Title	**Data collection strategies: CESSDA organisations and their relation to data collections outside CESSDA (D10.5a)**
Work Package	WP10
Authors	Dimitra Kondyli, Maria Tzortzopoulou, Katerina Vezyrgianni. With contributions from: Anne Cornilleau, Marie Cros, Roxane Silberman and Paola Tubaro, in Section 3, and Chryssa Kappi in Part II.
Source	CESSDA-PPP survey; web survey; supplementary survey on European data producers.
Dissemination Level	PU (Public)

Executive Summary

This report refers to the study of the European landscape in reference to data collections in the social sciences across data, i.e.: data producing, managing, providing. This investigation forms part of the strategy set by the CESSDA-PPP project, towards identifying the potential for registering and exploiting the existing data production initiatives across Europe and at the country level, so that the new CESSDA infrastructure will encompass the widest possible network of data producers and providers and act as a broker between data producers and data users. Currently, CESSDA is more connected to the academic sector of the European research community, while the prospect of liaising with actors from the extended research community is opening with its future establishment as a European Research Infrastructure.[1]

Empirical material for the study was collected from a variety of sources: a) The CESSDA-PPP survey, b) Web survey on the profile of CESSDA Archives, c) Web survey on data producers-providers-archives operating in nine European countries which are represented by a CESSDA archive.

The fact that social science research data are being produced by a variety of 'agents' such as research groups, governmental and non-governmental organisations, profit making companies, local administrative units, leads to great variety in data production,

[1] Academic sector research: research made specifically for education and research purposes in academic institutions; actors from extended research community: research conducted by agencies producing business administrative data, commercially used research, etc.

and to difficulties in identifying and registering data at the national level, and even more so at European and international levels.

CESSDA archives possess a significant share of the total European production in terms of numbers and size of datasets, but their collections do not reflect the entirety of production in the European landscape.

Data collections existing outside CESSDA are equally rich in microdata and metadata, but their availability is mostly restricted to the level of metadata. Other types of data as products of secondary analysis, i.e. research production that relies on the analysis of primary data and is available without the original data, form a large part of the production and are freely available.

Qualitative data are under-represented in CESSDA Archives; a large number of CESSDA members express the need for the acquisition of this kind of data. Qualitative data are difficult to trace and identify outside CESSDA, at the level of production; this is mainly due to the low rate of digitalization and archiving practices for this type of data.

There is general convergence between CESSDA and non-CESSDA organisations in terms of subjects available in their data collections. However compared to organisations outside of CESSDA, CESSDA members are lacking in the following subject areas: Economics, Trade, Industry & Markets, Education, Housing & Land Use Planning, Natural Environment.

There is an indication of a trend in specialising on certain Social Sciences and Humanities (SSH) fields at country level. This reflects different research traditions and further investigation is needed with special focus at the country level.

CESSDA archives do not require exclusive deposit. Collection policies generally rely on networking and personal contact with producers. The majority rely on liaising with the public sector.

The organisational structure of CESSDA archives in terms of central versus distributed establishment is equally represented between the two forms. The range of subjects in the collections is not differentiated by the organisational structure, although networking as an activity to attract data is the most preferred method, applied by organisations of both forms.

An impact of technological advances is the emergence of multiple actors in the new data landscape. The expectation of users to deal with data harvesting might affect generic data services. The challenge for CESSDA as a research infrastructure will be to keep a balance between high quality data resources –in terms of discovery- and high quality of the data provided.

It is clearly evident from the above that good knowledge of the producers existing in each country is necessary for the facilitation of these activities. Research on the producers in countries should ideally be constantly performed by representatives of the archives in each country. That way, a more complete landscape would be drawn.

Indeed, in cases where our key informants engaged seriously in editing and completing our web research, the corresponding country landscape became clearer and broader.

List of recommendations for the future cessda-ERIC in reference to data collection policies

Recommendation #1:

The cessda-ERIC must develop a strategic plan on networking activities spread across all types of producers at country level; the ERIC networking plan must take into account local –i.e.: national distinctiveness, while providing harmonised procedures for attracting and evaluating microdata collections produced by local agents lacking a centralised mechanism of data management and dissemination.

Recommendation #2:

Data archives with low visibility at country level must be supported by the cessda-ERIC networking and promotion initiatives, so that their collections can be expanded in terms of geographical coverage, and quantity.

Recommendation #3:

The disadvantages accompanying the requirement of non-exclusiveness in data deposit must be turned to advantages within the cessda-ERIC; professional handling of data, quality of services, abundance and variety of metadata, high performance in specialised tools –e.g. data harmonisation tools must be promoted and advertised, so that researchers will eventually be reinforced to prefer the cessda-ERIC over other sources for the same data collections.

Recommendation #4:

To take action to facilitate the acquisition of qualitative data collections along four lines: a) identify demand for qualitative data across all cessda-ERIC members; b) target research on specialised tools and services for qualitative data; c) facilitate provision of qualitative data collections by campaigning the advantages of archiving them; d) establish long term collaborations with other specialising in handling qualitative data.

Recommendation #5:

Outreach activities of the cessda-ERIC must focus on identifying research initiatives of other disciplines which not only touch upon social issues, but also provide ground for methodological developments to support social science data collections, such as geographical classification of data; co-operative activities must

be initiated and tools must be developed and/or adapted for applications in social science.

Recommendation #6:

The cessda-ERIC must liaise with data producers which do not fit its membership schema but are important in populating its collections. A distinctive group of such producers is the polling; their main strength lies on the abundance in production of data relevant to political behaviour, while their main weakness lies on the restricted usage of such production and poor exploitation of microdata for other research purposes. The cessda-ERIC must work on collaboration agreements, exchange expertise, e.g by providing data management and dissemination services in exchange for receiving polls data of historical and comparative value.

Recommendation #7:

Engage in research to identify the '*study topics markets*' existing within the CESSDA network, which reflect strong and consistent research traditions in a selection of topics; based on the identification of 'strong research traditions' across different topics among members, to capitalise on their strength by engaging in expertise exchanges.

Recommendation #8:

Observing the local production and how it can be accessed must be an ongoing goal for the new ERIC. The new organisation must engage in promotional initiatives to act as a gateway –even, in certain cases, at the level of information on the sources of data and facilitating a homogeneous culture of sharing and re-use of data across European countries.

Recommendation #9:

Equal consideration must be taken on data sources: the cessda-ERIC 'in-shop' acquisitions –i.e.: collections held by members, and outside sources –e.g.: journal publishers requiring deposit of data in reference to publications, or National Statistical Institutes (NSI) collections. We must ensure that both sources are visible through the Portal with analytic documentation on accessing, thus engaging in 'best practices' in reference to the wide research community and promoting cooperation with other data publishers. Possible modes of cooperation must be a permanent item on the agenda of the cessda-ERIC outreach activities.

Recommendation #10:

The cessda-ERIC, during its construction phase, must engage in research at country level with the purpose of identifying the localities of research traditions: from production to exploitation; case studies with in-depth approach in 'extreme' cases might also be needed.

Recommendation #11:

The ERIC must include in its governance schema an advisory body representing the countries and consisting of active researchers, with the purpose of monitoring the local research needs and activities and providing feedback to the ERIC.

Recommendation #12:

Given the fact that knowledge on the current situation in reference to the perspective of the users – their preferences in certain data collections, satisfaction on service provided, the end-products based on data acquired - is not currently available in an harmonised manner across CESSDA, it is recommended that cessda-ERIC will take action in: a) identifying users' needs; b) auditing their 'behaviour' for effective promotion of services. This can be accomplished in two phases: a) during the construction phase, doing comparative research on users' profiles; b) during the implementation phase, requiring homogeneous reporting on users' data. The details of this type of reporting can be included in the SLA as an amendment to reporting procedures relevant to user registration and authentication.

Recommendation #13:

It is recommended that cessda-ERIC shall capitalise on the collective expertise and human capital of the current CESSDA members in reference to existing relationships with various types of producers, in order to set up an *expert group* for the design of a strategy to attract producers across Europe and the world, provide conditions for improving poor relationships and stabilise high quality relationships. This expert group must consist of individuals specialised in an array of subject areas, so that data production of both 'conventional' and 'un-conventional' areas of research can be attracted.

Recommendation #14:

New data collections, expanding in numbers and varied in kinds, require specific technology for their management; it is particularly stressed that the cessda-ERIC must engage in activities to heighten the professional level of staff by investing in the employment of specialised experts and in training programmes. At the same time a constant course of action is needed for the development of tools which will support archiving of new types of data.

Recommendation #15:

It is recommended that a set of guidelines and procedures is produced, serving as the minimum actions to be taken for the management of data collections. This tool must also include operational definitions of key concepts of the data archiving profession in all languages, complementary to the work of WP4 on the ELSST thesaurus and the Controlled Vocabularies used for documentation.

Recommendation #16:

It is recommended that cessda ERIC will consider incorporating into the design of its Portal the facility of linking with research resources internationally, which can also be updated by authenticated users and screened through clearly set criteria of quality, in accordance with the quality criteria used for acquisitions in the infrastructure.

Recommendation #17:

It is recommended that cessda-ERIC will set-up a feedback mechanism – offering appropriate incentives - from users of datasets to the infrastructure, for referring back the products of their work, based on datasets acquired either by the infrastructure or directly disseminated to them through other web sources.

Contents

I. Task 4:

1. Introduction

According to recent statistics, more than 25,000 datasets are held across CESSDA Data Archives. Through the CESSDA Data Publishers 4,347 studies are currently available. Yet, the CESSDA network is not in a position to know the comparative significance of this figure within the European Social Sciences data landscape, because there is no registered indication of the number of datasets produced, as a whole, by research activities and generally datasets of interest for social sciences, across Europe.

Social science research data are being produced by a variety of 'agents' such as, university based research groups, governmental and non-governmental organisations, profit making companies, local administrative units, etc. Such disparity in data production, which is potentially useful to a wide user community, is difficult to identify and register at the local level, and even more so at European and international levels; yet, the growth of Research Infrastructures and the constantly developing need for comparative data on the wide range of social issues, require identification, access and usage facilities for the maximum possible range of data production. The demand was evident early enough in the process of setting a European Strategy for data in the Social Sciences and the Humanities and was explicitly expressed by the RISSH workgroup of ESFRI in their report of 2004. The group's central proposal was to establish a European Research Observatory for the Humanities and Social Sciences [EROHS], which would be guided by four core principles, two of which are highly relevant to this work, - namely, principle 1: "*The facilitation of access to and sharing of existing European and national data, thereby more efficiently and effectively linking data resources already available.";* principle 3: *"The generation of new and genuinely European data. This will involve both the collection of new data and the digitalization of materials not currently computerized"* (RISSH, 2004).

This strategic demand led CESSDA, which works towards becoming the major central facility in Europe for Social Science data, to set amongst its goals the ground for mapping the European data landscape. WP10 of CESSDA PPP and, particularly, Tasks 4 & 5 were designed to initiate this goal by collecting information on data production, data collections and their agents across CESSDA archives and non-CESSDA organisations producing data across Europe, within the preparative scope of the project.

This report is on the work of Tasks 4 & 5 of WP10, highlighting the existing situation in reference to Social Sciences data collections and data production in all CESSDA member, as well as a selection of non-CESSDA across nine (9) European countries;

the report includes recommendations for further actions towards the goal of incorporating all collections under the new CESSDA Hub within the governance and legal framework of cessda ERIC.

1. Structure of the report

The report is structured in two parts as follows:

Part I is the report on the work of Task 4 –data collection strategies and part II is the work on Task 5 –recommendations.

Part I is divided in 3 chapters. In Section 1 the basic terms used throughout the report, their frequency statistics and operational definitions are presented. This section introduces the working typology for Task 4; this typology is used as the framework for further analysis.

Section 2 is a short overview of the situation in Europe in reference to data collections, drawn from bibliographical evidence and the study of the sources used for this report as they are described in the Methodology section.

In Section 3 views and information are presented on the emerging landscape in Europe in a wide perspective of common and less-common kinds of data in the domain of Social Sciences.

Part II consists of 4 chapters; it is an overview of the findings presented in the report and includes a list of recommendations drawn from these findings.

Finally, Parts III to V consist of additional information on the research performed for Task 4.

2. Methodology

The central question for Tasks 4 - 5 can be summarized as follows:

What is the situation in Europe in terms of data collections?

To respond to it we used the following sources of information:

CESSDA survey

The CESSDA-PPP survey was conducted during June-July 2008. The part of the questionnaire which was relevant to the investigation of WP10-T4 consisted of 10 questions spread across three sections: profile of data collections, collaborations and strategy policies. The questions designed for T4 aimed at registering the current activities employed by the organisations for their data collections and providing information on their availability, management and usage.

The population of the survey included CESSDA and non-CESSDA organisations. Non-CESSDA organisations were identified by key informants, i.e., CESSDA and CESSDA

PPP members were asked to provide identification of organisations which are engaged in activities relevant to social science data, in the countries they represent. The CESSDA-PPP survey was designed to be administered in organisations performing archival activities. The majority of CESSDA Archives responded, whereas non-CESSDA organisations were under-represented, (18 and 5 respectively); therefore the survey alone could not be a reliable source for the identification of data collections by actors outside CESSDA. Thus, we turned to additional sources.

The attempt to map the data landscape in Europe demanded the use of a variety of sources for discovery, identification and classification of the maximum possible number of 'agents' engaging in the production of social science data. Since the CESSDA-PPP survey alone was not an adequate source for the purposes of WP10-T4, other sources used for discovery were:

- **The web**

Navigation and focused searching through the web pages of CESSDA and outside CESSDA organisations was performed in two phases.

Navigation through the web pages of CESSDA data archives (web survey A CESSDA)

Web navigation A was performed exclusively through the web pages of CESSDA members from February to April 2008 as a pre–survey in order to draw a 'members' profile; the pre-survey contributed to the formulation of the questions relevant to collection policies examined in the main survey. The main information collected was relevant to the legal character, activities of CESSDA member (i.e. archiving, research, training etc.) and type of collections they preserve. Also, we tried to identify those legal entities that produce, deposit and/or provide datasets to CESSDA Archives.

Exploration and navigation through the web pages of organisations outside CESSDA (web survey B (non-CESSDA)

Web exploration and navigation B was performed during March-April 2009 for the discovery of data producing organisations outside CESSDA. Since it was not possible to collect information from all countries participating in CESSDA –due to time restrictions, linguistic barriers and variable user friendliness of web pages- selection of countries was guided by the analysis of the CESSDA survey data according to a geographical grouping. Information was collected across 9 countries: France, Greece, Italy, Norway, Romania, Slovenia, Spain, Sweden and United Kingdom. The basic criterion for each country selection was its geographical place, so that all parts of Europe are represented (Northern-Central-East-West-South). Availability of information and contacts was another leading criterion for the selection of countries. Therefore, the sample of countries is not to be considered representative.

The web pages exploration and navigation aimed at identifying organisations which produce and/or provide data relevant, or potentially[2] relevant, to the social sciences. One hundred and ninety-five selected non-CESSDA organisations were examined with respect to the following questions:

- What types of data?
- Across which scientific topics?
- To which organisations are data located?
- How are they disseminated?

Based on the above, an inventory of data organisations was created, which is available in Access database and SPSS file format, but is not included in the deliverables of WP10 -Tasks 4 & 5.

- **Personal communication with key persons in each country**

In certain cases feedback and additional information was requested by key informants in the CESSDA organisations representing the countries. This source was used in support of the other sources where information was difficult to be traced due to language barriers and to confirm the reliability of information obtained.

- **Information exchange within the CESSDA-PPP work packages**

There have been cases where concurrent work on other work packages and/or tasks coincided with searching for WP10-T4 and revealed information useful for the partners involved.

- **Previous studies**

For the purposes of this report, background information from previous studies was considered, regarding SSH researchers' needs on data production as well as on the development of infrastructures at the national and/or European level.

[2] We use the term "*potentially relevant"* to denote those types of data which are not directly produced for research/academic purposes or are not in a physical condition for analysis, nevertheless could be used by researchers. Examples are: population records, incidence records, business reports, etc.

2. Report on surveys

1. Section 1

III.1.a. Typology of non-CESSDA and general activities profile

The CESSDA PPP survey allowed identification of key characteristics of the respondent, thus drawing their activities profile. Before engaging in discussion about each characteristic the general profile of non-CESSDA organisations will be drawn. The aim was to identify those *organisations that produce, and distribute data*[3] *to researchers in fields of Social Science and Humanities, either nationally or internationally.*

Towards this purpose operational definitions were provided concerning the main objects of our research, i.e.: the various types involved, their roles and functions in order to carry out a supplementary survey on organisations which produce social data.

In terms of roles, criteria were set for the definition of:

Data producers (PD)[4]: to be able to identify an organisation as PD, it must cover **all** the following criteria:

- Regularity in the production process;
- Data production for research or other purposes must be included in the organisation's actual practice, its aims and targets are referred to as such in its statutes and/or institutional regulations;
- The data produced must be in fields of SSH, or of potential[5] use for research purposes in the fields of SSH;
- A data cataloguing mechanism must exist, which allows the function of dissemination.

Data Providers (PV): to be able to identify an organisation as PV, it must:

- Serve as physical or electronic repository;
- Have a data cataloguing mechanism;
- Have data dissemination mechanism for either its own data or data produced by others;
- Data must be relevant to the fields of SSH, or of potential[5] use for research purposes in the fields of SSH.

Data Archives (DA): Data Archives have multiple functions; they possess all the functions arising from the role of Data Providers, that is:

- Serve as physical or electronic repositories;
- Have a data cataloguing mechanism;

---

[3] We refer to data produced by either quantitative and/or qualitative methods of research, or both.

4 See also the DDI2 definition: the producers of the data collection are persons or organisations with the financial or administrative responsibility for the physical processes whereby the data collection was brought into existence. (http://www.ddialliance.org/related/cessda-rec.pdf)

[5] See also footnote #2

- Have data dissemination mechanism for data produced by others;
- Are specialised in data relevant to the fields of SSH, or of potential[5] use for research purposes in the fields of SSH and in addition:
- engage in activities of ensuring the quality of data acquired;
- certificate the use of data;
- certificate the users of data;
- perform archiving activities with tools and services;
- develop and/or maintain systems and tools for classification and archiving;
- Organise supplementary documentation.

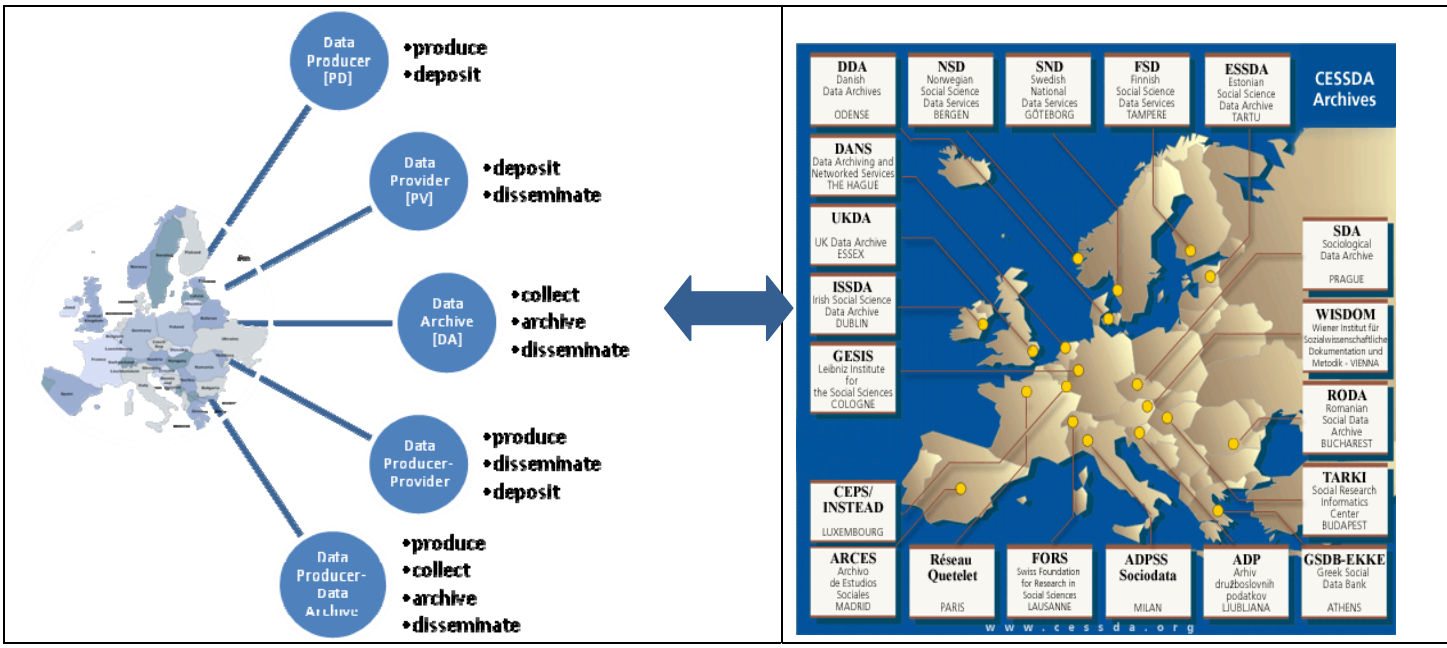


IMAGE 1: TYPOLOGY OF DATA

For the purpose of identifying data collections across the European landscape, the preferred focus is on data production; therefore the role of related to data production is important. We divided non-CESSDA organisations into subcategories related to their activities in relation to CESSDA organisations, -i.e.: Actor in Re to CESSDA Archive'. The values were:

- **Parallel actor defined as:**

Those organisations in the country which have data collections in a wider range of topics relevant to the social sciences; these organisations will be defined as parallel actors if and only if their archiving/disseminating activities are not performed through the national organisations or the corresponding CESSDA organisation.

