.ucsd.edu/lij/	1	1	0	0	elections	48 studies in Dataverse	
B	Pippa Norris Data	Harvard	USA	http://ksghome.harvard.edu/~pnorris/Data/Data.htm	1	1	0	0	elections		
B	TIMSS & PIRLS International Study Center	International Association for the Evaluation of Educational Achievement	USA	http://timss.bc.edu/	1	1	0	0	education		
B	Institute for Research on Poverty	University of Winsconsin	USA	http://www.irp.wisc.edu/	1	1	0	0	poverty		Required to submit a research proposal
B	Publishing Network for Geoscientific & Environmental Data	University of Bremen	Germany	http://www.pangaea.de/	1	0	0	1	environment		
C	UBC Library	University of British Columbia	Canada	http://data.library.ubc.ca/	1	1	1	1	social surveys		

Group	Data Source	Institution	Country	Internet Address	Quality	Topic	Relevance	Size	Notes (topic)	Notes (collection)	Access
C	Data Center	Computing in the Humanities and Social Sciences (CHASS) at the University of Toronto	Canada	http://dc1.chass.utoronto.ca/	1	1	1	1	social surveys		
C	National Digital Archive of Datasets	National Archives	UK	http://www.ndad.nationalarchives.gov.uk/	1	1	1	1	government digital data		Registration needed
C	Data Archive Services	The Odum Institute	USA	http://152.2.32.107/odum/jsp/content_node.jsp?nodeid=7	1	1	1	1	social surveys		
C	ICPSR	Institute for Social Research	USA	http://www.icpsr.umich.edu/	1	1	1	1			
C	ISSR Social Science Data Archive	UCLA	USA	http://www.sscnet.ucla.edu/issr/da/	1	1	1	1		Earthquake related data, among others	
C	Time Series Data Library	Monash university	Australia	http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/	1	1	0	1	economic time series		
C	EDAC**wowe**	EU Network of Excellence	EU	http://www.edacwowe.eu/en/frmIndex	1	1	0	1	work and welfare	META-DATA-SHELL	
C	The European Health and Safety Database (HASTE)	Finnish Institute of Occupational Health	Finland	http://www.ttl.fi/internet/partner/haste	1	1	0	1	health and safety at work	Summaries of descriptions	Data can be provided on discs for 35 EUR
C	Archaeology Data Service	ADS	United Kingdom	http://ads.ahds.ac.uk/	1	1	0	1	archaeology		

Group	Data Source	Institution	Country	Internet Address	Quality	Topic	Relevance	Size	Notes (topic)	Notes (collection)	Access
C	EDINA	University of Edinburgh	UK	http://edina.ac.uk/	1	1	0	1	geographic		
C	Davidson Data Center and Network	William Davidson Institute	USA	http://ddcn.prowebis.com/general/overview.asp	1	1	0	1	economy	data from transition and emerging market economies	
C	Harvard Geospatial Library	Harvard University	USA	http://peters.hul.harvard.edu:8080/HGL/jsp/HGL.jsp	1	1	0	1	GIS data		
C	IPUMS	Minnesota Population Center	USA	http://usa.ipums.org	1	1	0	1	socioeconomic panel		
C	International Archive of Education Data	Institute for Social Research	USA	http://www.icpsr.umich.edu/IAED/welcome.html	1	1	0	1	adult education		
C	Social Science Electronic Data Library	Sociometrics Corporation	USA	http://www.socio.com/edl.htm	1	1	0	1	health and social sciences	over 500 datasets from more than 250 studies	About 200 for a data set
C	Association of Religion Data Archives	ARDA	USA	http://www.thearda.com	1	1	0	1	religion	survey and administrative data	
C	Wharton Research Data Services	University of Pennsylvania	USA	https://wrds.wharton.upenn.edu/	1	1	0	1	finance, accounting, banking, economics, management, marketing and public policy		Access limited to academic and non-commercial research purposes

Group	Data Source	Institution	Country	Internet Address	Quality	Topic	Relevance	Size	Notes (topic)	Notes (collection)	Access
C	SHARE - Survey of Health, Ageing and Retirement in Europe		Germany	http://www.share-project.org/	1	1	0	0	ageing		
C	Psychology Data Archive	Institute for Psychology Information (ZPID)	Germany	http://www.zpid.de/index.php?wahl=products&uwahl=frei&uuwahl=psychdatainfo&lang=EN	1	1	0	0	Psychology		
C	RELU Data Support Service	University of Essex	United Kingdom	http://www.data-archive.ac.uk/relu/	1	0	0	1	interdisciplinary	links to data	
C	Depository of Unpublished Data	Canada Institute for Scientific and Technical Information	Canada	http://cisti-icist.nrc-cnrc.gc.ca/eng/ibp/cisti/collection/collection-development-policy/collections.html	1	1	1	1	unpublished data	collection of smaller studies, unknown quality	Committed to providing permanent access
C	Center for International Earth Science Information Network (CIESIN)	Earth Institute at Columbia University	USA	http://www.ciesin.columbia.edu/	1	1	0	1	environment		
D	The Henry A. Murray Research Archive	Harvard University	USA	http://www.murray.harvard.edu/	1	1	0	1	mixed research data		
D	Max Planck Institute for Demographic Research	Max-Planck-Gesellschaft	Germany	http://www.demogr.mpg.de/	1	1	0	0	demographic		