- Concurrent actor defined as:

Those organisations in the country which produce data and/or engage in archiving/dissemination activities in a restricted or situational manner. They may be organisations of any legal character and of any size; use of limited resources and

infrastructure for various purposes (e.g. self-publishing of data on the web, projects, etc.) is more likely.

Further refinement in the typology of organisations was made, which was based on information related to the kind of data produced, the subjects covered in the collections and modes of dissemination used. Thus, parallel non-CESSDA organisations were assigned either of two properties:

- Engaging with large scale collections at the national level[6]
- Other Data organisation.

Finally these were coded according to the typology based on three different types of roles and their combinations: 'Data Producers', 'Data Providers', and 'Data Archives'. Each role is defined by the main activities of the organisation (Image 1).

Thus, the 'ontology' of data production could be represented as follows:

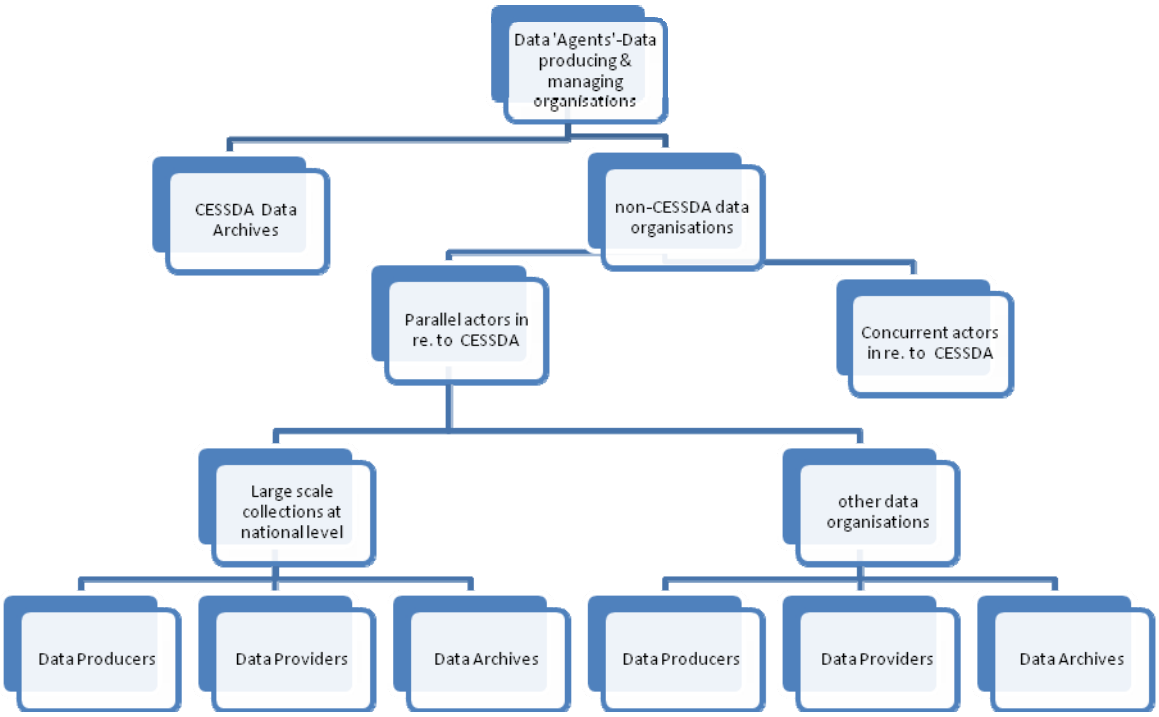


Image 2: A PROPOSED ONTOLOGY OF DATA

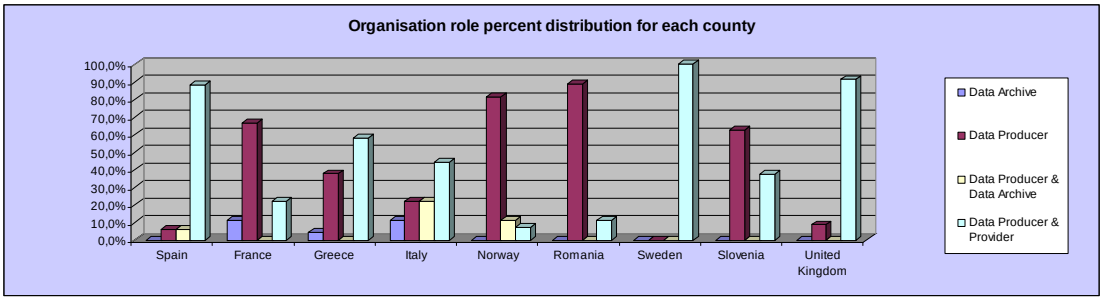
[6] See: definitions -Appendix A

III.1.b. Characteristics of the organisations based on their activities

- **Organisation role in the data landscape**

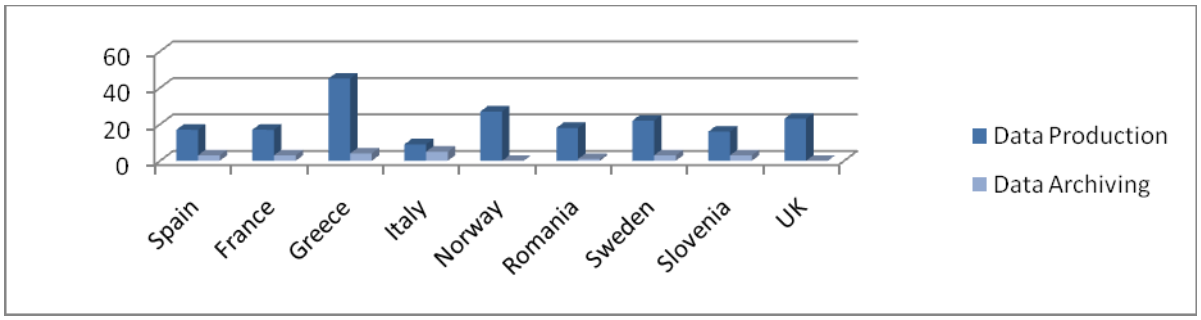
Figure 1 shows the distribution of the different roles across the countries of study. Since the CESSDA Data Archives representing each country were studied separately, they are not shown in this Figure.

Figure 1



Figures 1 and 2 show the distribution of 195 non-CESSDA studied: 51.3% are 'Data Producers and Providers', 42.1% are 'Data Producers', whereas only 2.6% are 'Data Archives'. Specifically, the majority of registered organisations of Spain, Greece, Italy, Sweden and United Kingdom is 'Data Producers and Providers', while in France, Norway, Romania and Slovenia there are mainly 'Data Producers'. Organisations with the exclusive role of Data Provider, as was defined for the purposes of this study, were not identified in the sample.

Figure 2

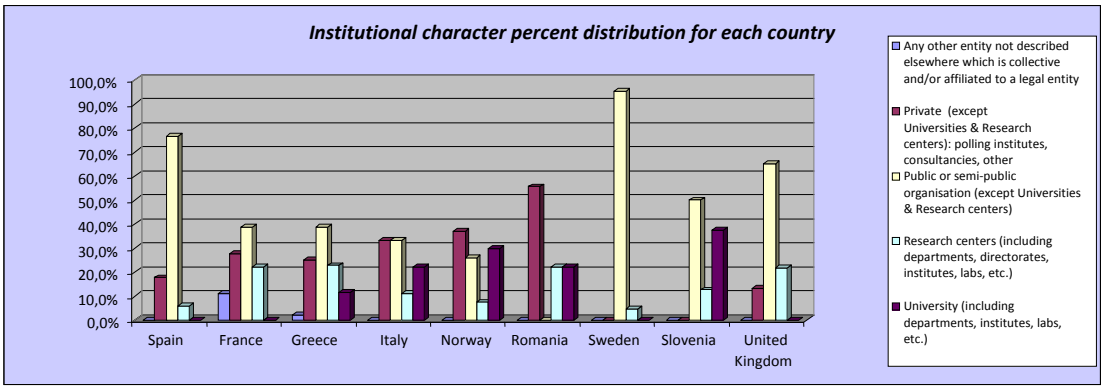


- **Institutional character**

Of the non-CESSDA organisations studied, 46.9% were 'Public or Semi-Public organisations' (excluding Universities and Research Centres); 23.2% were 'Private Organisations', such as Polling Institutes and Consultancies. 'Research centres' and 'Universities' were registered to a lesser extent (15.5% & 12.9% respectively). Viewed across countries, more 'Private Organisations' are recorded in Romania, Norway and to

a lesser extent Italy. This finding cannot be accurately interpreted since the factor of web visibility was significantly filtering our research.

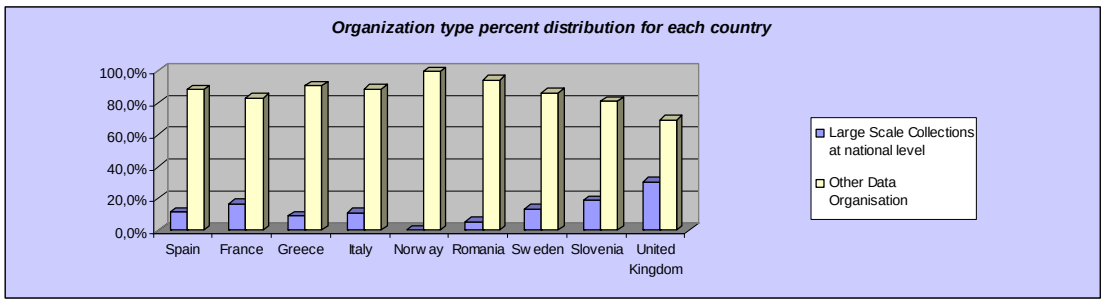
Figure 3



- **Organisation type**

The majority of the cases registered for this report, is 'Other Data Organisations'; However, in some cases, such as in UK and to a lesser extent in Slovenia and France there are of 'Large Scale Collections at National Level (LC)'.

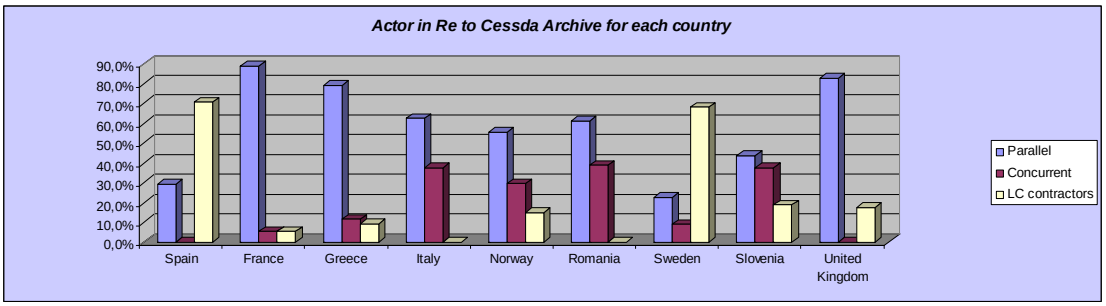
Figure 4



- **Organisation activity in relation to CESSDA**

The majority of the present sample in total was coded as 'Parallel Actors'. However, there are cases such as Spain, as well as Sweden in which 'Large Scale Collections at national level contractors' are much more compared to 'Parallel Actors'. Also, the percentage of 'Concurrent Actors' tends to be rather high in cases of Italy, Romania, Norway and Slovenia, whereas in Spain and UK this type of actor is not observed at all.

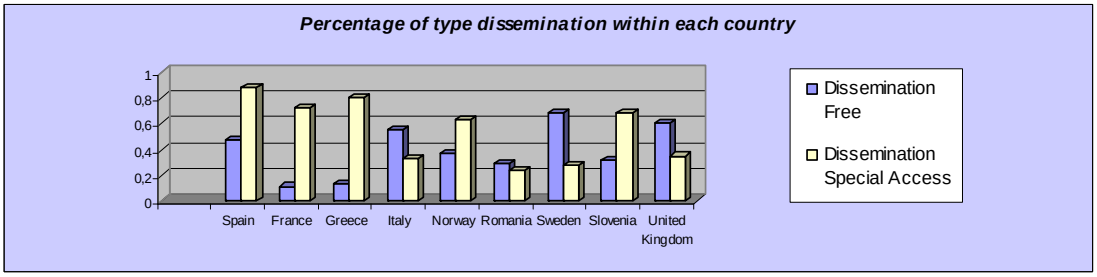
Figure 5



- **Dissemination**

183 out of the 195 non-CESSDA organisations disseminate their production; the manner of dissemination was categorized as "dissemination free" and "dissemination special access". Figure 6 shows the distribution of these two categories in the 9 countries sample.

Figure 6



The majority of cases disseminate their data with 'special access'. However, across countries there are cases (e.g. Italy, Romania, Sweden and UK) in which more data are available for 'free' than for 'special access'.

2. Section 2

The situation in Europe in terms of data collections

The European landscape concerning SSH data production has been changing rapidly in the past 10 years because of the development of e-infrastructures, e-repositories, and adequate tools. These applications have contributed to a 'burgeoning data deluge'[7].

[7] Lyon L. , Dealing with Data: Roles, Rights, Responsibilities and Relationships Consultancy Report, 19/5/2007, UKOLN, University of Bath http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf , downloaded on the 12/06/2009.

From reports on projects targeted to research and SSH from a variety of viewpoints, the following can be summarized:

- SSH scientists were mostly occupied in the public sector.

The public sector, including public research institutes and universities were the main domain in which SSH scientists were almost exclusively occupied[8]. This could possibly explain the large percentage of public institutes registered as data producers in the web Survey B. As private organisations started emerging in the knowledge production landscape, the situation changed; technological changes have facilitated the marketing of information[9].

- The emergence of private organisations widens the research landscape

Research is privatised, follows the rules of the market and dissemination and distribution functions have an ad hoc character[10].

- Connection with higher education

In certain European countries there is a better management of higher education and research functions; i.e. UK and northern European countries. The investment in R & D in Sweden is the highest internationally (4.2% of the GDP) and in Finland the percentage of researchers is the highest internationally[11].

III.2. Data production across Europe: CESSDA and non-CESSDA data collections

This part describes the main line of actions taken for the task of "drawing a map" of data production in Social Sciences and Humanities across Europe.

The following analysis is descriptive, based on frequencies and percentages; further statistical measures are not possible, due to the heterogeneity of samples and procedures. The population used consists of two separate sets:

CESSDA organisations –Archives (N=20);

non-CESSDA -producers and providers relevant or potentially relevant to SSH data (N=195 across 9 countries - see Appendix C).

In this section results will be presented in the following way:

I. Results based on CESSDA organisations;

---

[8] Report (2008) of the Project: Social Sciences and Humanities for Europe (SSH futures) Instrument: Specific Targeted Research Project, Thematic Priority: 7 Citizens and Governance in a Knowledge-Based Society, http://www.iccr-international.org/sshfutures/docs/SSH-FUTURES, downloaded on the 25/4/2009.
[9] op.cit 2
[10] op.cit 2
[11] D. Sotiropoulos (2006) Survey for the higher educational system in Greece, Athens, ELIAMEP www.eliamep.gr downloaded on the 21/01/2008.

II. Results based on non-CESSDA organisations;

III. Discussion of similarities and differences.

A basic aim of Task 4 was a typology of organisations according to data production and provision properties. The basic classifying element for this typology is the production and management of data collections. Data collections were studied according to the following main characteristics:

a. Number of datasets held (information is available only for CESSDA Archives)

b. Kinds of data

c. Social science subjects covered by data collections

d. Data deposit practices and regulations

e. Data collection strategies (information is available only for CESSDA Archives)

f. The organisations' needs in data collections in terms of kind, number, subjects (information is available only for CESSDA Archives).

III.2.a. Characteristics according to geography

An important dimension in our analysis is the geographical dimension, since we are discussing similarities and differences across countries in Europe. One of the hypotheses adopted was that different parts of Europe would differ in traditions, culture and practices in handling research data. To test this hypothesis the CESSDA Archives were assigned to four geographical groups:

- Northern Europe[12]
- Southern Europe[13]
- Central Europe[14]
- North-West Europe[15]

Separate analysis of results was carried out based on this grouping as supplementary to the general analysis. The results revealed differences in certain practices among the geographical groups. These differences are summarised in Table 1. Findings based on geographical groupings were used as leading criteria in the selection of countries for the survey of non-CESSDA, i.e.: the web pages exploration and navigation (web survey B) sample contains countries from all geographical groups.

[12] Finland, Sweden, Norway, Denmark.
[13] Spain, Italy, Romania, Greece
[14] Germany, Czech Republic, Hungary, Slovenia, Switzerland
[15] Netherlands, Ireland, U.K., France, Luxembourg

Table 1: Main characteristics of CESSDA Archives by geographical group

	Subjects of data collections	Qualitative vs quantitative data	Data collection policies	Networking activities
Northern Europe	*Equal distribution of subject areas*	*Some qualitative data*	*A variety of data collection strategies*	***Express need for expansion of cooperation***
Southern Europe	***Unequal distribution of subject areas***	***Very little qualitative data***	***Attract data mainly with networks***	*Have cooperation with a variety of*
Central Europe	*Equal distribution of subject areas*	*Some qualitative data*	*A variety of data collection strategies*	*Have cooperation with a variety of*
North-West Europe	*Equal distribution of subject areas*	*Some qualitative data*	*Mostly bi-lateral co-operations*	*Express need for cooperation with private*

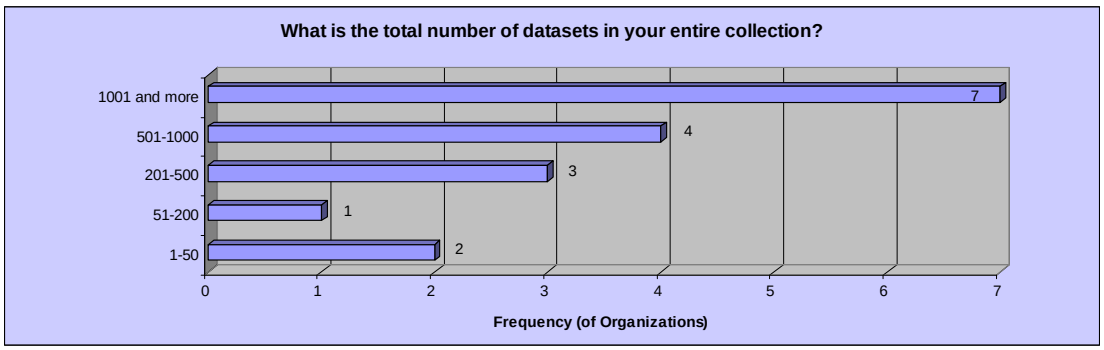
Note: bold characters in some cells are used for stressing differences in characteristics

III.2.b. Data Collections across CESSDA and non-CESSDA

Number of datasets held (information is available only for CESSDA Archives)

Information on the number of datasets available is only relevant as part of the profile of data archives; this information does not represent anything about the actual production of social science data, their availability and the conditions of dissemination at a European level. Further, it is a measure bearing the possibility of misinterpretation, in the sense that the term "dataset" can have various meanings. It has already been said that CESSDA Archives have a significantly large collection of datasets.

Figure 7 (Only for CESSDA Archives)



If number of datasets is an indication of the 'size' of an organisation, then from Figure 7 we can conclude that CESSDA largely consists of 'large' archives. This of course is not the case, in reality. To be able to interpret the distribution of Figure 7 we should have been able: a) to ensure that the term "dataset" is generic and clear across all respondents and b) to have a clear view of the role of each CESSDA organisation across Europe in relation to collecting the total production in the social sciences in their countries. The

measure 'number of datasets' was impossible to be obtained for organisations outside CESSDA; comparisons could only be meaningful with organisations engaging in archiving activities and a few organisations of this type were traced. Out of a 195 non-CESSDA organisations studied across 9 countries, five (5) were characterized as Data Archives.

Overview box #1

Most CESSDA Archives have more than 1000 datasets; this finding alone is not a reliable measure of the size of an organisation. Combined information on data production, types of data and usage statistics is needed for a better view in terms of data production across Europe.

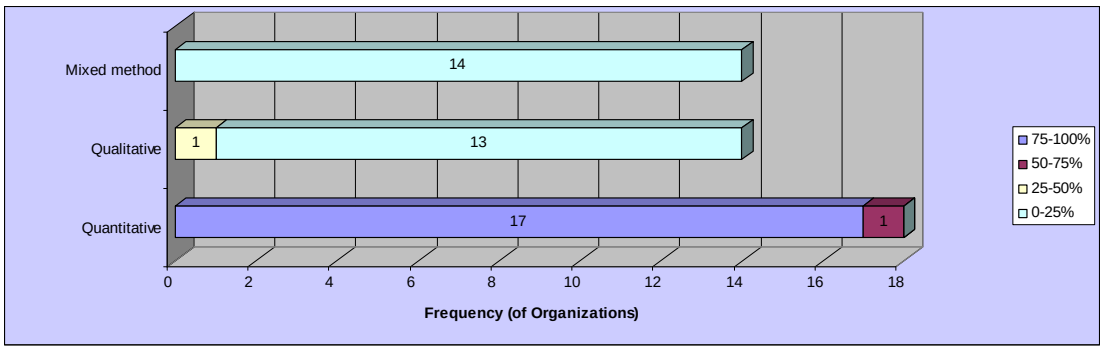
Kinds of data

Kinds of data were studied in terms of method of production i.e.: quantitative vs qualitative research methods, and physical properties of data i.e.: microdata, macrodata, metadata or other types.

i. *All CESSDA organisations*

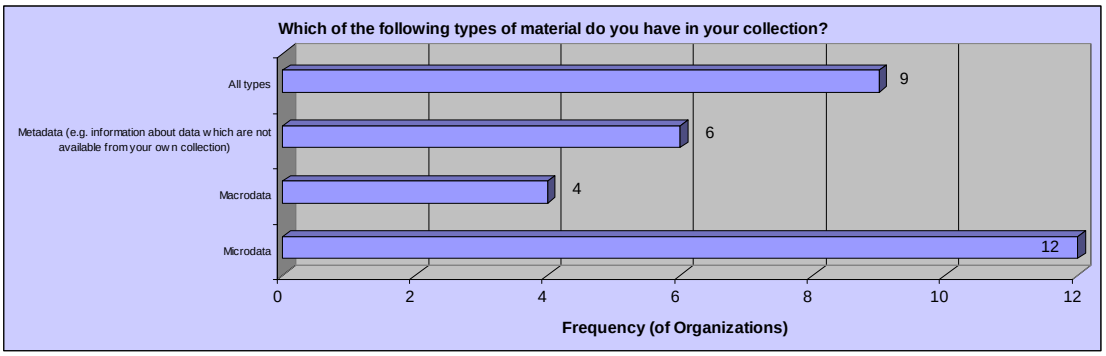
Figure 8 shows the percentages of quantitative & qualitative data collections across CESSDA organisations. Collections on qualitative data are of the smallest percentage in the majority of CESSDA organisations.

Figure 8 (For CESSDA Archives only)



As naturally expected of Archives, microdata form the majority of the CESSDA collections (figure 9).

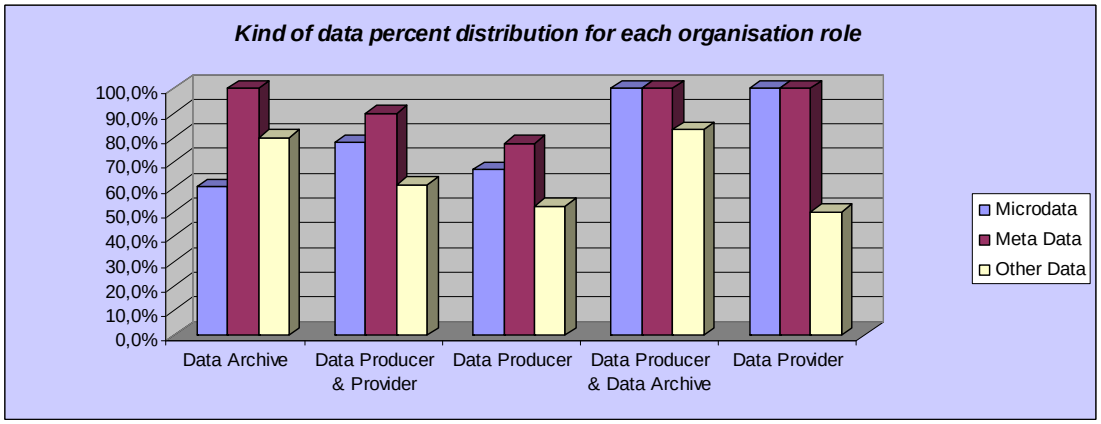
Figure 9 (For CESSDA Archives only)



ii. non-CESSDA sample

The production that exists outside CESSDA (at least in the countries studied) also has an abundance of microdata (figure 10). What is very important about the view outside CESSDA is that metadata production is larger, compared to microdata, and there is also significant production of other types of data; these are usually reports, tables, indexes, etc. Across countries, there is the same tendency with the exception of Sweden in which 'raw data' production is larger than 'microdata'.

Figure 10 [non-CESSDA]



Overview box #2

Qualitative data are under-represented in CESSDA Archives, whereas in terms of physical properties all types of data are offered (micro-meta-macro-data); in outside CESSDA, although microdata production is large, more collections of metadata exist, while other types of data are a significant figure. Tracing what these "other types of data" are and how they can be accessed must be a goal for the new cessda-ERIC.

Social science subjects covered by data collections

i. All CESSDA organisations

The classification of social science subjects adopted for the purposes of this study was the CESSDA Topic Classification Catalogue. In general, all topics are included in the CESSDA archives' collections, but there is uneven distribution among the topics; some are over-represented and this pattern is evident in most archives. For example, 18 out of 20 organisations have collections in "Politics".

The ranking of topics follows, naturally, the most traditional areas in the Social Sciences; thus, we see that the following topics have a large share in the collections of most organisations:

Ranking	Topics
1	Politics
2	Social Stratification and Groupings
3	Society and Culture
3	Labour and Employment
4	Demography and Population
5	Social Welfare Policy and Systems
6	Education
7	Economics
7	Health

In contrast, less than 50% of the organisations have collections in the following topics:

Ranking	Topics
1	History
1	Information and Communication
1	Transport, Travel and Mobility
2	Housing and Land Use Planning
2	Law, Crime and Legal Systems
3	Trade, Industry and Markets
4	Psychology
4	Reference and Instructional Resources
4	Science and Technology
5	Natural Environment

Conclusions for the distribution of topics in the data collections across the organisations cannot be easily drawn based on one question only; data archives populate their collections on the basis of country and institutional regulations. It can also be the case that the profile and identity of the organisations is specifically targeted to certain areas, whereas other subjects are covered by other institutions; this remains to be further investigated country by country. We might also hypothesize that *research production* towards certain subjects can be affected by the *research tradition* of each country.

ii. non-CESSDA sample

Production was investigated in the selected countries sample. The same list of topics was used. In comparison with the grouping of subjects based on the distribution of collections within CESSDA, we observe that the same grouping applies for the non-CESSDA but with one additional subject: 'Trade, Industry & Markets'; figure 11 shows the distribution of collections in the organisations operating as parallel actors outside CESSDA. We selected the subset of parallel actors because of their activity in a wide range of subjects within SSH; this characteristic of parallel actors is mostly relevant to our analysis of the range of subjects existing outside CESSDA. Table 2 is a comparative list showing the ranking of the most frequent subjects in CESSDA and non-CESSDA organisations.

Figure 11 [non-CESSDA]

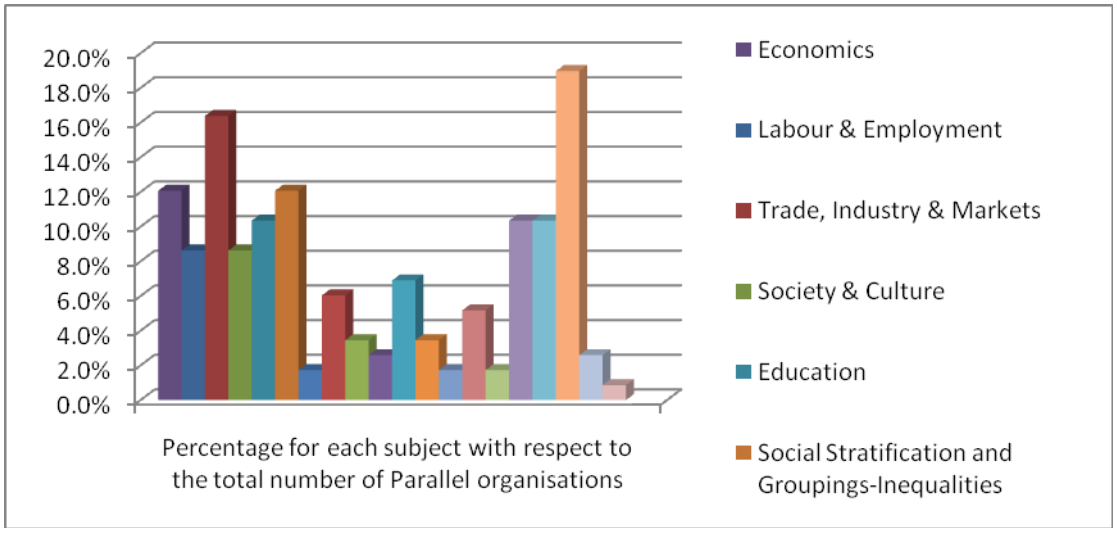


Table 2: (comparative ranking CESSDA and non-CESSDA)

Subjects	**CESSDA**	**NON-CESSDA**
Politics	1st	1st
Trade, Industry and Markets	n/a	2nd
Social Stratification and Groupings	2nd	3rd
Social Welfare Policy and Systems	3rd	4th
Education	4th	4th

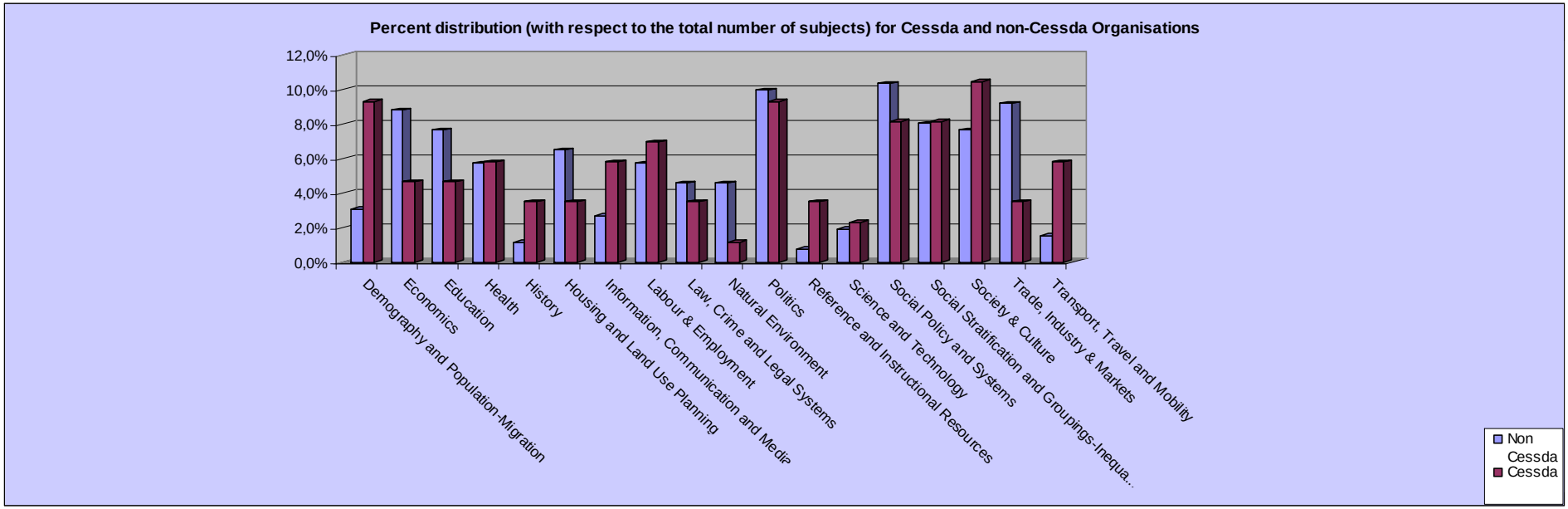
Health	5th	4th

To be able to make reliable comparisons, the distribution of subjects studied in the CESSDA survey was restricted to the 9 countries for which non-CESSDA organisations are also recorded. In Figure 12 we observe the distribution of subjects in data collections of CESSDA and non-CESSDA organisations in 9 European countries. The overall view is that the CESSDA archives in the corresponding countries are not lacking in data collections across this range of subjects. Looking in context though, we observe that data collections in certain subjects are considerably outside CESSDA.

The subjects for which a large amount of data collections exist outside CESSDA in the particular countries studied, are:

- Economics
- Trade, Industry & Markets
- Education
- Housing & Land Use Planning
- Natural Environment

Figure 12


Percent distribution (with respect to the total number of subjects) for Cessda and non-Cessda Organisations
12,0%
10,0%
8,0%
6,0%
4,0%
2,0%
0,0%
Demography and Population-Migration
Economics
Education
Health
History
Housing and Land Use Planning
Information, Communication and Medi
Labour & Employment
Law, Crime and Legal Systems
Natural Environment
Politics
Reference and Instructional Resources
Science and Technology
Social Policy and Systems
Social Stratification and Groupings-Inequa...
Society & Culture
Trade, Industry & Markets
Transport, Travel and Mobility
Non Cessda
Cessda

Comparing this finding with the expressed needs of CESSDA Archives for certain kinds of data we find significant convergence in the subjects needed (Figure 13)

Figure 13 (For CESSDA Archives only)

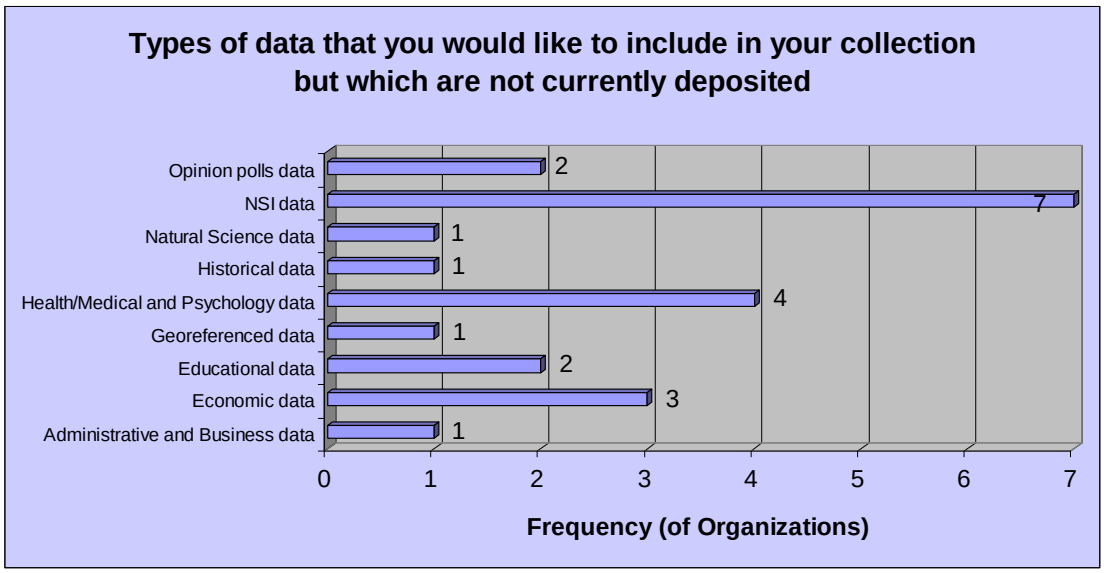


Figure 14 shows graphically the calculations based on data collections in subjects appearing more than 10 times to the total number of non-CESSDA organisations within countries. These subjects are 12, distributed across countries. Analysis showed that there is a tendency for each country (except Italy and Greece), to be oriented towards a specific topic more than to others. Thus, Spain tends to focus more on 'Social Stratification and Groupings - Inequalities'. France focuses on 'Health issues', Norway focuses in 'Trade, Industry and Markets', Romania focuses on 'Politics', Sweden focuses on 'Social Policy and Systems', Slovenia focuses on 'Society and Culture' and UK focuses on 'Education'.

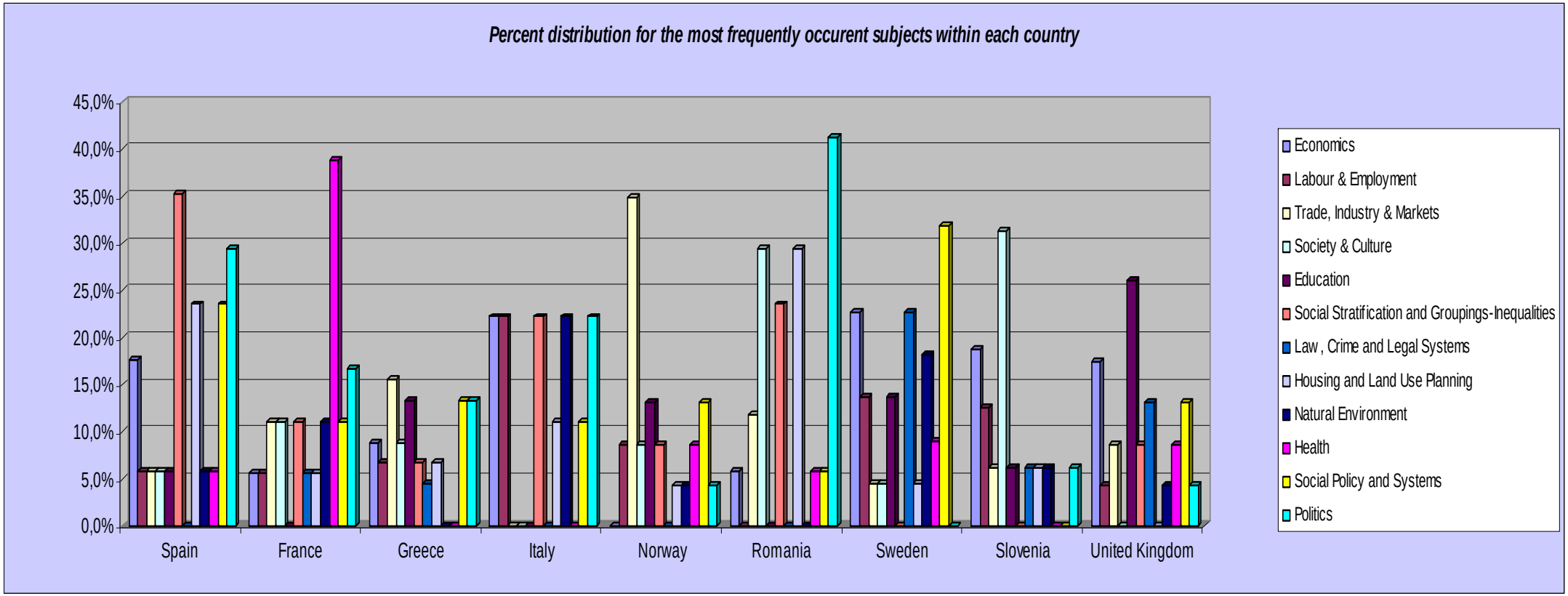
For further and detailed information on data collections by country and organisations, there are tables in Appendix B.

Overview box # 3

There is convergence between CESSDA and non-CESSDA organisations in terms of subjects in their data collections. Collections lacking in CESSDA organisations, compared to the data production outside CESSDA, are in the subjects: Economics, Trade, Industry & Markets, Education, Housing & Land Use Planning, Natural Environment.

In the countries studied a tendency is observed towards focusing production on a core topic, which is different in each country; this could be an indication of a particular research tradition in terms of favourite subjects in the country, reflecting strengths and weaknesses in the research area.

Figure 14 [non-CESSDA sample)



Data deposit/collection practices and regulations [non-CESSDA organisations only]

No CESSDA archives require exclusive deposit of data. This practice is to be considered positive, in terms of attracting data. Information on data deposit policy of the archives in the non-CESSDA sample could not be obtained.

Non-exclusiveness practically means that the same datasets might theoretically exist in different places across the data landscape. How then do CESSDA organisations attract data in competition with other organisations? Study on the data collection strategies in CESSDA revealed that most Archives engage in personal contacts & networking with producers. Bilateral co-operations and linking to projects producing data are also employed at large. The vast majority of archive-producer co-operations refer to liaising with the 'Public Sector, Academic and/or Research Centres' –their relationship is characterized as 'good' and 'excellent'. Relationships with the private sector are rated neutrally. Long-term and/or permanent licences of CESSDA Archives with producers are the major trend and they correlate with good-to-excellent relationships.

It is clearly evident from the above that good knowledge of the producers existing in each country is necessary for the facilitation of these activities. In relation to this issue a word of caution is needed. Research on the producers in countries should ideally be constantly performed by representatives of the Archives in each country. That way, a more complete landscape would be drawn. Indeed, in cases where our key informants engaged seriously in editing and completing our web research, the corresponding country landscape became clearer and broader.

Overview box # 4

CESSDA Archives do not require exclusive deposit. Collection policies rely on networking and personal contact with producers. The majority of co-operations refer to liaising with the public sector.

The Organisational Structure of CESSDA Archives: Central v/s Network

The division of labour which characterises production of social data, is enforcing its network characteristics. Basically, contemporary production of social data is based on the co-existence of two sides which constitute the complete research network: the side of a network of basic producers and providers of data and the side of a network of analysts and users of data. The second network uses and reuses data for testing of scientific and/or operational hypotheses and provides feedback to the first group.

Research infrastructures are those which can connect these two types of networks. The existence of a network of analysts and users presupposes the existence of research infrastructure which ensures access and distribution of the data; while the development of the infrastructure is based on the network of producers-providers, who ensure the continuous flow and accumulation of data, as part of the infrastructure.

For the purposes of this report we attempted to create an inventory of producers/ providers of data in selected countries. This work touches upon the first side of the research network; for a complete view of the pragmatics and dynamics of this network in the future, the second side of analysts and users must be an area to be investigated.

The importance of the development of networks within research infrastructures has proven to be critical mainly for two reasons:

First, contemporary research presupposes large scale infrastructures which cannot be sustained by a single research organisation due to the considerable developmental costs and the significant costs for maintenance, as well as difficulties in financing activities. Although the developmental cost can be covered from research projects, this is particularly difficult for the maintenance cost (SSH futures);

Second, research in the social science is characterised by a basic idiomorphy: the production and analysis of data has traditionally been organised in a decentralized manner by independent research projects; whereas, the development of research infrastructures leads to a network type of research around a core where research material exists and is accessed by all (Kallas 2002). CESSDA, operating successfully as a network so far can expand networking towards concrete collaboration with independent producers and analysts, relying mainly on current or future national networks.

Based on the above, we attempted to investigate how CESSDA organisations operate with respect to their organisational structure in terms of networking activities, i.e.: whether they operate as network organisations or as centralized organisations, and which is the commonest type.

From information extracted by navigation through the websites during Web survey A, 10 CESSDA members out of 20 operate in a network structure. We can distinguish between more institutionalized types of networking i.e.: UKDA and RQ (France) and less formal types of cooperation i.e.: FORS-Switzerland. Archives of the networking type are: CEPS/INSTEAD (Luxembourg), DANS (the Netherlands), DDA (Denmark), FORS (Switzerland), GESIS (Germany), NSD (Norway), RQ (France), RODA (Romania), SSD (Sweden), UKDA (UK).