Group	Data Source	Institution	Country	Internet Address	Quality	Topic	Relevance	Size	Notes (topic)	Notes (collection)	Access
D	Data Services and Data Bases	DIW Berlin	Germany	http://www.diw.de/deutsch/data_service/32763.html	1	1	0	0	economy, demography, SOEP Germany		Parts of datasets, otherwise a special contract is needed
D	Institute for Employment Research	Bundesagentur für Arbeit (	Germany	http://www.iab.de/	1	1	0	0	labour market, inequality	contact point became a module of the Research Data Centre	Scientific use files (SUF)
D	Archiv für Lebenslaufforschung, ALLF	Universitat Bremen	Germany	http://www.lebenslaufarchiv.uni-bremen.de/index.php?id=567&no_cache=1&L=1	1	1	0	0	life-course		
D	IDSC - Labour Market Data	Institute for the Study of Labour	Germany	http://idsc.iza.org/metadata/	1	1	0	0	labour market data		Direct access to microdata is only possible for registered and obligated users following acceptance of their application
D	CentERdata	CentERdata	Netherlands	http://www.centerdata.nl/en/TopMenu/Databank/	1	1	0	0	labour and general, household panel data		
D	The Norwegian Historical Data Centre	Faculty of Social Science at the University of Tromsř	Norway	http://www.rhd.uit.no/indexeng.html	1	1	0	0	historical data		
D	Demographic Data Base	Universitat Umea	Sweden	http://www.ddb.umu.se/index_eng.html	1	1	0	0	demographic		

Group	Data Source	Institution	Country	Internet Address	Quality	Topic	Relevance	Size	Notes (topic)	Notes (collection)	Access
D	Constituency-Level Elections (CLE) Dataset	Washington University in St. Louis	USA	http://cle.wustl.edu/	1	1	0	0	elections		
D	NLS User Services	Center for Human Resource Research	USA	http://www.nlsinfo.org/	1	1	0	0	longitudinal surveys	public use files	
D	Cranfield Network on Comparative Human Resource Management (Cranet).	Cranfield School of Management	United Kingdom	http://www.cranet.org/data/data.htm	1	0	0	1	human resources management		
D	European Bioinformatics Institute	European Bioinformatics Institute	United Kingdom	http://www.ebi.ac.uk/Information/	1	0	0	1	bioinformatics data	FTP server provides a comprehensive archive of databases and software	FTP
D	Natural Environment Research Council	Natural Environment Research Council	United Kingdom	http://www.nerc.ac.uk/	1	0	0	1	environment		
D	Time Series Center	Harvard IIC	USA	http://timemachine.iic.harvard.edu/	1	0	0	1	time series		
D	Center for Hazards and Risk Research	Columbia University	USA	http://www.ldeo.columbia.edu/chrr/index.html	1	0	0	1	natural disasters		
D	National Ecological Observatory	NEON, Inc.	USA	http://www.neoninc.org/	1	0	0	1	ecology		

Group	Data Source	Institution	Country	Internet Address	Quality	Topic	Relevance	Size	Notes (topic)	Notes (collection)	Access
	Network										
D	International survey 2008	Critical Care nutrition :	Canada	http://www.criticalcarenutrition.com/index.php?option=com_content&task=view&id=18&Itemid=41	0	0	0	0	nutrition / nutrition practices		

Sources:

http://www.nessie-essex.co.uk/dataResources_comResearch.asp
http://www.gesis.org/en/services/data/portals-links
http://www.esds.ac.uk/International/access/guide.asp
http://www.sociosite.net/databases.php
http://www.ciser.cornell.edu/info/polls.shtml
http://www.ciser.cornell.edu/info/datasource.shtml
http://www.intute.ac.uk/socialsciences/statistics/
http://www.disc.wisc.edu/newcrossroads/links.asp?cat=Economic
http://3stages.org/idata/
http://libraries.ucsd.edu/ssds/econ.html
http://www.gesis.org/en/research/research-stay/european-data-laboratory/data-resources/context-information/
http://www.ifdo.org/network/index.html
http://www.edacwowe.eu/en/frmIndex
http://www.diggingintodata.org/Repositories/tabid/167/Default.aspx
http://oad.simmons.edu/oadwiki/Data_repositories