Centralized organisations operate without, at least clearly, formally defined cooperation schemas at country level. CESSDA members of central character are 10 out of 20; they all are departments of academic institutions or research centres, i.e.: ARCES (Spain), ADP (Slovenia), ADPSS (Italy), ESSDA (Estonia), FSD (Finland), GSDB-EKKE

(Greece), ISSDA (Ireland), SDA (Czech Republic), TARKI (Hungary), WISDOM (Austria).

When CESSDA Archives are grouped geographically[16] we observe that there is a tendency towards a networking character in organisations located on the North-West of Europe whereas in the South the character is more centralised. As it was shown in Table 1[17] in terms of data collection policies organisations in the Southern part of Europe report that they facilitate networking for the acquisition of data, a strategy which seems to be an efficient model as applied in the other parts of Europe.

Archives of network character might have a wider range of subjects in their data collections. To test this assumption CESSDA Archives were divided into two groups, i.e.: central-network according to which their collections were studied. The subjects were also grouped in five main categories as follows:

1. Economy, Trade, Industry and Markets

2. Social Stratification and Groupings

3. Housing and Land Planning

4. ICTSc -Information & Communication, Science & Technology

5. Health, Social Welfare Policy & Systems

It appears that there is no difference in the subjects of their collections between the two groups. Historical reasons and institutional tradition may well be the main determinants. The most common subjects in data collections in organisations are "Social Stratification and Groupings", followed by data on Economics and Politics. Datasets on health and ICTS are rarer on both groups of organisations.

Overview box # 5

The organisational structure of CESSDA Archives in terms of central vs networking establishment is equally represented between the two forms. The range of subjects in the collections possessed are not differentiated by the organisational structure, although networking as an activity to attract data is the most preferred method, applied by organisations of both forms.

[16] See: Section 2
[17] Table 1

3. Section 3

The new landscape [18]

This section provides an overview of recent and ongoing changes in the data landscape that are likely to affect the future activity of CESSDA organisations. Some of these are changes in the requirements imposed by depositors, owing among other things to growing concerns about protection of respondents' personal data; others are due to changes in the demand of users who increasingly need international datasets for comparative cross-country research, and who now avail themselves of a wider range of data, including subjects such as health and the environment which used to be outside the scope of social science in the past. Another transformation results from technological developments, most prominently the availability of web-based tools to manage and disseminate data, and to an evolution of the data culture and professional standards of many disciplines, with a raising demand by funding agencies and by scientific journals to make available the data used in publicly-funded research projects and/or in published articles.

Some of these changes are already visible and reflected in the landscape surrounding CESSDA as described in the previous sections. Especially the role of some of the parallel or concurrent actors may be referred to these evolutions. Some CESSDA organisations are already trying to cope with these new needs but further, more extensive changes can be expected in future and may well modify the CESSDA environment in a more radical way.

It is impossible to fully map the extent of these changes, many of which are ongoing and may give rise to further (and possibly, deeper) transformations in the years to come. For this reason, this section does not attempt an exhaustive description of actual and anticipated novelties throughout Europe, but adopts a case study approach. While focusing only on a limited number of examples, it engages in an in-depth analysis of their various aspects in order to bring to light potential constraints and opportunities that may arise if the observed changes spread to a larger number of countries organisations. By so doing, this section aims to provide preliminary insight into some of the challenges that the future Infrastructure may face, and to suggest follow-up activities for the next few years. In particular, it is argued that despite the fact that the emerging needs for broader coverage are difficult to address in the short run, CESSDA should consider creating a task force in charge of surveillance of the evolution of the European data landscape in the near future. This solution will allow sufficient time to assess the nature, the extent and the repercussions of changes, and to explore the different opportunities available to CESSDA in order to set up a comprehensive, long-term strategy on how to satisfy a more diversified and more exigent demand of data-related services.

The remainder of this section is organised as follows. Section 1 considers the impact of recent technological changes, with focus on the effects of increasing data availability through the

[18] This section has been written by Anne Cornilleau, Marie Cros, Roxane Silberman, and Paola Tubaro.

Internet. Section 2 focuses on growing confidentiality concerns and their potential impact on CESSDA members. Section 3 examines the availability of international and comparative datasets across and outside CESSDA and considers possible scenarios for the future. Finally, section 4 explores some "non-conventional" types of data that were not traditionally part of CESSDA Data Archives' collections, but are now increasingly requested by social science researchers. These include subjects such as health, the environment, geo-referenced data, qualitative data, historical data, and psychology data. At present, some CESSDA Data Archives distribute these data nationally at least in part, while other countries have developed parallel systems of distribution with little or no participation of CESSDA members.

3.1 The society of information and the impact of technological change: direct dissemination, internet actors.

In recent years, the development of web-based tools for data management and data dissemination has substantially increased the availability of highly anonymised individual data, aggregate data, and tabulations. Parallel to these technological changes, access to information has become a major issue in the so-called information society with high pressure on public actors to make publicly funded data easily available for social partners, economic actors and citizens.

A first consequence of this evolution is that data producers have set up systems of data dissemination from their own web sites. In particular, National Statistical Institutes have largely increased their capacity to provide tabulations through the Internet: for instance in the Netherlands, public tabulations are provided through the web-based application "Statline" (http://statline.cbs.nl/statweb/) developed by CBS, the Central Bureau of Statistics. Similarly, Statistics Norway has developed "StatBank Norway", a system which allows users to select scope and content of each table, and then to export results with different format options (http://statbank.ssb.no/statistikkbanken/default_fr.asp?PLanguage=1); a similar tool exists in Denmark (http://www.statbank.dk/statbank5a/default.asp?w=1024). Statistics Austria is currently testing its new web-based tabulation system "SuperSTAR" (http://www.statistik.at/web_en/publications_services/superstar_database/index.html. Statistical departments of ministries and other government agencies are also increasingly pushed by governments to provide more data from their websites with similar tools, as can be seen for instance in France. International organisations also make increasing use of the web to disseminate their own tabulated data, e.g. the Geneva-based World Health Organisation, WHO (http://www.who.int/whosis/en/index.html).

Tabulations are of particular interest as producers can more easily meet specific needs of users without giving access to individual data, thus removing an important barrier. Nevertheless, one must note that together with tabulations, the Internet has also brought about rapid changes in the distribution of highly anonymised microdata, under the denomination of Public Use Files (PUFs). Once difficult to be obtained, PUFs can now be downloaded directly from some

producers' websites, e.g. in the case of INSEE, the National Statistical Institute of France (http://www.insee.fr/fr/bases-de-donnees/).The tendency to make anonymised data downloadable through the web is now flourishing and might spread further in future, although it is not yet universal (for instance in Germany, only PUFs designed for teaching purposes are freely downloadable, while other PUFs are still sent to users via postal services on CD support, see http://www.forschungsdatenzentrum.de/campus-file.asp). Thus, the debate on privacy protection across the European Union, the blossoming of laws protecting confidentiality that have impacted the degree of anonymisation even for researchers' access, has been paralleled by increasingly direct access to highly anonymised datasets with less intermediaries. The problem is that highly anonymised data are often sufficient for students and the general public, but do not always meet the expectations of researchers for whom they may not be sufficiently informative. In this sense CESSDA may still play an important role, focusing on an effort to provide data tailored to the need of researchers for which it has long experience and expertise; the traditional role of data archives would then be complementary with respect to web-based dissemination directly by producers.

Regarding datasets produced by researchers, a significant change is the data availability policy that some top-level scientific journals have recently implemented. From 2005 onwards, the *American Economic Review*, a leading economics journal, has adopted a strict policy requiring authors to provide, prior to publication, the data, programs, and other details of the computations sufficient to permit replication. The basic purpose of this policy is to ensure high-level scientific standards. Datasets are posted on the ***Review***'s Web site (http://www.aeaweb.org/aer/contents/index.php). Other journals have followed suit, for instance the *Journal of Political Economy*, the *Canadian Journal of Economics*, and in Europe, the *Journal of the European Economic Association* (http://www.eeassoc.org/index.php?site=JEEA&page=41). This tendency is not yet widespread across all disciplines but may lead to a situation in which each scientific journal has a web page containing datasets or links to datasets available from individual researcher's web pages. Journal may thus play an increasing role in data dissemination.

In addition, public-sector funding agencies increasingly require researchers to make available the data collected and used in the research projects for which they award grants. Commitment to the principle that the various forms of research data collected with public funds belong in the public domain is justified on the basis of various arguments, including not only replicability, but also an effort to provide resources for research and educational purposes, and more generally, a requirement of openness and accountability. This is an established policy of, for instance, the Social Sciences and Humanities Research Council of Canada (www.sshrc-crsh.gc.ca) and in Europe, of the Economic and Social Research Council of the UK (http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/index_academic.aspx); it is a recent evolution in France at the ANR (National Agency for Research) that now explicitly includes these requirements. Although the adoption of such policies is not yet universally established, other countries may follow suit in future.

These developments sometimes reinforce the role of CESSDA Data Archives, to the extent that they can be designated as repositories for data from publicly-funded research projects, as is currently the case in the UK. In other cases, however, these policies encourage new actors to emerge in the social science data landscape, thanks to the possibilities offered by technological changes. For instance, the Institute for Quantitative Social Science (IQSS) of Harvard University has developed a tool for data deposit and linkage with related publications, the Dataverse network (http://thedata.org/). This platform allows any researcher or institution to create their own page with the data used in publications. The documentation is based on a few DDI fields and long-term preservation of data is guaranteed by IQSS. One of the main ideas underlying this project is to enhance the visibility of researchers who work with data and to respond to the surging demand of journals to deposit data used in articles.

It also needs to be emphasized that recent tools based on Web 2.0 principles, blur traditional roles and functions. One could previously distinguish between passive "consumers" of available data and "producers" and/or providers; today, instead, the content is not merely consumed by users but can be, to some extent, produced by users. These developments are welcome as they encourage a culture of transparency and data sharing, with participation of all actors to the whole data production, archiving and dissemination process; yet they may challenge the traditional role of CESSDA Data Archives as central providers of data documentation, preservation, and dissemination services. They may also, somehow paradoxically, lead to a higher degree of opacity in the system as a multiplicity of sources and actors may make it difficult for data producers and users to locate the services they need.

Hence, it is important for CESSDA to reflect on how to upgrade its role to provide meaningful and up-to-date information and mediation of data services in a landscape in which numerous and diverse actors operate simultaneously, and new tools allow traditional users to have a more active role in data production and dissemination.

3.2 Impact of growing confidentiality concerns

Since the 1990s, growing concerns about privacy protection have intensified problems of access to detailed microdata. A wider range of issues are regarded as sensitive and several types of data that could be in the public domain in the past are now subject to increased protection.

In particular regarding government microdata, National Statistical Institutes (NSIs) are now often reluctant to authorise access through third parties, including CESSDA Data Archives. New computing technologies offer NSIs an opportunity to provide data on their own under secure conditions. Some of them have created on-site safe centres to allow researchers to access very detailed microdata on their premises. Usually in a dedicated room, researchers can use a special IT environment with no downloading or e-mail facilities; intermediate printing is often allowed but only final outputs can be taken out once results have been checked for confidentiality. Eurostat and the National Statistical Institutes of Finland, France, Germany, Hungary, Ireland, Italy, the Netherlands, Slovenia, Spain, and the UK currently have this mode of access. Other public administrations have safe data centres as well: for instance in Germany,

BA-IAB (the research and statistics service of the Federal Employment Agency). Another option (often co-existing with the former) consists in creating secure remote connection facilities based on special IT environments with enhanced firewall protection; no availability of printing, downloading or e-mail facilities; control of inputs; and output confidentiality checks; the advantage with respect to on-site arrangements is that researchers can use the data from their own institutions. Such systems are available at the NSIs of Austria, Denmark, Finland, France (as a pilot at the moment), Germany, Ireland, the Netherlands, Slovenia, Sweden, and the UK. Other producers of governmental statistics have such facilities as well: it is the case, for instance, of the Bank of Italy, BA-IAB in Germany, and LIS (Luxembourg Income Study).

On the whole, it can be said that secure modes of access to detailed datasets (both on-site and remote) have spread quickly and may develop further as an increasing number of institutions are planning to implement them in the near future, for instance Statistics Austria and the Italian ISTAT. Government data producers might in future even offer coordinated access through a European network; this is not only an obvious problem for CESSDA which might be relegated to a secondary role in the dissemination of governmental microdata but also for data users, who may have to face a more fragmented landscape with subsequent opacity of the system and potential gaps in data documentation, information, and other services.

In this perspective, CESSDA data archives must rethink their policies with respect to their role as intermediaries for government microdata. The renewed forms of cooperation that have been recently set up between National Statistical Institutes and some Data Archives –notably NSD in Norway, Réseau Quetelet in France and UKDA in the UK– for enhanced access to detailed data for purposes of scientific research are a promising example. In future, these archives might provide support to others in their efforts to prepare agreements with official statistics providers. (For more details see the WP10-T1 and T2 report).

The difficulties for CESSDA Data Archives are not limited to possible competition from National Statistical Institutes and other public-sector agencies. Another problem is that a wider range of academic and private-sector data producers require stronger guarantees for privacy protection and are increasingly likely to set up their own remote access. In particular, it may be the case for longitudinal studies (see for instance the ELFE cohort in France that plans to set up its own platform for access).

While the tradition of most data archives was to only (or primarily) handle data that could be put into the public domain, they now have to undergo substantial changes and to upgrade their ability to manage confidential data. To some extent, this transformation is already under way. Outside Europe, ICPSR has substantially increased its capacity to offer restricted-access data under secure conditions. Among the datasets it is offering in this form are two major longitudinal studies, the Panel Study of Income Dynamics of the University of Michigan (from September 2009) and the National Longitudinal Study of Adolescent Health of the University of North Carolina at Chapel Hill (from January 2010). Within Europe, UKDA is implementing a secure data service system due to start in October 2009 as a pilot project, which is expected to

offer not only data from ONS (the Office for National Statistics) but also longitudinal studies produced by academic research.

Again, the CESSDA members that already have acquired expertise in handling confidential data might consider offering coordinated support services to other archives that currently lack these competencies but might need to develop them in the near future.

3.3 International and comparative data

Another change in the data landscape is the increase in the production of international comparative surveys and furthermore, the increase of the number of users of these surveys. As an indicator, one can look at European Social Survey (ESS) users were more than 20 000 in 2009, up from around 9 000 in 2006. Another indicator is the number of publications based on these data: for instance, the International Social Survey Programme (ISSP) website gives a trend of publications with ISSP data, showing an increase of their use since the beginning of the programme (http://www.issp.org/trends2.shtml).

Thus, it is important for CESSDA to position itself with respect to these surveys. Some of them are already managed and/or distributed by CESSDA as a whole or by some of its members, as described in sub-section 3.3.a and in Tables 3 and 4; a substantial heterogeneity across member archives emerges, which will need to be addressed by the future Infrastructure. Other surveys are produced and distributed by non-CESSDA actors as outlined in section 3.3.b, and the problem for the future Infrastructure will be to identify forms of collaboration and partnership with these other actors so as to reinforce its role as an intermediary, albeit indirectly. Finally, section 3.3.c focuses on the closely related question of the CESSDA Trans-Border Agreement, which is meant to facilitate comparative research through the use of national data from different countries, by allowing researchers accredited by one CESSDA member to access the data collections of another member. In practice, most member archives have limited experience of the Agreement and little practical knowledge of how to implement it; hence, the future CESSDA will need to find ways to ensure more extensive and systematic application of it, and to provide more transparent and regular flows of information on the number and type of cross-country data users.

3.3.a CESSDA's role in the dissemination of international comparative surveys

Within CESSDA, access to major comparative surveys seems to be heterogeneous as it appears in Tables 3 and 4 at the end of this section.

International surveys as ISSP, Eurobarometers or European Value Survey (EVS) are archived, documented and disseminated by GESIS. Dissemination has been facilitated since the launch of ZACAT, the NESSTAR server of GESIS but heterogeneity in access for users persists.

Indeed, these surveys are available in different modes from several CESSDA members. Some countries disseminate the national data of the survey. For instance, the French ISSP data are disseminated by Réseau Quetelet to all users as is the case in Sweden, where they are made available from the Swedish National Data Service.

Another case is the dissemination of complete datasets of these surveys but only for national users: as an example, ESDS International (UKDA service) gives access to ISSP and Eurobarometer datasets only for "UK residents or those registered at UK institutions of higher or further education only". On the study description of these surveys, it is mentioned that they could be accessed from ZACAT. Similarly ISSDA, the Irish CESSDA member, disseminates several waves of ISSP and Eurobarometers and points to access through ZACAT for the most recent waves.

In France, the Eurobarometers were the object of an agreement with GESIS to disseminate to French users the datasets that are not on ZACAT and are no longer under embargo.

Differences in access occur from one survey to another and also from one country to another. Several issues could be pointed out:

- Ideally, heterogeneity should disappear with the construction of CESSDA as a European Infrastructure, where as much homogeneity as possible should be ensured.
- The current situation could be confusing for the scientific community, since researchers might not know exactly where to search and what they can obtain from the different modes of access.
- Different forms of access lead to disparity for users from different countries, which is paradoxical with international comparisons.
- CESSDA members might encounter problems in determining how many researchers or students of their own country are using these data.

Some international surveys are also archived and disseminated by CESSDA members, but on behalf of a project, as it is the case for ESS (NSD archive) or the CSES (Comparative Study of Electoral Systems, GESIS archive). If the situation is not really confusing for users, since only one gateway exists to access these data (even if some national data are available in national archives), the infrastructure CESSDA should closely follow those projects, mainly in order to provide clear information about these surveys.

3.3.b International comparative surveys outside of CESSDA

Some international surveys have developed outside CESSDA. The Luxembourg-based LIS (Luxembourg Income Study) harmonises and provides access to income surveys, including a number of European surveys, through a remote execution system. MTUS (Multinational Time Use Study) is another well-known example of an international platform, based at a European

Union University, harmonising and providing access to a comparative database. We must of course also mention IPUMS (Integrated Public-Use Microdata Series), based in the USA, which regularly collects, harmonises and distributes a great number of censuses across the world, and plans to develop a European platform focussing more on European census comparisons. More recently SHARE (Survey of Health, Ageing and Retirement in Europe), which is part of the ESFRI process towards European Research Infrastructures, has developed its own platform to provide access for researchers to data collected in different countries. Whereas LIS, MTUS and IPUMS mainly collect datasets from NSIs (see the WP10-T1 and T2 report), SHARE is an academic project. Regarding education, PISA (Programme for International Student Assessment) is another example of comparative survey, developed in the environment of International institutions, namely OECD, which gives easy and free access to its microdata. How CESSDA will develop links with these European and international platforms is an important issue for the future ERIC, as comparative platforms outside CESSDA may well become more numerous in the near future, and develop their own standards.

3.3.c Circulation of data among CESSDA members: the Transborder Agreement

The rise of comparative research is not only linked with international survey programs but also with the use of national surveys in order to carry out comparison.

The CESSDA Transborder Agreement (TBAA) should facilitate access for users to data from foreign CESSDA members.

Initially (in the non-digital age) the arrangement for data circulation was more or less as follows: if a user from country X needed a dataset from country Y, the data archive in X would ask its partner in Y to provide a copy; the X data archive would generally keep the copy for further use, i.e. in case another user would require the same dataset in future.

However, this made little sense in the digital age: the idea then was to allow for circulation of users from one data archive to another, with no (or little) circulation of data files. The CESSDA TBAA was designed to achieve precisely this. The principle is that the CESSDA member in country X accredits a user from its home country who is then allowed to request access to foreign datasets from the partner archive in country Y. Under the assumption that each CESSDA member is likely to have better knowledge of the higher education and research institutions of its home country than of any other countries, the TBAA allows for this knowledge to be shared within CESSDA.

This should happen without any additional expenses, so that users from any CESSDA country would benefit from the data at the same fees as national users (or at no fees if native users are not charged). This agreement applies only to data that are not subject to any restrictions (i.e. aggregate data or anonymised individual data, typically from large international surveys).

Technically, the system should work smoothly under software systems supported by, for example, NESSTAR and an authentication system (like Athens or Shibboleth). As of today, not all CESSDA organisations are using NESSTAR but more may do so in future.

To what extent is the TBAA really applied? The number of users who avail themselves of TBAA provisions might prove to be rather low. One reason for this is that some of the data sets are provided by ICPSR with a slightly more straightforward procedure, so that European researchers from ICPSR member institutions may prefer to direct their requests to ICPSR rather than to CESSDA members. There is no estimate of how many do so at the moment.

From the users point of view, a brief examination of CESSDA members' websites shows that information is not always available on how to get data with TBAA for foreign users, and if the information exists, the procedure is not clearly indicated.

The very existence of the agreement makes it difficult to monitor how many users from one European country access data directly from archives in other member countries. (For instance, someone from country X who was accredited once to have data from archive Y within CESSDA may go directly to Y for a second dataset that he/she needs, without mediation from archive X.) To prevent such problems, the TBAA requires member archives to provide a yearly report on the number of non-native users who access their data. But it is unclear whether or not CESSDA members do prepare such reports annually, and if they do, with what degree of detail.

Issues raised about the international surveys case are quite similar in the case of data circulation within CESSDA.

- TBAA should be more systematically applied than it is now the case.
- Information for users is not sufficiently clear and should be improved in the new infrastructure.
- Again, the issue of user statistics is present. Transparency on user data within CESSDA has to be taken into account in the construction of a European infrastructure.

Table 3: ESS, ISSP, EVS, Eurobarometres

	ESS	ISSP	EVS	Eurobarometres
Main Dissemination	data available through the project website essdata. Freely available after registration on the website.	Via ZACAT	Via ZACAT	Via ZACAT
GESIS	NO DISSEMINATION	Official archive	Official archive	Official archive
ISSDA	NO DISSEMINATION	Dissemination restricted to user in Ireland and for social science research only	Dissemination restricted to users in Ireland and for social science research only	Dissemination restricted to users in Ireland and for social science research only
RQ	NO DISSEMINATION	Dissemination of the French survey	NO DISSEMINATION	Dissemination restricted to users in France and for social science research only
UKDA (ESDS International)	Researchers directed to ESS site	Dissemination restricted to user in UK (and Athens institutions) and for social science research only	Dissemination restricted to users in UK (and Athens institutions) and for social science research only	Dissemination restricted to users in UK (and Athens institutions) and for social science research only
NSD	Official archive but dissemination on behalf of the project.	NO DISSEMINATION	NO DISSEMINATION	Dissemination restricted to users in Norway and for social science research only

FORS	Dissemination of the Swiss datasets	Dissemination of Swiss survey	NO DISSEMINATION ?	Dissemination the Swiss survey (parallel survey)
DANS	NO DISSEMINATION	Dissemination of the Dutch survey (when carrying out with another Dutch survey?)	Dissemination of integrated datasets (1999-2002 EVS-WVS and 1999 EVS)	NO DISSEMINATION
SND	NO DISSEMINATION	Dissemination of the Swedish data	NO DISSEMINATION	Few datasets available
RODA	?	NOT IN THE SURVEY PROGRAMME	?	?
GDSB/EKKE	Dissemination of the Greek dataset (2002 survey only)	NOT IN THE SURVEY PROGRAMME	NO DISSEMINATION	Few datasets available
FSD	Dissemination of Finnish datasets	Dissemination of Finnish datasets	NO DISSEMINATION	Dissemination of integrated datasets. Finnish researcher only?
ESSDA	?	NOT IN THE SURVEY PROGRAMME (new member for last wave)	?	?
CIS	NO DISSEMINATION	Dissemination of Spanish datasets	NO DISSEMINATION	NO DISSEMINATION
DDA	Dissemination of Danish datasets	Dissemination of Danish dataset	Dissemination of Danish datasets	?

ADP	Dissemination of Slovenian datasets	Dissemination of Slovenian datasets	Dissemination of Slovenian datasets	Few datasets available
SDA	NO DISSEMINATION	Dissemination of Czech dataset	NO DISSEMINATION	NO DISSEMINATION
WISDOM	NO DISSEMINATION	Dissemination of all datasets (only Austrian researcher?)	NO DISSEMINATION	Dissemination of all dataset (only Austrian researcher?)
TARKI	Dissemination of Hungarian datasets	Dissemination of Hungarian datasets and some international datasets	Dissemination of Hungarian datasets and some internationa datasets (WVS integrated datasets)	NO DISSEMINATION
APDSS	Dissemination of integrated datasets	Dissemination for Italian researchers	Dissemination for Italian researchers	Dissemination for Italian researchers
CEPS	?	NOT IN THE SURVEY PROGRAMME	?	?

Table 4: SHARE, PISA, CSES

	SHARE	PISA	CSES
Main Dissemination	Via centerdata (NL) after engagement form	Freely available on pisa website (oecd website)	ICPSR and GESIS for Europe
GESIS	NO DISSEMINATION	NO DISSEMINATION	Archiving and dissemination for European researchers on behalf of the project.
ISSDA	NO DISSEMINATION	NO DISSEMINATION	NO DISSEMINATION
RQ	NO DISSEMINATION	NO DISSEMINATION	Dissemination of French datasets
UKDA (ESDS International)	NOT IN THE SURVEY PROGRAMME	NO DISSEMINATION	NO DISSEMINATION
NSD	NOT IN THE SURVEY PROGRAMME	NO DISSEMINATION	NO DISSEMINATION
FORS	NO DISSEMINATION	Dissemination of the Swiss datasets	NO DISSEMINATION
DANS	NO DISSEMINATION	Dissemination of the Dutch datasets	NO DISSEMINATION

SND	NO DISSEMINATION	NO DISSEMINATION	NO DISSEMINATION
RODA	NOT IN THE SURVEY PROGRAMME	?	?
GDSB/EKKE	NO DISSEMINATION	NO DISSEMINATION	NOT IN THE SURVEY PROGRAMME
FSD	NOT IN THE SURVEY PROGRAMME	NO DISSEMINATION	Dissemination of Finnish datasets
ESSDA	NOT IN THE SURVEY PROGRAMME	?	NOT IN THE SURVEY PROGRAMME
CIS	NO DISSEMINATION	NO DISSEMINATION	NO DISSEMINATION
DDA	NO DISSEMINATION	?	Dissemination of Danish datasets
ADP	NO DISSEMINATION	NO DISSEMINATION	NO DISSEMINATION
SDA	NO DISSEMINATION	NO DISSEMINATION	NO DISSEMINATION
WISDOM	NO DISSEMINATION	NO DISSEMINATION	NOT IN THE SURVEY PROGRAMME

TARKI	NOT IN THE SURVEY PROGRAMME	NO DISSEMINATION	NO DISSEMINATION
APDSS	NO DISSEMINATION	Dissemination of integrated datasets	NO DISSEMINATION
CEPS	NOT IN THE SURVEY PROGRAMME	?	NOT IN THE SURVEY PROGRAMME

3.4 Non-conventional data (environment, health, history, geo-tagged data, qualitative, psychology)

This section deals with data traditionally belonging to disciplines that are outside the boundaries of the social sciences, such as health and the natural environment. Of course the social sciences have long been interested in say, perceptions of health or environmental risk, attitudes towards environmental protection, etc., which is why a number of CESSDA data archives classify some of their data under these headings; yet little attention was given to epidemiology, medical and clinical data, as well as data on air pollution, deforestation, waste production, etc. These data existed outside of the social science network of data production and distribution. Today, however, social scientists have developed an interest in these data and often undertake research that needs access to them together with more traditional sources. Social science research on medicine, health economics, and the design and evaluation of environmental policies are just some examples of areas in which social scientists increasingly need data from the medical and the natural sciences. Conversely, epidemiologists are more and more interested in having socio-economic variables and combining different datasets. This is why access to these data is becoming a crucial issue for CESSDA. While a solution is unlikely to be found in the short run, surveillance of new developments in these areas and an effort to provide information on availability and accessibility of health and environmental data to interested social scientists can be taken as more immediate goals.

History is a peripheral discipline for CESSDA to the extent that it belongs to the humanities rather than the social sciences which constitute member archives' main business. However, there are intersections between historical and social science research that require scholars to access data from both recent and past sources. What's more, some CESSDA archives have historical data in their holdings and have experience and expertise in this area. Thus, it may be useful for the new Infrastructure to consider possible forms of cooperation and inter-operation with historical research institutions and historical archives.

Qualitative data are not the main focus of CESSDA data archives but they are of interest to many social science researchers in European countries. While there is a limited tradition of archiving and sharing qualitative data, some initiatives are currently being undertaken (notably in Germany and the UK), and the future CESSDA may consider joining new developments in this area.

Geo-tagged data are of recent creation and still unsystematically available throughout Europe. However, production of these data is quantitatively increasing and qualitatively improving at an astonishingly high speed, which suggests that this area must be monitored closely in the near future in order for CESSDA to be part of the process.

Finally, other data that are little found in CESSDA data archives so far are psychology and still other data (genealogical, names, etc.).

What follows is a more detailed presentation of what CESSDA data archives offer in these areas, together with an outline of the outside landscape (based on a few examples and case studies), bringing to light the presence of a multiplicity of parallel operators and in some cases, of parallel systems.

3.4.a Health

Health is an area where data have existed for a long time, but they were previously used only in medicine. Researchers in the humanities and social sciences increasingly need to work on these data. Health policy evaluations are a major issue across the European Union. Socio-economic characteristics are necessary to understand differences in health and ageing. Thus there are more and more datasets combining epidemiological and socio-economic datasets. Long term cohorts now increasingly mix these kinds of data, used by both research communities. It is therefore important to understand the current dissemination of these data. What role can CESSDA data archives take in it?

First, most data archives indicated in the CESSDA survey that they had "health data" in their collection[19]. However, it appears most of the time that the data disseminated are data about health insurance, policy, education, care services, behaviours, while few epidemiological data are available. It reveals that there is no clear, commonly shared understanding of what "health data" are in the social science and archiving community.

Several forms of dissemination exist depending on the country. Here are a few examples:

- Parallel systems of distribution with no (or little) mediation from CESSDA members.

Germany
In Germany, GESIS does not distribute any data on health. The Federal Health Monitoring is a central actor. It offers more than 1 billion numbers displayed in tables free of charge. The data are derived from more than 100 different sources, among which there are many statistics from the Statistical Offices of the Länder and the Federation. There are also data from numerous other institutions from the health sector. Besides data from and about Germany the information system also offers international tables of the OECD and the WHO.

(http://www.gbe-bund.de/gbe10/pkg_isgbe5.prc_isgbe?p_uid=gastd&p_sprache=E)

[19] UKDA, FSD, NSD, FORS, Réseau Quetelet, SDA, CIS, DDA, Tarki, DANS, WISDOM (Sources: CESSDA-Survey and CESSDA members' websites).

France
In France, IRDES (Institute for research and information in health economics) disseminates health data bases (on status of the population, expenditures, professions, hospital, care insurance), while DREES (Department for research, studies, evaluation and statistics of the Ministry in charge of Health) provides epidemiological data in addition to numerous specific surveys or administrative data on health behaviours. These data are mainly available for research through research projects. Some surveys from IRDES are now available through Réseau Quetelet and negotiations are going on with DREES.

Two other actors play a major role in access to health data. IRESP has been raised in conjunction with INSERM, the major actor in Health research, to organize access to health data, especially long term cohorts. IRESP particularly aims at raising a technological platform facilitating access to these data as well as to the administrative basis of SNIRAM (national register for individual health data) whereas IDS (Institut de la Santé) which gathers public health actors, researchers and the health industry, is in charge of regulating access to this major database.

- Both CESSDA Archives and national institutes provide health data.

UK
In the UK, UKDA provides access to health data, but the National Health Service (NHS) is England's central, authoritative source of health and social care information. A particular service, the Secondary Uses Services (SUS) is designed to provide anonymous patient-based data for purposes other than direct clinical care (http://www.ic.nhs.uk/services/the-secondary-uses-service-sus).

Norway
Researchers in Norway have access to health data from many governmental registers and other sources. One of the most central in this respect is the Norwegian Institute of Public Health (NIPH - http://www.norgeshelsa.no/norgeshelsaen/). This is a national centre of excellence in the areas of epidemiology, mental health, and control of infectious diseases, environmental medicine, forensic toxicology and drug abuse. Another example of a central source is the Cancer Registry of Norway (http://www.kreftregisteret.no/en/). It holds data collected from clinicians and pathologists, as well as from administrative discharge and mortality sources. It is updated continuously with information on new cases as well as on cases diagnosed in previous years. Researchers can have access to these data.

NSD is also an important provider of health data to the research community, at regional level, and as surveys, register data, etc. at individual level. These data are collected from various sources such as Statistics Norway, governmental agencies, academic surveys, and research projects. NSD's agreement with Statistics Norway on the dissemination of data for research purposes is important in this respect as it gives NSD the possibility to offer a series of surveys on health issues for research purposes. This is also the case when it comes to regional data.

The Directorate for Health and Social Affairs has assigned NSD the task of running a data service for research on the system of regular GPs. The data are provided by the Norwegian Labour and Welfare Organisation. NSD is also on commissioned by the

same directorate to offer a web-based statistics system containing information about the case-processing procedures of the Patient Ombudsmen.

Denmark
In Denmark, "DDA Health" has been implemented as an integrated part of the Danish Data Archive. Data Health disseminates original data collected by interviews or register as part of human-based scientific health research.

The National Board of Health is the supreme health care authority in Denmark. The Danish National Board of Health uses an active IT strategy to, among other things, give Danes access to statistical information on health and illness. The most recent access point is a Web site (www.sst.dk) with statistical information relating to such subjects as hospital treatments, incidence of cancer, number of births and causes of death.

The Danish National Board of Health develops and uses a wide range of registers within the health sector that are used for health monitoring and planning, as well as research and administration. On the basis of the extensive data contained in the registers, the Danish National Board of Health draws up the Danish health statistics that are now available online.

Based on the current shortcomings and the ongoing developments, recommendation for the future CESSDA could be to set up a clear classification and description of the different types of health data, then finding some way to provide comprehensive information to social scientists on where and how these data can be accessed.

3.4.b Environment

Many CESSDA archives have survey data, or interviews, on values, attitudes, behaviours towards the environment. Topics are e.g, eco-energy attitudes, environmental attitudes and behaviour, environmental concerns. Yet environmental data are more difficult to find. A mixed picture of environmental data dissemination across countries emerges.

- Environmental data in CESSDA archives.

Environmental data are not widespread, especially among CESSDA Archives. Only two CESSDA archives mentioned[20] these data in their collection, namely GESIS and UKDA. GESIS provides indicators such as damaged forest area, air pollution, or household waste produced; UKDA distributes among other things data of the International Energy Agency on energy production and consumption as well as on greenhouse gas emissions. Sub national researches (for example on woods, aircraft noise…) are also proposed for scientific use.

[20] CESSDA survey.

The UKDA's Rural Economy and Land Use (RELU) data support service provides researchers with information about access to data which are stored at and distributed from different sites but which are all collected under a research programme funded jointly by the Social and Economic, the Natural Environment, and the Biotechnical and Biological Research Councils.

- Parallel actors

France
In France, different organisations linked to ministries provide data on the environment with no intermediation of the CESSDA Archive. These institutes[21] give access to key-figures and data on subjects such as nature and biodiversity, water quality, risks, land use, etc. An increasingly relevant issue is how these data will be accessed for use as contextual variables in a number of social as well as health surveys. ELFE, which will follow a cohort from birth to adult age, combines classical socio-economic survey, epidemiologic datasets, and environment contextual data. It is an interesting example of the blurring frontiers between disciplines, raising new challenges for Archives.

Germany
In Germany the information system PANGAEA - Publishing Network for Geoscientific & Environmental Data- is operated as an Open Access library aimed at archiving, publishing and distributing geo-referenced data from earth system research (http://www.pangaea.de) Most of the data on water, sediment, ice, and atmosphere are freely available and downloadable.

Norway
The Norwegian Pollution Control Authority (SFT) is a government agency that manages and enforces the Pollution Control Act, the Product Control Act and the Greenhouse Gas Emission Trading Act. One of its main tasks is to monitor and inform about the state of the environment. SFT has established an online data portal that has information about all data that has been used in it, with hyperlinks to the portal "State of Environment Norway" (http://www.environment.no/), which contains data from other environmental agencies, too. The purpose is to improve access to environmental data for the government and other agencies, the research community and others. Data can be downloaded from the portal without any restrictions.

The Directorate for Nature Management is the national governmental body for preserving Norway's natural environment. The directorate serves as an advisory and executive agency under the Norwegian Ministry of Environment. The Directorate runs a data service that gives access to maps, databases and references. It provides a

[21] The French Environment and Energy Management Agency (ADEME), IFREMER (French Research Institute for Exploitation of the Sea), Coastal Observatory (Ministry of Ecology, Energy, Sustainable Development and Planning).

large amount of data about the Norwegian nature and landscape. Data are accessible online without any restrictions (http://english.dirnat.no/)

3.4.c Geo-tagged data

Geo-tagged data do not seem to be widespread across countries. They are not currently available from CESSDA organisations which may become a critical issue as researchers are increasingly interested in mapping socio-economic phenomena.

These data are generally available through other actors. As an example we can first focus on an initiative built to integrate data on geo-tagged systems, and then on the lGN French experience of cartography-related data. The INSPIRE initiative may also be an example to follow or to join.

UK
The Centre for Advanced Spatial Analysis (CASA) is an initiative of researchers at the University College London to develop emerging computer technologies in several disciplines which deal with geography, space, location, and the built environment. As an interdisciplinary research centre, expertise is drawn from archaeology, architecture, cartography, computer science, environmental science, geography, planning, remote sensing, geomatic engineering, and transport studies (www.casa.ucl.ac.uk)

France
In France the Geographical National Institute (www.ign.fr) is in charge of producing, maintaining and disseminating reference geographic information. IGN has been involved in all cartography-related operations in France and its territories, since 1940. Due to its legal status that combines public and private participations, access for researchers is still very costly and no general agreement has been secured till now. One consequence is for instance that Réseau Quetelet which provides access to National Statistical Institute (INSEE) surveys and censuses, cannot provide access to the related IGN units under a specific licence between INSEE and IGN.

The INSPIRE (Infrastructure for Spatial Information in the European Community, http://inspire.jrc.ec.europa.eu/index.cfm*) initiative*

INSPIRE endeavours to address some gaps in spatial information in Europe. The general situation is one of fragmentation of datasets and sources, limited availability, lack of harmonisation between datasets at different geographical scales, and duplication of information collection. These problems make it difficult to identify, access, and use data that are available.

INPSIRE aims to fill the need for quality geo-referenced data. The initiative intends to trigger the creation of a European spatial information infrastructure that delivers integrated spatial information services. These services should allow users to identify and

access spatial or geographical information from a wide range of sources, from the local to the global levels, in an inter-operable way for a variety of uses. The target users of INSPIRE include policy-makers, planners and managers at European, national and local levels, and the citizens and their organisations. Possible services are the visualisation of information layers, overlay of information from different sources, spatial and temporal analysis.

The Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) was published in the Official Journal on the 25th April 2007. It entered into force on the 15th May 2007. This was seen as a milestone for the use of Geographical Information in Europe as a contribution to environmental policy and sustainable development. It was the first step in a co-decision procedure that should lead to the formal adoption of the INSPIRE Directive, which then will have to be implemented in every EU Member State. At the moment, it remains a process outside CESSDA though researchers will need to have bridges between spatial and statistical information.

3.4.d Historical data

According to the CESSDA-survey, several CESSDA archives provide historical data: GESIS, ISSDA, NSD, DDA, TARKI, UKDA, RQ, DANS.

GESIS gives for instance access to studies and data of historical social research, which are characterized by a high heterogeneity of the collected data. The historical sources of these studies can be both text documents (e.g. church registers, written documents of finance departments, court documents) and official statistics or data of institutions like e.g. the German 'Reichsbank'.

In the UK, UKDA gives access to a various range of historical studies such as administrative, agricultural, religious, education, legal and local, and to a rich collection of historical and contemporary census material. It must be noted that the HDS (History Data Service) is the successor of the AHDS History, a service of the Arts and Humanities Data service created in 1966. It is now housed in the UKDA.

DDA distributes historical data which are mostly linked to demographic history and deal with subjects such as immigration and minorities. In addition, the Danish State Archives Filming Centre digitises parish registers and population censuses in order to make them accessible via the internet (since 1787).

In Norway, historical data can be accessed from a wide range of local and national based sources. At national level The National Archival Services of Norway, The Norwegian Historical Data Centre (NHDC) and Statistics Norway should be seen as central.

As mentioned above, history is a peripheral discipline for CESSDA to the extent that it belongs to the humanities rather than the social sciences; in general, most historical data are accessed through National Archives, and in some countries in regional archives, under provision of the Archives laws. Nevertheless, historians constitute databases which are available in some cases through CESSDA members, in other cases through other bodies, or are kept by the researchers with uneven access. An example that demonstrates the need for a more proactive policy in this area is the TRA project, which collected French data over two centuries for families whose names begin with TRA. Today, National Archives are sometimes involved in programmes aiming to build vast data bases for historians. Finally in the contemporary period, regular surveys constituting historical series are available from CESSDA members' catalogue and/or NSIs' catalogues. Labour Force Surveys for instance are now available for more than 30 years, and can be of interest for historians. Where frontiers and intersections should be between social science and historical data, what forms of cooperation should be set up between CESSDA and the major actors in this field, notably the National Archives, are issues that CESSDA should consider in the near future. At least in countries where CESSDA members are considered as part of the National Archives, as is the case for UKDA, collaboration should be relatively smooth. In any case, cooperation with National Archives is needed in this field.

3.4.e Qualitative data

Regarding qualitative data, several CESSDA Archives provide data, for example UKDA, FSD, DDA.

UK
ESDS Qualidata acquires data created during the course of qualitative research across a wide range of social science disciplines. Data supported include: in-depth and semi-structured interviews; focus groups; field notes and observations; personal documents and photographs. The service provides access to some of the classic post-war studies of British society. (http://www.esds.ac.uk/qualidata). It must be noted that the recent evolution towards a network structure in the UK has led to include Qualidata in the same national service, ESDS, as UKDA. Though, currently, only UKDA is member of CESSDA.

Finland
FSD gives access to interviews, to open-ended questions of surveys and other materials (in Finnish).

Denmark
An example of Danish qualitative data held at DDA are "Danish craftsmen recollections, 1880-1914" (http://ddd.dda.dk/ddakatalog/sdfiler/R00676.htm) .

Germany

The Archive for Life Course Research (ALLF) in Germany (http://www.lebenslaufarchiv.uni-bremen.de) is not part of GESIS. It provides access for a wide range of documented and digitised social science data collections from empirical research projects for secondary use in research, learning and teaching. At present ALLF holds a collection of about 700 qualitative, documented and anonymised interviews, and thus, it is probably the largest archive for qualitative interview data from social science research in Germany. Most data collections derive from the Special Collaborative Centre 186 'Status Passages and Risks in the Life Course', a research programme with longitudinal projects on different transitions and status passages in the life course conducted between 1988 and 2001 which was funded by the German Research Foundation, DFG.

(http://www.qualitative-research.net/index.php/fqs/issue/view/13).

Discussion about the place of qualitative data in CESSDA will develop in the near future in a context where open archives policies are setting up local platforms in the Universities that could increasingly be a place for archiving qualitative data produced in the academic world. In France, the ARSH (Archives de la Recherche en Sciences Humaines) could aim at using the network of the MSH (Maisons des Sciences Humaines) located on various campuses to collect and make available qualitative data.

A detailed assessment of the availability and accessibility of qualitative data and of relationships between Qualitative data providers and CESSDA is provided by the report from the Bremen Workshop (April 2009) 'Qualitative and Quantitative Longitudinal Resources in Europe: Mapping the Field and Exploring Strategies for Development'. The workshop was supported by the PPP and the full report is available as D10.5b.

3.4.e Other non-conventional data

Some CESSDA archives have developed competencies in archiving and distributing very specific data. UKDA give access to psychological studies and FSD to personality tests for instance. Denmark has developed the Danish Demographic Database. It is based on censuses, data from the Danish Emigration Archives, police records, church books, departure lists and so on. It "allows you to search for your ancestors on the Internet" (http://www.ddd.dda.dk/ddd_en.htm). Also in Denmark, the Filming Centre is in charge of the development of electronic accessibility records following scanning from microfilm and microcards or directly from records (http://www.sa.dk/content/us/about_us/filming_centre). The Danish National Business Archives collect and preserve archival material with the purpose of clarifying the history and development of Danish trade and industry.

(http://www.sa.dk/content/us/about_us/danish_business_archives). Finally, DANS provides access to archaeological data.

3.4.f Conclusions

The diversity of fields of research that are now emerging as new interests for social science research should prompt CESSDA to 1) discuss frontiers and intersections in the near future, and 2) secure a role as an intermediary in these fields, despite the fact that it has less experience with health, the environment, history, and qualitative research than its traditional domain of social surveys. This does not mean that CESSDA should include these data in its own collections but it might consider providing indirect support to social scientists that need to access them. Forms of support may range from information services to building actual partnerships with external providers that have more specific expertise for these data. In areas where some CESSDA members already have expertise such as history, one solution might be an internal arrangement to allow exchange of advice and knowledge-sharing between members. Finally in areas that are rapidly developing such as geo-referenced data, CESSDA should first consider monitoring new developments closely in the next few years.

II. Task 5

Overview and Recommendations

II. 1.: Overview on strengths, weaknesses, areas of expertise in data collections (CESSDA & non-CESSDA)

To state the problem in one sentence: ***Data exist in as much as they can be identified and accessed.*** As long as we separate the processes in the culture of *data creation* from those in the culture of *data provision,* we can set up a strategy for identifying these two key states in the life-cycle of data. In general terms, social sciences have been suffering from a poor culture of data provision in comparison to other scientific fields, and of polysemy as to the boundaries on what constitutes social sciences research. The first is due to fragmentation of research and inadequate infrastructure; the second is due to the interdisciplinary nature of social sciences.

In reference to the culture of data creation, it has been historically the case in the social sciences that the term "research infrastructure" has been interpreted in a technical sense, and thus ignored as a need for being the supporting environment for this type of research. Multiplication and overlapping of research activities, the uneven distribution of SSH research infrastructures across Europe and, what we may call for the purposes of this work, '***a veiled landscape of research production',*** are the outcomes of the lack of a central mechanism for registering all kinds of SSH research production in Europe. The existence of such a mechanism would be the ideal state for European research; significant positive steps towards this purpose can be seen in the *culture of European wide comparative large collections* arising from such projects as the European Social Survey, European Value Study, Luxembourg Income Study, etc.

From the interdisciplinary point of view, we can dare to characterise the blurring of boundaries as an 'inherent weakness' of the social sciences. This weakness is stressed by the fact that an extended array of data arising from subjects of other disciplines can be interpreted as being relevant to the social sciences, in contrast with other scientific domains where the research boundaries are clear and agreed. From a broad perspective, any research production which touches upon the behaviour and ways of life of individuals could potentially be exploited through the viewpoint and methods of social scientists. Thus, we observe the attentiveness of sociologists and other social scientists to research on environmental issues, health issues, urban & agricultural planning, to name a few.

In reference to the culture of data provision, the existence of CESSDA has contributed significantly to the promotion of data archives in European countries. As a consequence,

the large number of datasets deposited at CESSDA data archives is already considered a significant "mass" of social sciences data for Europe; this, along with parallel research infrastructures like ESS, large scale data like EuroStat and OECD, and European wide panel programmes, constitute a vast data production landscape accessible at a sufficient level.

But, in as much as Europe is about similarities it is also about differences. In other words, differences in the cultures of data sharing across countries reflect not only differences in corresponding research cultures, but hinder the potential for the 'opening' of the communities towards revealing these differences. Thus, a major challenge/dilemma is rising: to facilitate large infrastructures for wide-scale comparisons, versus focusing on the 'uncommon', nation-private, possibly not harmonisable qualities of certain parts in Europe.

Currently, the usual practice is to focus on easily accessed data nationally and/or internationally, unconsciously reinforcing production towards 'popular' research subjects, abundance of certain data types over less usual types, and provision to large scale dissemination infrastructures. On the other hand, 'small scale' projects, culturally focused research, uncommon data types, facilities with restricted visibility and national production by agents not included in the infrastructures fall on the *veiled side of the data landscape,* suffering promotional inequality, poor usage statistics, and therefore low prospects for further financing and development.

For the future cessda-ERIC a clear view of the data landscape means:

- Focused targeting on data not currently accessible;
- Strategies for expanding acquisition policies;
- Harmonising dissemination policies;
- Realistic planning for networking with data agents at various phases in the life-cycle of data;
- Enhancing the professional character of the 'data business';
- Strengthening the bonds among research actors; i.e. academic-administrative-business actors;
- Internationalising research in terms of provision, while focusing at national/cultural boundaries in terms of production.

This report aimed at shedding light to the *landscape of social sciences data production* in Europe, focusing on data collections which *exist within* CESSDA and those which *exist outside*. The basic source of information has been what is believed to be the 'specialists' in this production – namely, CESSDA member organisations.

The enquiry motivating this task can be summarised under three broad questions:

1. *How far is social research production covered by data archives the members of CESSDA?*
2. *What is produced, how is it provided and where – if not through CESSDA?*

3. *What can CESSDA do to enhance acquisition and dissemination of the collections existing in the European research area?*

II. 2.: Summary of findings and recommendations for the future CESSDA ERIC

Currently CESSDA data organisations spread across 20 countries. Their affiliation in the CESSDA network ensures that more than 25000 datasets are visible to the social science research community. To what degree does the actual research production in social sciences reaches the CESSDA member? To respond to this question we must take either of two stances:

a) find ways to register all research production in each country;
b) study the collection policies in each country and in-depth analysis of their effectiveness.

It is clearly unrealistic to register all research production, 1) because it is a dynamic figure, 2) because the boundaries in social sciences research are not clear. Therefore the method is to investigate which collection policies are applied, how effective they are and how their products – namely, data collections - relate to the needs of the users. The CESSDA PPP survey provided information on collection policies the analysis of which is in the report relevant to Task 4 [T4]. In addition, more sources were used for a 'case study type' investigation on the research production by a variety of agents in selected countries. The case study results are inconclusive and can be used as a pilot for further study. A fact worth considering though, is that there is opacity about research production at local/national level, as viewed from outside (*see: Notices -follow page link >>98).*

From the CESSDA PPP survey it is not clear whether the collection policies applied are effective and how they relate to covering the needs of the end-users; the most striking findings refer to the following facts:

a) The absolute lack of the requirement of exclusiveness for deposited data;
b) The large percentage of data archives engaging in networking activities for tracing and acquiring data;
c) The small number of datasets held by most CESSDA archives (from 1 to 1000 by 12 out of 20 CESSDA organisations);
d) The scarcity of holdings in qualitative datasets in most organisations;
e) The concentration of data collections around a few study topics compared to the topics which are potentially relevant to social research (section 2 pages 26-29).

In terms of research production which is not available through the CESSDA archives the case study on 9 countries, out of 20 currently participating in the CESSDA network, revealed that the majority of producers are in the public sector, followed by universities and research centres. This is also the prevailing 'clientele' of CESSDA organisations, in the role of depositors, rated with the best quality of relationships. Their production is

mostly available in '*metadata*' and other types of data i.e. reports, tables, products of secondary analysis, etc., which means that microdata are difficult to access outside a centralised infrastructure. This finding has important implications for the future of cessda ERIC, since it can set up the controlled environment for the management of microdata arising from small producers at country level and provide the means for their safe dissemination.

Strengths and weaknesses based on the above are listed in Table 5.

Table 5:

	Known Facts	**Strengths**	**Weaknesses**	**Unknown facts**
Current CESSDA Collection policies	The absolute **lack of the requirement of exclusiveness** for deposited data	Professional handling: • data management • quality control • centrality of information about production • facilitating contacts among researchers	Technological advances allow users to turn to other sources for the same data	How are the same collections disseminated outside CESSDA?
	The large percentage of data archives engaging in **networking activities** for tracing and acquiring data	Production at the country level would not be visible if a central collection point did not exist to show interest in the production	The direction of networking activities depends on the "visibility" of the organisation: large and established attract producers, small need to initiate networking activities	How effective are these activities? What is their spread in terms of geography? i.e. local, national, international. How are they financed? What is their regularity?
	The small number of datasets in archives	Perhaps holdings reflect the needs of researchers, so archives tend to collect what is circulated and requested	Filtering data collections might increase the level of quality, but at the same time decrease the number and frequency of users	Usage statistics
	The scarcity of holdings in qualitative datasets in most	Technical and methodological knowledge accumulated over the long period of archiving activities for quantitative	A vastly emerging area of research in SSH is not covered. Data collections including both types of data are difficult to handle with	What is the interest for this type of data? How effective is the 'data archiving' model for qualitative data?

		data can be capitalised for the same purpose on qualitative datasets	the current methods	
	The concentration of data collections around a few study topics compared to the topics which are potentially relevant to social research (pages 19 & 21)	High level of expertise; allows comparative use; enhances international cooperation; allows data harmonisation	There are 'untouched' subject areas which are difficult to find without a central collection point; small 'favourability' in certain subjects leads to lower frequency of use, less visibility, and reinforces the concentration of research on a restricted number of subjects, thus narrowing the scope of social sciences.	What does this uneven distribution reflect? Preferences in certain types of research? Inadequate infrastructures? A "study topics market" reflecting research traditions?
Collections existing outside CESSDA	The majority of producers are in the public sector seconded by universities and research centres.	This is the prevailing 'clientele' of CESSDA, in the role of depositors, rated with the best quality of relationships.	There is low visibility of producers in the private sector.	What are the conditions under which the producers not currently cooperating with CESSDA would agree to provide their collections?
	Production is mostly available in '*metadata*' and other types of data i.e.: reports, tables, products of secondary analysis, etc., which means that microdata are difficult to be accessed outside a centralised infrastructure.	The cessda-ERIC can set up the controlled environment for the management of microdata arising from small producers at country level and provide the means for their safe dissemination.	Microdata produced by non-authenticated producers and for purposes other than research might be of low quality; regularity of production is unknown.	The conditions and procedures under which microdata is produced are unknown in the case of 'small' and/or irregular production; researchers' interest on microdata produced under uncontrolled circumstances is unknown.

Questions which could not be answered by the use of the sources in this work are:

-When researchers apply for microdata outside the central national infrastructures, are the collections provided?

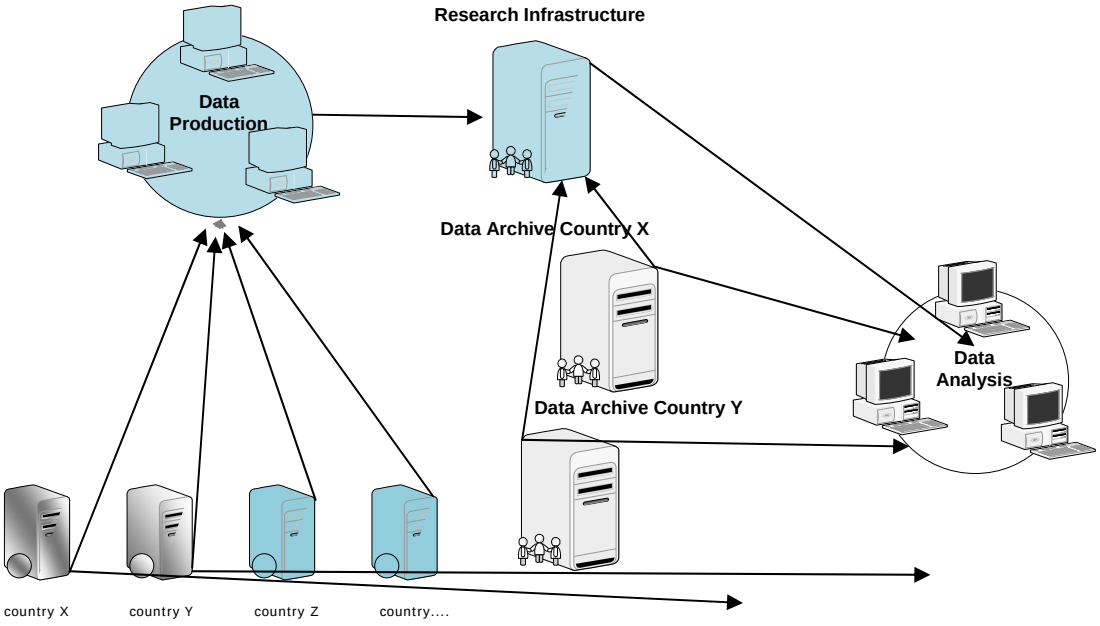
-By which process?

-How long do they need to wait?

-Can they find the same datasets in an established infrastructure at another European country?

The research network is a complicated schema with flows at different points during the production-analysis continuum; the basic flow is to be traced on the disparate directions the research production takes, despite the existence of local or central infrastructure. The schema in image 3 graphically represents this disparity and eventual loss of part of the production.

Image 3. The Research Network



An indirect finding during the work for this report is alarming: there is restricted knowledge of production outside the archive network at national level. The researcher from another country has limited access to information by national research producers, unless they are part of a national infrastructure; more alarmingly, the native researcher is almost equally uninformed, unless there is personal or institutional connection with the producer. Visibility of production relies on the promotional plan of the producer. Large scale producers such as administrative bodies, banks, private businesses use variable transparency of their production targeted to different users.

Generally speaking, production is significant in terms of quantities and use value over three lines:

a) As components of European/international projects;

b) As research of national scope; those two are followed by;

c) Irregular academic research consumed mainly within the academic sector.

While line 'a' is generally accessible (see chapter 3), lines 'b' and 'c' are widely obscured.

Recommendations arising from the summary of results so far are as follows:

Recommendation #1:

cessda ERIC must develop a strategic plan on networking activities spread across all types of producers at national level; the cessda ERIC networking plan must take into account local –i.e.: national distinctiveness, while providing harmonised procedures for attracting and evaluating microdata collections produced by national agents lacking a centralised mechanism of data management and dissemination.

Recommendation #2:

Data archives with low visibility at national level must be supported by cessda ERIC networking and promotion initiatives, so that their collections can be expanded in terms of geographical coverage, and quantity.

Recommendation #3:

The disadvantages accompanying the requirement of non-exclusiveness in data deposit must be turned to advantages within the cessda ERIC; professional handling of data, quality of services, abundance and variety of metadata, high performance in specialised tools – e.g. data harmonisation tools must be promoted and advertised, so that researchers will eventually be reinforced to prefer cessda ERIC over other sources for the same data collections.

Throughout the T4 report it is stressed in many different ways that there is an open ground for 'non-conventional' or 'unusual' data collections. These can be categorised in two classes: a) data arising from qualitative methods of research, b) data produced by other scientific fields not directly connected to the social sciences. In a collective perspective CESSDA is lacking in data of these kinds. Especially in the case of qualitative data, where production is most usually evident in scientific fields in the domain of Humanities, infrastructure resembling the data archives and the CESSDA network does not exist. Therefore, the cessda-ERIC must work towards a two-fold purpose: a) to invest in attracting qualitative datasets with the shared investment in projects and tools to provide support on documentation and handling of qualitative data,

b) to invest in enhancing co-operation with parallel existing and developing infrastructures in Humanities, to promote the use of their data by social scientists.

Recommendation #4:

To take actions for facilitating the acquisition of qualitative data collections along four lines: a) identify demand for qualitative data across all cessda ERIC members, b) target research on specialised tools and services for qualitative data, c) facilitate provision of qualitative data collections by campaigning the advantages of archiving them, d) establish long term collaborations with other organisations specialised in handling qualitative data.

In further focus on 'unusual' or 'unconventional' data collections, T4 work identified initiatives arising from other scientific disciplines which relate to social sciences data collections either in an interdisciplinary perspective, – as is the case with health, historical, environmental and psychology data, or in a methodological perspective, – as is the case with geo-tagged data. Although this report does not reveal any evidence about the demand and use of such types of data by sociologists and other social scientists, the fact that projects and collaborative initiatives exist across many European countries (see, section 3.4) is indicative of an emerging class of data collections which CESSDA is currently lacking.

Recommendation #5:

Outreach activities of cessda ERIC must focus on identifying research initiatives of other disciplines which not only touch upon social issues, but also provide grounds for methodological developments to support social sciences data collections, such as geographical classification of data; cooperation activities must be initiated and tools must be developed and/or adapted for applications on social sciences data.

CESSDA organisations work closely with the academic sector, yet there is still room for further co-operation. At the same time, private polling institutes produce large quantities of data but are lacking wide dissemination activities and preservation technology. Given the fact that a large number of polling institutes in Europe deal with political data, – a fact certified by this work as well, - while political data are the most favourite collections among CESSDA members, grounds for co-operation exist not only in terms of collections, but also in the form of knowledge exchange.

Recommendation #6:

cessda ERIC must liaise with data producers which do not fit its membership schema but are important in populating its collections. A distinctive group of such producers is the polling organisations; their main strength lies on their

abundant production of data relevant to political behaviour, while their main weakness lies in the restricted usage of such production and poor exploitation of microdata for other research purposes. cessda ERIC must work on collaboration agreements on the basis of expertise exchange in the form of providing data management and dissemination services in exchange for receiving polls data of historical and comparative value.

A further finding worth to be discussed and followed up is that in the countries studied, there was a trace of tendency towards focused production on a core topic; this must be studied further to correct for the non-representativeness of the sample used[22]; if there is actually a trend in specialising in certain research topics at country level, the implications for the cessda ERIC are vast, in terms of facilitating and promoting local networks for exploiting expertise in particular research areas.

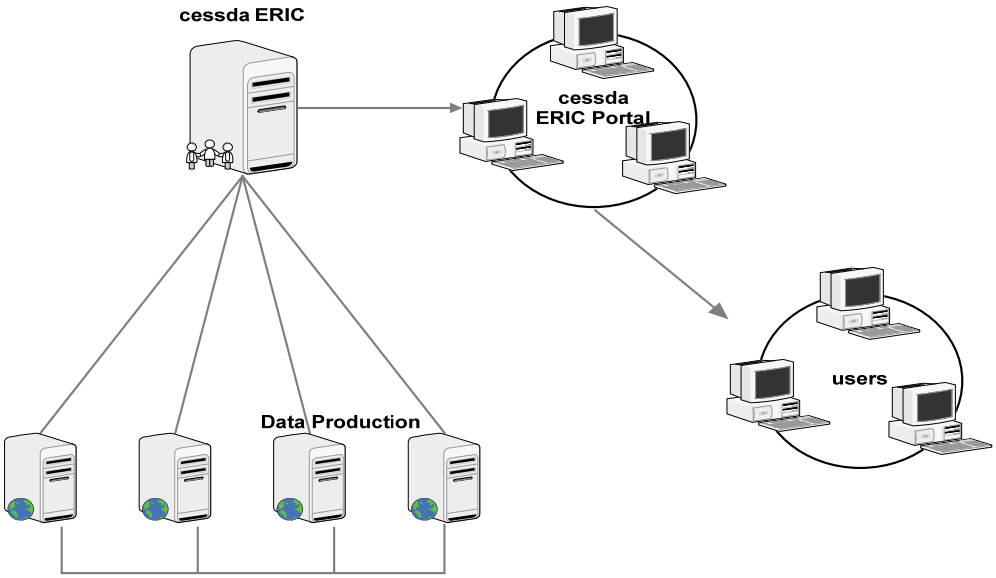
Recommendation #7:

Engage in research to identify the '*study topics markets*' existing within the CESSDA network, which reflect strong and consistent research traditions on a selection of topics; based on the identification of 'strong research traditions' across different topics among members, to capitalise on their strength by engaging in expertise exchanges.

Basically, the role of the national archives is to provide research data for secondary analysis. Research production is rich in microdata and metadata, a fraction of which reach the data archives. It needs to be realised that it is in the interest of its members to attract national production and promote it through the infrastructure. Ideally, the situation cessda ERIC will tend to is graphically presented in image 4.

[22] Figure 14.

Image 4. Research Network through central infrastructure



Recommendation #8:

Observing the local production and how it can be accessed must be an ongoing goal for the new cessda ERIC. cessda ERIC must engage in promotional initiatives to act as a gateway, – even, in certain cases, at the level of information on the sources of data and facilitating and homogeneous culture in sharing and use of data across European countries.

The fact that the new cessda-ERIC is not expected to be the 20+ network from its beginnings does not threaten the prospect of expanding co-operation across all European countries, – be they full members or of some other type of membership. 'In-shop' data collections as well as mechanisms to direct cessda-ERIC users to other sources will be the outcome of harmonised strategic activities for discovery, acquisition (of actual data or information on their sources), dissemination. The experience of large and long established CESSDA partners will be valuable towards this direction.

Recommendation #9:

Equal consideration must be taken on data sources: cessda ERIC 'in-shop' acquisitions –i.e.: collections held by cessda ERIC members, and outside sources – e.g.: journal publishers requiring deposit of data in reference to publications, or NSI collections. We must ensure that both sources are visible through the Portal with analytic documentation on accessing, thus engaging in 'best practices' in reference to the wide research community and promoting positive co-operation with other data publishers. Possible modes of cooperation must be a permanent part of the agenda of the cessda ERIC's outreach activities.

II.3: Needs identified and recommendations for action by cessda ERIC towards satisfying these needs.

The actions recommended to be taken by the new ERIC in terms of widening the data collections to cover the majority of research production across Europe can be summarised in 5 main lines:

- Researching
- Networking
- Promoting
- Managing resources
- Implementing procedures

In this section we give justification for these actions, through the needs identified, as they arise from the work of T4, followed by relevant practical recommendations.

II.3.1: Need to expand the data collections in number, time, space & subjects

To accomplish this expansion we need **targeted research** in two key areas:

a. The social science research activity at national level; its path from the producer to the user

As mentioned already, the work set up in Task 4 only partially identified the research activity at national level, and provided indications only that local factors affect this research activity and its products. An in-depth approach with the involvement of representatives from each country studied will reveal reliable information on local cultures and research traditions. This kind of approach is necessary for the **management of resources** and the **implementation of procedures** contributing to the network character of the new ERIC. Further, the network character of the cessda-ERIC as designed so far allows for the inclusion of several types of actors contributing to the creation of data. Thus, ideally, representatives of the national research communities i.e. experienced researchers, participating in the governance structure of the cessda-ERIC would take up the roles of informants and advisors on research activities, needs, trends and uses at national level.

Recommendation #10:

Cessda ERIC, during its construction phase, must engage in research at national level with the purpose of identifying the localities of research traditions: from production to exploitation; case studies with in-depth approach in 'extreme' cases might also be needed.

Recommendation #11:

Cessda ERIC must include in its governance schema an advisory body representing the countries and consisting of active researchers, with the purpose of monitoring the national research needs and activities and providing feedback to the ERIC.

ii. *The users' needs; their profile, their preferences, their research products*

A large area of research is lacking from the design of this Work Package and the CESSDA PPP in general. It is research on the users' behaviour and needs in terms of the data collections. Users are yet another clientele group of the ERIC and collective knowledge about them will allow effective **promotion** of services, better **management of the resources** and targeted **implementation of procedures** in relation to data acquisition. Knowledge of the perspective of the users is important and it is recommended that the ERIC acquires this either by future research or by the contribution of ERIC members in the form of national reports on current usage of their services; the reports must contain both quantitative and qualitative information and must allow analysis in a harmonised manner.

Recommendation #12:

Given the fact that knowledge on the current situation in reference to the perspective of the users: their preferences in certain data collections; satisfaction of service provided, the end-products based on data acquired; is not currently available in a harmonised manner across CESSDA, it is recommended that cessda ERIC will take action in: a) identifying users' needs, b) auditing their 'behaviour' for effective promotion of services. This can be accomplished in two phases: a) during construction phase, doing comparative research on users' profiles; b) during implementation phase, requiring homogeneous reporting on users' data. The details of this type of reporting can be included in the SLA as a component of reporting procedures for user registration and authentication.

II.3.2: Need to strengthen existing relations and attract new actors -producers in the infrastructure.

To accomplish this, proactive strategies need to be designed and adapted to targeted groups of producers. See for example relevant recommendations for NSIs cooperation in T1-T2 (Tubaro, et al., 2009) and recommendations on "interoperating with external data resources" in WP7, Part II- (Hausstein, et al., 2009)**.**

The cessda-ERIC must decide on the borderlines of what type of research and which producers will form its future network for data provision; this decision is going to be affected by the results of research as in II.3.1.i & II.3.1.ii. The existing relationships

with various groups of actors need to be further analysed in a more refined manner; good relationships need to be justified as a source of 'best practice' conditions. Poor relationships need careful attention as well; what are the common factors in the quality of relationships across current CESSDA organisations? Which business models are appropriate for strengthening relationships?

Recommendation #13:

It is recommended that cessda ERIC shall capitalise on the collective expertise and human capital of the current CESSDA members in reference to existing relationships with various types of producers, in order to set up an *expert group* for the design of a strategy to attract producers across Europe and the world, provide conditions for improving poor relationships and stabilise high quality relationships. This expert group must consist of individuals specialised in an array of subject areas, so that data production of both 'conventional' and 'un-conventional' areas of research can be attracted.

II.3.3: Need to invest in technology for hosting production in new areas of research

To accomplish this, the cessda-ERIC must sensitise and occupy its human resources in setting needs, specifications, demands and courses of action for incorporating new and 'unusual' types of data into the data archiving model. Further research on tools and the recruitment of specialised professionals is also needed in the future. While the identity of the infrastructure must remain focused on exploiting experts' knowledge of data archiving, data management and dissemination, where training and professionalization practices must be applied[23], it must also explore the potential of incorporating individuals with specialised knowledge in other areas. A decentralised system of managing the infrastructure is perhaps more promising in this respect, because it allows the use of expertise from the 'periphery' of the ERIC, as well as from external networks; in the philosophy of ***best management of resources***, programmes of expert exchanges, training in special topics, consultancy and subcontracting must be considered.

Recommendation #14:

New data collections, expanding in numbers and varied in kind, demand specific technology for their management; it is particularly stressed that cessda ERIC must engage in activities towards heightening the professional level of staff by investing in the employment of specialised experts and in training programmes.

[23] Specific reports relevant to training have been produced as part of the CESSDA PPP project, and specifically in WP6 (Krejci, et al., 2009), (Krejci, et al., 2009).

At the same time a constant course of action is needed for the development of tools which will support archiving of new types of data.

II.3.4: Need to set guidelines and harmonise procedures for data collection across its members

It is evident from the work under Task 4 that CESSDA is suffering from variation in procedures at all levels of the archiving process; this is also true for collection and acquisition policies which are in focus. This weakness has two facets:

a- lack of clear, agreed and operational definitions on key concepts used in the work flow of archives,
b- great variability in day-to-day practices across countries for accomplishing the same task, e.g. data acquisition, promotion of services.

It is of course realistic to expect and accept variation in a diversity of cultures, which is often welcome, but the idea of an infrastructure inherently implies convergence in, at least, core practices. The provision of guidelines as the simplest kind of tools is a resourceful activity, especially when targeted at less-experienced members of the network, and contributes to the **implementation of harmonised procedures** and **promotion of professionalism**.

Recommendation #15:

It is recommended that a set of guidelines and procedures is produced, serving as the minimum actions to be taken for the management of data collections. This tool must also include operational definitions of key concepts of the data archiving profession in all languages, complementary to the work of WP4 on the ELSST thesaurus and the Controlled Vocabularies used for documentation.

Examples of such key concepts relevant to the work of T4 are:

- Types of data
- Kinds of data
- Data producers
- Dataset

II.3.5: Need to set-up a mechanism for updating on new production including primary production and production arising from the use of datasets through the cessda ERIC and the web

Evidently, it is only a display of realism to state that in a world of fast and effective technological advances any knowledgeable individual can have direct access to large amounts of information on research projects and their products, through the web. One could argue that a mediating agent, such as the cessda-ERIC, with a ***mechanism for updating on new production*** would only add overlapping information to the already existing mass.

The justification for this action does not focus on information; it focuses on **quality control** and **evolution of research**. As mentioned in T4 -section 3.1, the recent trend is towards direct dissemination, and open access. The cessda-ERIC must be in line with the trend while advancing a culture of quality in the management of research products. Certain applications developed within the CESSDA-PPP, such as the "CHARMCATS" by WP9, are implementations contributing to this purpose.

Recommendation #16:

It is recommended that cessda ERIC will consider incorporating in the design of its portal the facility of linking with research resources internationally, which can also be updated by authenticated users and screened through clearly set criteria of quality, in accordance with the quality criteria used for acquisitions in the infrastructure.

Further, practice so far has shown that most researchers who have used the service of data archives for acquiring data, do not 'return' to deposit the new product of their research, neither do they provide information to their source, the archive, about the physical state of their product. Their research falls into the "veiled landscape", thus hindering further use.

Recommendation #17:

It is recommended that cessda ERIC will set-up a feedback mechanism – offering appropriate incentives - from users of datasets to the infrastructure, for referring back the products of their work, based on datasets acquired either from the infrastructure or directly disseminated to them through other web sources.

II.4.: Concluding remarks

The original incentive for initiation of Tasks 4 & 5 was to gain knowledge about the data collection strategies across current CESSDA organisations, to identify their relation with the data producers outside CESSDA, point to their strengths, weaknesses and expertise and provide a set of recommendations for the future cessda-ERIC.

Research and recommendations on the subject are not exhaustive but hopefully provide the basis for the initiation of actions towards a new approach in ***researching***, ***networking***, ***promoting***, ***managing resources*** and ***implementing procedures***, in accordance with the general goals set for the new infrastructure.

The critical elements of the new approach lie on the idea of a network in the model of an organism, where no part is the same but all are needed for its proper functioning, enjoying common proprietary rights, shared expertise, distributed facilities, common goals and targets, complementary activities.

In reference to a strategy on data collections, all the critical elements apply but refinements in plans and actions are needed for its implementation, while the time factor is essential for their hierarchical ordering, especially in relevance to entering into new grounds of research production; thus, some of the recommendations in T5 do not in any way imply actions to be taken in the immediate future, but take the form of either guidelines or a 'wish list' for the functioning of an infrastructure which will constantly assess and modify its targets depending on the developments in the international research community.

III. Reference material

DDI Alliance Making XML files for the new IDC of CESSDA: Mandatory and recommented fields [Conference] // Minutes from the CDG meeting at Essex, 5-6 September. - Colchester, UK : DDI Alliance, 2001. - http://www.ddialliance.org/related/cessda-rec.pdf.

Declaration Berlin Berlin Declaration on Open Access to Knowledge in the sciences and Humanities [Online]. - 2003. - 6 16, 2009. - http://oa.mpg.de/openaccess-berlin/berlin declaration.pdf.

ESFRI European Roadmap for Research Infrastructure [Report]. - 2006.

FORS FORS Bremen Workshop. "Qualitative Longitudinal Research and Qualitative Resources in Europe: Mapping the Field and Exploring Strategies for Development, with support from CESSDA [Conference].

Hausstein Brigitte [et al.] WP7 Final Report: Widening the cessda ERIC: Inclusiveness and Comprehensiveness of the upgraded European Research infrastructure [Report]. - 2009.

Heiko Tjalsma (ed.) Valentijn Gilissen, Henk Koning and Dirk Roorda (DANS) Data Preservation in the Social Sciences: Recommendations for a CESSDA Research Infrastructure [Report]. - 2009.

Kallas John Issues on the Design of Empirical Research (in greek) [Book]. - Athens : National centre for Social research & Nefeli publishing Co., 2002. - ISBN-960-211-654-4.

Krejci Jindrich and Cizek Tomas Current training Practices and Training Opportunities for the CESSDA [Report] : Interim Report. - 2009.

Krejci Jindrich, Hausstein Brigitte and Dusa Adrian Establishing a CESSDA Training Programme: Challenges and Possible Strategies [Report] : Interim report. - 2009.

Lyon L. Dealing with Data: Roles, Rights, Responsibilities and Relationships [Online] // Consultancy Report, UKOLN, University of Bath. - 5 19, 2007. - 6 12, 2009. - http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf.

Martinez-Uribe L. Digital Repository Services for Managing Research Data: What Do Oxford Researchers Need? [Journal]. - [s.l.] : IASSIST Quarterly, 2007. - Fall & Winter.

Mochman E. International Social Science Data Service: Scope and Accessibility [Report] : Report for the International Social Science Council. - Cologne : [s.n.], 2002.

MORESS Report [Online]. - 6 12, 2009. - http://www.moress.org/report.html.

RISSH Research Infrastructure in the Humanities & Social Sciences WG Blueprint for the European Research Observatory for the Humanities and Social Sciences -EROHS [Report] = EROHS Report. - Copenhagen : Danish Research Council for the Humanities & Danish Social Science Research Council, 2004. - p. 19. - ISBN: 87-90201-72-8.

Roundtable I -Data for Comparative European Socio-Economic Research: Challenges, Opportunities and Constraints [Conference] = NESSIE Report. - Colchester : Network of Economic and Social Science Information, 2002. - http://www.nessie-essex.co.uk/roundtable_1Report.pdf.

Sotiropoulos Dimitris Survey for the Higher Educational System in Greece [Online] // ELIAMEP. - 2006. - 1 21, 2008. - www.eliamep.gr.

SSH Futures EU Project Social Sciences and Humanities for Europe (SSH futures); Instrument: Specific Targeted Research Project, Thematic Priority 7: Citizens and Governance in a Knowledge-Based Society [Online]. - 2008. - 4 25, 2009. - http://www.iccr-international.org/sshfutures/docs/SSH-FUTURES.

Timescapes Qualitative longitudinal Study and Archive [Online]. - 6 12, 2009. - http://www.timescapes.leeds.ac.uk.

Tubaro Paola [et al.] Audit Report on Access Mechanisms and Availability of Official Statistics across the European Research Area, with Recommendation on Role of CESSDA Resaerch Infrastructure in Facilitating Greater Access [Report] : Final Report. - 2009.

IV. Appendices

A. Terminology and Definitions

Large Scale Collections at National Level (LC)

National Archives are organisations assigned for the collection of data and documents on social sciences, humanities and other disciplines, recognized as official repositories of data in the country.

CESSDA organisation

Any organisation representing its country in the CESSDA network. CESSDA organisations are defined as those which are members of CESSDA because they contribute to the aims of CESSDA as they are described at Article II: Membership in the CESSDA Governance (CESSDA, 2004).

Large Scale Collections at national Level contractors

Those organisations which produce microdata on behalf and/or under the framework of LC production; the data produced are relevant to a narrow scope of topics reflecting the specialization of the organisation.

Metadata

The metadata referred to in this project is the physical representation of meta-information including all elements of information which effectively guide and support the process of identification and extraction of relevant survey data and those which are needed for their valid interpretation (def. used on MetaDater project 2002-2005)

B. Variable Values and Value Names

Organisation Role	Organisation Function	Organisation Type	Actor in RE to CESSDA Archive	Kinds of Data	Subjects (SSH)
PD: Data Producer	Data production	LC: Large Scale Collections at National Level	Parallel	Microdata	Economics
PV: Data Provider	Dissemination (free, special access)	ODO: Other Data Organisation	Concurrent	Metadata	Labour & Employment
DA: Data Archive	Data archiving		LC (Large Scale Collections at National Level) contractors	Other (publications, reports, indicators, links, references, as well as other forms of contextual metadata)	Trade, Industry & Markets
PD-PV: Data Producer-Provider					Society & Culture
PD-DA: Data Producer-Data Archive					Education
					Social Stratification and Groupings
					Inequalities
					Reference and Instructional Resources
					Law, Crime and Legal Systems
					Demography and Population
					Migration
					History
					Housing and Land Use Planning
					Natural Environment
					Transport, Travel and Mobility
					Information and Communication & media
					Science and Technology
					Health
					Social Policy and Systems
					Politics
					All or most of the categories above
					Other

C. Case Studies
Lists of SSH data producers in each country studied

France

Name	Role	Org_ Type	Actor	Main Subject(s)	Web-site
Archives nationales (national archives)	DA	LC	parallel	Labour and Employment, Social Policy and Systems	http://www.archivesnationales.culture.gouv.fr/
Archives de la recherche en sciences humaines et socials (ARSHS)	DA	ODO	parallel	Society and Culture, Social Stratification and Groupings-Inequalities	http://www.msh-reseau.fr/spip.php?article34
Centre de recherche pour l'étude et l'observation des conditions de vie (CREDOC)	PD	ODO	parallel	Society and Culture, Social Policy and Systems	http://www.credoc.fr/
La Fédération nationale des observatoires régionaux de la santé (FNORS)	PD	ODO	parallel	Health	http://www.fnors.org/index.html
L'Institut du Développement Social (IDS)	PD	ODO	parallel	Social Stratification and Groupings-Inequalities, Health	http://www.ids.fr
l'Institut Français d'Opinion Publique (IFOP)	PD	ODO	parallel	Health, Politics	http://www.ifop.com/europe
Institut Géographique National (IGN)	PD-PV	ODO	parallel	Natural Environment	http://www.ign.fr/
Institut national de prévention et d'éducation pour la santé (INPES)	PD-PV	ODO	LC contractors	Health	http://www.inpes.sante.fr/
Institut de la recherche agronomique (INRA)	PD	ODO	parallel	Housing and Land Use Planning, Natura Environment	http://www.inra.fr/
Institut national de recherche sur les transports et leur sécurité (INRETS)	PD	ODO	parallel	Transport, Travel and Mobility	http://www.inrets.fr/
l'Institut national de la santé et de la recherche	PD	LC	parallel	Health, Science and Technology	http://www.inserm.fr/r/

médicale (INSERM)					
IPSOS	PD:DA	ODO	concurrent	Social Stratification and Groupings-Inequalities, Health	http://www.ipsos.fr/
Institut de recherche et de documentation en économie de la santé (IRDES)	PD	ODO	parallel	Trade, industry and Markets, Politics	http://www.irdes.fr/
Institut de recherche en Santé Publique (IRESP)	PD-PV	ODO	parallel	Health	http://www.iresp.net.php
MEDIAMETRIE	PD	ODO	parallel	Health	http://www.mediametrie.fr/
Observatoire de la Délinquance	PD	ODO	parallel	Law, Crime and Lega Systems	http://www.inhes.interieur.gouv.fr/Observatoire-national-de-la-delinquance-6.html
TNS SOFRES	PD	ODO	concurrent	Law, Crime and Lega Systems	http://www.tns-sofres.com/
Banque de France	PD	ODO	parallel	Trade, industry an Markets, Politics	http://www.banque-france.fr

Greece

Name	Role	Org_ Type	Actor	Main Subject(s)	Web-site
Ministry of National Education and Religious Affairs, Institute for the Greek Diaspora Education and Intercultural Studies (IPODE)	PD-PV	ODO	Parallel	Education, Demography and Population-Migration	http://www.ipode.gr
Research Centre for Gende Equality (KETHI)	PD	ODO	Parallel	Social Stratification and Groupings-Inequalities	http://www.kethi.gr/nglish/indexen.htm
Hellenic Foundation of European and Foreign Policy (ELIAMEP)	PD	ODO	Concurrent	Politics	http://www.eliamep.gr/eliamep/content/home/research/en/
Institute of Greek & Roma Antiquity	PD-PV	ODO	Parallel	History	http://www.eie.gr/nhrf/institutes/igra/index-gr.html
Institute for Neohelleni Research	PD-PV	ODO	Parallel	Society and Culture History	http://www.eie.gr/nhrf/institutes/inr/index-gr.html
Institute for Byzantin Research	PD-PV	ODO	Parallel	History	http://www.eie.gr/nhrf/institutes/ibr/index-gr.html
Focus Bari	PD-PV	ODO	Parallel	Trade, industry an Markets, Social Polic and Systems	http://www.focus.gr/default.asp?id=100000001&lcid=1032
Metron Analysis	PD-PV	ODO	Parallel	Politics	http://www.metronanalysis.gr/web/html/index.asp?language=greek&page=about
ICAP	PD-PV	ODO	Parallel	Trade, industry an Markets	http://www.icap.gr/index_uk.asp
Research Internationa Hellas	PD	ODO	Concurrent	Trade, industry an Markets	http://www.research-int.com/worlds/worlds.asp?cou=20&id=
AGB Nielsen Medi Research	PD-PV	ODO	Parallel	Information, Communication an	http://www.agbnielsen.net/whereweare/dynPage.asp?lang=loca

				Media	l&id=316&country=Greece
Centrum Research	PD-PV	ODO	Parallel	Economics, Information, Communication an Media	http://www.centrum.gr/
VPRC	PD-PV	ODO	Parallel	Politics	http://www.v-prc.gr/index_gr.html
Employment Observator Research- Informatics SA (PAEP)	PD-PV	ODO	Parallel	Labour an Employment	http://www.paep.org.gr/eng/index.php
University Researc Institute of Applie Communication (Universit of Athens)	PD	ODO	Parallel	Information, Communication an Media	http://www.media.uoa.gr/institute/pages/eng/identity.html
QUANTOS	PD	ODO	Parallel	Economics, Trade, industry and Markets	http://www.quantos-stat.com/
Regional Development Institute (Panteion University)	PD	ODO	Parallel	Housing and Land Use Planning, Social stratification and Groupings-Inequalities	http://www.ipa.panteion.gr/en/html/main.html
FOUNDATION FOR ECONOMIC & INDUSTRIAL RESEARCH (IOBE)	PD-PV	ODO	Parallel	Economics, Trade, industry and Markets	http://www.iobe.gr/index.asp?a_id=131
Laboratory of demographic and social analysis, University of Thessaly	PD-PV	ODO	Parallel	Demography and population-Migration	demography-lab.prd.uth.gr
Attiko Metro SA,	PD	ODO	Parallel	Transport, Travel and Mobility	http://www.ametro.gr/page/default.asp?id=4&la=2
The Greek Ombudsman	PD-PV	ODO	Parallel	Law, Crime and Lega Systems	http://www.Synogoros.gr/en.index.htm
Opinion SA	PD-PV	ODO	Parallel	Politics	http://www.opinion.gr/
Democritus University of Trace Department of Social Administration	PD	ODO	Concurrent	Social policy and systems	http://www.socadm.uth.gr/

Institute for Language and Speech Processing (Research and Innovation Centre "Athena") – ILSP	PD	ODO	Parallel	Society and Culture	www.ilsp.gr
Industrial Systems(Research and Innovation Centre "Athena") – ISI	PD	ODO	Concurrent	Trade, industry and Markets, Science and Technology	www.isi.gr
"Athena" Cultural and Educational Technology Institute – CETI	PD	ODO	Parallel	Society and Culture, Education	www.ceti.gr
"Athena" Institute for Research on Networking Technologies -IRNET	PD	ODO	Parallel	Science and Technology	www.irnet.gr
"Athena" Institute: Institut for the Management o Information Systems-IMIS	PD	ODO	Parallel	Information, Communication an Media	www.ipsyp.gr
MRB	PD-PV	ODO	Parallel	Politics	http://www.mrb.gr/ (web page under construction)
Archive of State Gazette	PD-PV	LC	Parallel	Law, Crime and Lega Systems	www.et.gr
General Archives of the State	DA	LC	Parallel	Politics, Society and Culture	www.gak.gr
Directorate of Statistical and Actuarial Studies, Social Security Foundation	PD-PV	ODO	LC Contractors	Labour and Employment, Social Policy and Systems	http://www.ika.gr/gr infopages/stats/stat_ eport.cfm
Ministry of National Education and Religious Affairs, Directorate of planning and operational research, Department of statistics and operational research	PD-PV	ODO	LC Contractors	Education	http://www.ypepth.g /el_ec_page6653.htn
National Documentation Centre	PD-PV	LC	Parallel	All or most of the categories above	http://www.ekt.gr/en
Laboratory for Monitoring Social Cohesion Policies, National Center for Social Research	PD	ODO	Concurrent	Social policy and systems	http://www2.ekke.gr main.php?id=352
Laboratory for Gender Issues (Panteion University	PD	ODO	Parallel	Social Stratification and Groupings-	http://www.genderpa nteion.gr/gr/ergastiri

				Inequalities	o.php
Panteion University, Institute of Urban Environment and Human Resources	PD	ODO	Parallel	Housing and Land Use Planning	http://www.uehr.panteion.gr/site/gr/index.php
Institute of Labour, Genera Confederation of Greek Workers	PD-PV	ODO	Parallel	Labour and Employment	http://inegsee.gr
Institute of Social Protection and Solidarity	PD	ODO	Parallel	Social policy and systems	http://www.ikpa.gr/html/system.htm
Historical Archive University of Athens	DA	ODO	Parallel	Education	www. archive. uoa.g
EDUCATION RESEARCH CENTRE OF GREECE (KEE)	PD-PV	ODO	Parallel	Education	http://www.kee.gr/html/english_main.php
Athens Organisation, Urban planning and environmenta protection	PD-PV	ODO	Parallel	Housing and Land Use Planning	http://www.minenv.gr/3/31/313/31303/g3130304.html
Organisation for Professional Education and Training	PD-PV	ODO	Parallel	Education	http://www.oeek.gr/
Bank of Greece	PD-PV	LC	LC Contractors	Economics, Trade, industry and Markets	http://www.bankofgreece.gr/en/
Statistics and Studies Department, Organisation of Agricultural Insurances	PD-PV	ODO	LC Contractors	Social policy and systems	www.oga.gr

Italy

Name	Role	Org_ Type	Actor	Main Subject(s)	Web-site
University of Trento-Department of Sociology & Social Research, Italian Data Archive for the Social Sciences [IDAss]	PD:DA	ODO	Concurrent	Social Stratification and Groupings-Inequalities, Health	http://portale.unitn.it/dsrs/
Banca d'Italia	PD-PV	LC	Parallel	Economics	www.bancaditalia.it/bancaditalia
Centro Ricerche e Studi Direzionali (CERISDI)	PD	ODO	Parallel	Politics, Natural Environment	www.cerisdi.it/
ICstat -International Cooperation Center for Statistics	PD-PV	ODO	Parallel	Demography and population-Migration, Natural Environment	www.icstat.org/
Instituto Nazionale Previdenza Sociale	PD	ODO	Concurrent	Labour and Employment, Social Policy and Systems	www.inps.it/home/default.asp
Institute "Carlo Cattaneo" – Foundation of sociological and political research	PD-PV	ODO	Parallel	Politics, Social Stratification and Groupings-Inequalities	www.cattaneo.org/default.asp
Institute "Guglielmo Tagliacarne" for the promotion of economic culture	PD:DA	ODO	Concurrent	Economics	www.tagliacarne.it/sito/link.asp
Planning Center - Emilia Romagna	DA	ODO	Parallel	Housing and Land Use Planning	www.regione.emilia-romagna.it/planningcenter/
Universitŭ degli Studi di Torino - LABORatorio Riccardo Revelli WHIP - Work Histories Italian Panel	PD-PV	ODO	Parallel	Labour & Employment	http://laboratorioreverelli.it/

Norway

Name	Role	Org_ Type	Actor	Main Subject(s)	Web-site
The University Centre in Svalbard (UNIS)	PD	ODO	concurrent	Natural Environmen	http://www.unis.no/
Nielsen Norge	PD	ODO	parallel	Trade, industry and Markets	http://no.nielsen.com /site/index.shtml (Norwegian)
Norfakta	PD	ODO	parallel	Trade, industry and Markets	http://www.norfakta com
Norstat	PD	ODO	parallel	All or most of the categories above	http://www.norstat.n o
Norway's Central Health Registers	PD:DA	ODO	parallel	Health	http://www.fhi.no
Norwegian Directorate for Education and Training	PD:DA	ODO	LC contractors	Education, Social policy and systems	http://www.udir.no
Norwegian Directorate o Immigration (UDI)	PD:DA	ODO	LC contractors	Demography and population-Migration, Social policy and systems	http://www.udi.no
Norwegian Institute for Urban and Regional Research (NIBR)	PD	ODO	parallel	Housing and Land Use Planning	http://www.en.nibr.n o
Norwegian Labour and Welfare Organisation (NAV)	PD	ODO	LC contractors	Labour and Employment	http://www.nav.no
Norwegian Social Research (NOVA)	PD	ODO	parallel	Social policy and systems	http://www.nova.no
Norwegian University of Life Sciences (UMB)	PD	ODO	concurrent	Natural Environment, Housing and Land Use Planning	http://www.umb.no/ avd=30
Norwegian University of Science and Technology (NTNU)	PD	ODO	concurrent	Society and Culture, Information, Communication and Media	http://www.ntnu.no/ nglish/
Respons	PD	ODO	parallel	Trade, industry and Markets	http://www.responsa nalyse.no
Sentio	PD	ODO	parallel	Politics, Trade, industry and Markets	http://www.sentio.no
Synovate	PD	ODO	parallel	Trade, industry and	http://www.synovate

				Markets	com
The Brønnøysund Register Centre	PD-PV	ODO	LC contractors	Trade, industry and Markets	http://www.brreg.no
The Institute of Transpor Economics (Transportøkonomisk institutt, TØI	PD	ODO	parallel	Transport, Travel and Mobility	http://www.toi.no
The Norwegian Institute for Studies of Research and Education (NIFU STEP)	PD-PV	ODO	parallel	Reference and Instructional Resources, Education	http://english.nifustep.no/
The Norwegian Institute of Public Health	PD	ODO	parallel	Health	http://www.fhi.no
THE SINTEF GROUP	PD	ODO	parallel	Trade, industry and Markets	http://www.sintef.no
The Work Research Institute Ltd. (AFI-WRI)	PD	ODO	parallel	Labour and Employment	http://www.afi-wri.no/index.asp?iLang=1
TNS Gallup	PD	ODO	parallel	Trade, industry and Markets	http://www.tnsgloba.com/global/europe/norway/
University of Agder	PD	ODO	concurrent	Society & Culture, Economics	http://www.uia.no/en
University of Bergen (UiB)	PD	ODO	concurrent	Education, Social Stratification and Groupings-Inequalities	http://uib.no/info/english/
University of Oslo (UiO)	PD	ODO	concurrent	Natural Environment, Social Stratification and Groupings-Inequalities	http://www.uio.no/english/
University of Stavanger	PD	ODO	concurrent	Social Stratification and Groupings-Inequalities	http://www.uis.no/frontpage/
University of Tromsø	PD	ODO	concurrent	Society and Culture	http://uit.no/informasjon/english?Language=en

Romania

Name	Role	Org_ Type	Actor	Main Subject(s)	Web-site
Center of Urban and Regional Sociology	PD	ODO	concurrent	Politics, Housing and Land Use Planning	http://curs.ro/?lang=en
CURS - Centre of Urban and Regional Sociology	PD	ODO	concurrent	Politics, Housing and Land Use Planning	www.curs.ro
Faculty of Social Humanistic Science from the University of Oradea	PD-PV	ODO	concurrent	Society and Culture, Social Stratification and Groupings-Inequalities	www.uoradea.ro/english/
Faculty of Sociology and Social Work (SAS), University of Bucharest.	PD	ODO	parallel	Society and Culture, Social Stratification and Groupings-Inequalities	www.unibuc.ro/en/home
Gfk Romania	PD	ODO	parallel	Politics, Information & Communication-Media	http://www.gfk-ro.com/new/index.html
Infomass	PD	ODO	parallel	Politics, Trade, industry and Markets	http://www.infomass.ro/
INSOMAR	PD	ODO	parallel	Politics, Housing and Land Use Planning	http://www.insomar.o/
Institute for Quality of Life Research (IQLR) [under the aegis of the Romanian Academy of Sciences]	PD	ODO	parallel	Social Stratification and Groupings-Inequalities, Social policy and systems	www.iccv.ro/
International Institute for the Advanced Studies of Psychotherapy and Applied Mental Health	PD	ODO	concurrent	Health	www.psychotherapy.ro/component/option,com_frontpage/Itemid,91/
IRECSON	PD	ODO	parallel	Reference and Instructional Resources	http://www.irecson.ro/index.php?module=info&page=articol&parent_id=131&s_cat=133&articol_id=100
Mercury Research	PD	ODO	parallel	Politics, Trade, industry and Markets	http://www.mercury.ro/index.php
MMT - Metro Media Transilvania.	PD	ODO	parallel	Politics, social Stratification and Groupings-	www.mmt.ro/Engleza/despre.htm

				Inequalities	
National Institute for Economic Research	PD	LC	concurrent	Economics	www.ine.ro
Soros Foundation Romania	PD-PV	ODO	concurrent	Demography and Population, Housing and Land Use Planning	http://soros.ro/en/index.php
The Center for Research on Culture	PD	ODO	parallel	Society and Culture	http://www.culturant.ro/eng/
The Galloup Organisation Romania	PD	ODO	parallel	Politics, Housing and Land Use Planning	http://www.gallup.ro/
The Research Centre on Interethnic Relations	PD	ODO	parallel	Society and Culture, Demography and Population-Migration	http://www.ccrit.ro/index_eng.htm
University of Babes-Bolyai	PD	ODO	concurrent	Society and Culture	http://www.ubbcluj.ro/en/index.html

Slovenia

Name	Role	Org_ Type	Actor	Main Subject(s)	Web-site
University of Ljubljana	PD	ODO	Concurrent	Society and Culture, Natural environment	www.uni-lj.si
Employment Service of Slovenia	PD-PV	ODO	LC contractors	Labour and employment	www.ess.gov.si
Ministry of Culture Archives of the Republic of Slovenia	PD-PV	LC	Parallel	Society and Culture	www.archiv.gov.si
Ministry of Finance	PD-PV	ODO	LC contractors	Economics	www.mf.gov.si
Znanstvenoraziskovalni center Slovenske akademije znanosti in umetnosti	PD	ODO	Concurrent	Society and Culture	www.zrc-sazu.si
University of Maribor	PD	ODO	Concurrent	Housing and Land Use planning, society and culture	www.uni-mb.si
Agency of the Republic of Slovenia for Public Legal Records and Related Services	PD-PV	LC	Parallel	Economics, Trade, industry and Markets	www.aipes.si
Bank of Slovenia	PD	LC	Parallel	Economics	www.bsi.si
Faculty of Social Sciences, University of Ljubljana	PD	ODO	Concurrent	Science and Technology	http://english.fdvinfo .net
Human Rights Ombudsman of the Republic of Slovenia	PD-PV	ODO	Parallel	Law, Crime and Legal Systems	www.varuh-rs.si
Institute for Cultura Studies-University o Nova Gorica	PD	ODO	Concurrent	Society and Culture	www.ung.si
Ministry of Educatio and Sport - Inspectorat for Education and Sport	PD	ODO	LC contractors	Education	www.mss.gov.si
Organisation o Employment Service o Slovenia	PD-PV	ODO	Parallel	Labour an Employment	www.ess.gov.si
Public Opinion Centr	PD	ODO	Parallel	Politics	www.zrs-kp.si

(POC) - Science an Research Centre of th University of Primorska.					
Slovenian Researc Agency	PD	ODO	Parallel	Other	www.arrs.gov.si
University of Primorska	PD	ODO	Concurrent	Other	www.upr.si

Spain

Name	Role	Org_ Type	Actor	Main Subject(s)	Web-site
ASEP/JDS	PD-PV	ODO	Parallel	Politics, Social Stratification and Groupings-Inequalities	http://www.jdsurvey net
Bank of Spain	PD-PV	LC	Parallel	Economics	http://www.bde.es
Centro de Investigacione Sociologicas (CIS)	PD:DA	LC	LC contractors	Politics, Science and Technology, Social Stratification and Groupings-Inequalities	http://www.cis.es
Elcano Royal Institute	PV	ODO	Parallel	Politics, Social Stratification and Groupings-Inequalities	http://www.realinstit utoelcano.org
Instituto de Crédito Oficial (ICO)	PD-PV	ODO	LC contractors	Economics, Trade, industry and Markets	http://www.ico.es/
Instituto de Estudios Sociales Avanzados (IESA-CSIC),	PD-PV	ODO	Parallel	Social Stratification and Groupings-Inequalities, Politics Social policy and systems	http://www.iesa.csic es/es
Instituto de la Mujer/National Women's Institute	PD-PV	ODO	LC contractors	Social Stratification and Groupings-Inequalities	http://www.migualda d.es/mujer/
Juan March Institute-Center for Advanced Study in the Social Sciences	PD	ODO	Parallel	Social Stratification and Groupings-Inequalities, Politics Social policy and systems	http://www.march.es /ceacs
Ministerio de Economía y Hacienda Secretaría de Estado de Economía	PV	ODO	LC contractors	Economics	http://www.meh.es/e s
Ministerio de Fomento / Ministry of Public Works	PD-PV	ODO	LC contractors	Housing and Land Use planning	http://www.fomento es/
Ministerio de Sanidad y Consumo	PD-PV	ODO	LC contractors	Health	http://www.msc.es/
Ministry of Agriculture,	PD-PV	ODO	LC	Housing and Land	http://www.mapa.es

Fisheries and Food			contractors	Use Planning	
Ministry of Culture	PD-PV	ODO	LC contractors	Society and Culture	http://en.www.mcu.e s
Ministry of Education, Social Policy and Physical Education	PD-PV	ODO	LC contractors	Education, Social policy and systems	http://www.mepsyd. s
Ministry of Labour & Immigration- Social Security-	PD-PV	ODO	LC contractors	Labour and Employment, Social Policy and Systems	http://www.mtas.es/e s/seg_soc/
Ministry of the Environment, Rural & Marine Affairs	PD-PV	ODO	LC contractors	Housing and Land Use Planning, Natural Environmen	http://www.marm.es
National Geographic Institut, Ministry of Public Works	PD-PV	ODO	LC contractors	Housing and Land Use planning	http://www.ign.es

Sweden

Name	Role	Org_ Type	Actor	Main Subject(s)	Web-site
National Board of Health and Welfare	PD-PV	LC	Parallel	Health, Social policy and systems	www.socialstyrelsen.se
National Council for Crime Prevention	PD-PV	ODO	LC Contractors	Law, Crime and Legal Systems	www.bra.se
National Institute of Economic Research	PD-PV	ODO	Parallel	Economics, Labour and employment	www.konj.se
National Mediation Office	PD-PV	ODO	Parallel	Labour and Employment, Law, Crime and Legal Systems	www.mi.se
Swedish Agency for Economic and Regional Growth	PD-PV	ODO	Parallel	Economics, Trade, industry and Markets	www.tillvaxtverket.se
Swedish Arts Council	PD-PV	ODO	LC Contractors	Society and Culture	www.kulturradet.se
Swedish Board of Agriculture	PD-PV	ODO	LC Contractors	Housing and Land Use Planning, Social policy and Systems	www.sjv.se
Swedish Board of Fisheries	PD-PV	ODO	LC Contractors	Social policy and systems, Natural environment	www.fisheriverket.se
Swedish Chemicals Agency	PD-PV	ODO	LC Contractors	Natural environment	www.Kemi.se
Swedish Environmental Protection Agency	PD-PV	ODO	LC Contractors	Natural Environment, Law, Crime and Legal Systems	www. naturvardsverket.se
Swedish Financial Supervisory Authority	PD-PV	ODO	Concurrent	Economics	www.fi.se
Swedish Forest Agency	PD-PV	ODO	LC Contractors	Social policy and systems, Natural environment	www.svo.se
Swedish Institute for Growth Policy Studies	PD-PV	ODO	LC Contractors	Social policy and systems	www.tillvaxtanalys.se
Swedish Institute for Transport and Communications Analysis	PD-PV	LC	LC Contractors	Transport, Travel and Mobility	www.sika-institute.se

Swedish National Agency for Education	PD-PV	ODO	LC Contractors	Education	www.skolverket.se
Swedish National Agency for Higher Education	PD-PV	ODO	LC Contractors	Education	www.hsv.se
Swedish National Board for Study Support	PD-PV	ODO	LC Contractors	Social policy and systems, Education	www.csn.se
Swedish National Dept Office	PD-PV	ODO	LC Contractors	Law, Crime and Legal Systems, Economics	www.riksgalden.se
Swedish National Financial Management Authority	PD-PV	ODO	LC Contractors	Economics	www.esv.se
Swedish Social Insurance Agency	PD-PV	ODO	LC Contractors	Social policy and systems	www. forsakringskassn.se
Swedish Work Environment Authority	PD-PV	ODO	Parallel	Labour and employment, Health	www.av.se
National Courts Administration	PD-PV	LC	Concurrent	Law, Crime and Legal Systems, Economics	http://www.domstol.se/

United Kingdom

Name	Role	Org_ Type	Actor	Main Subject(s)	Web-site
Bank of England	PD-PV	LC	Parallel	Economics	http://www.bankofeng land.co.uk
British Market Research Bureau BMRB	PD	ODO	Parallel	Natural Environment, Information, Communication an Media	www.bmrb.co.uk/abo ut/
Centre for Regional Economic and Social Research (CRESR)	PD-PV	ODO	Parallel	Economics, Social Policy and Systems	www.shu.ac.uk/resea ch/cresr/
Department for Children, Schools and Families (DCSF)	PD-PV	ODO	Parallel	Education	http://www.dcsf.gov.u k/index.htm
General Register Office For Scotland	PD-PV	LC	LC Contractors	Demography and Population-Migration	http://www.gro-scotland.gov.uk
Government Social Research	PD-PV	ODO	Parallel	Social Policy and Systems	http://www.gsr.gov.ul /
Learning Skills Development Agency	PD-PV	ODO	Parallel	Education	http://www.lsda.org.u k/home.asp
Low Pay Commission	PD-PV	ODO	Parallel	Economics	http://www.lowpay.g v.uk
Ministry of Justice	PD-PV	LC	Parallel	Law, Crime and Legal Systems	www.justice.gov.uk
National Centre for Social Research (NatCen)	PD-PV	ODO	Parallel	All or most of the categories above	www.natcen.ac.uk/
National Institute of Economic and Social Research	PD-PV	ODO	Parallel	Economics	www.niesr.ac.uk/abou t/about.php
Northern Ireland Department For	PD-PV	ODO	Parallel	Labour and Employment,	http://www.delni.gov.

Employment And Learning				Education	uk
Northern Ireland Neighbourhood Information Service (NINIS)	PD-PV	ODO	Parallel	Education, Demography and Population-Migration	www.ninis.nisra.gov.uk/
Northern Ireland Statistics And Research Agency	PD-PV	LC	LC Contractors	All or most of the categories above	http://www.nisra.gov.uk
Office for Standards in Education	PD-PV	ODO	Parallel	Education	http://www.ofsted.gov.uk/
School of Social Policy, Sociology and Social Research	PD-PV	ODO	Parallel	Health, Social Policy and Systems	www.kent.ac.uk/sspssr/
Scottish Official Statistic	PD-PV	LC	LC Contractors	All or most of the categories above	http://www.scotland.gov.uk/Topics/Statistics/
SIRC (Social Issues Research Centre)	PD	ODO	Parallel	Trade, industry and Markets, Health	www.sirc.org/
Statistics Wales	PD-PV	LC	LC Contractors	All or most of the categories above	http://wales.gov.uk/topics/statistics
The Scottish Centre for Social Research (ScotCen)	PD-PV	ODO	Parallel	Education. Social Stratification and Groupings-Inequalities	www.natcen.ac.uk/scotland/index.html
UK Parliament	PD-PV	LC	Parallel	Law, Crime and Legal Systems, Politics	http://www.parliament.uk
Women And Equality Unit	PD-PV	ODO	Parallel	Law, Crime and Legal Systems, Social Stratification and Groupings-Inequalities	http://www.womenandequalityunit.gov.uk
Work-Life Balance	PD-PV	ODO	Parallel	Trade, Industry and Markets	http://www.dti.gov.uk/work-lifebalance

V. Notes

A list of difficulties in the process of collecting information on European data:

1. Our information sources were not 'direct', i.e. we had to rely on published information on the profile of the organisations to explore their characteristics.

2. Not all websites among those visited were organized in a way that provided clear information on the activities and products of the organisations.

3. The variables we used were of a qualitative type and involved interpretation of information from various sources, to ensure that the codes assigned to them are comparable in the range of organisations studied.

4. Our access to the organisations which are relevant to data production in each country depended on the availability of information in English – (to a large extent) - and French (to a lesser extent); therefore, any websites offering information in other languages were considered 'not accessible' for the purposes of this research.

5. Due to time restrictions we did not engage in tracing data organisations from all European countries; therefore, our sample is biased and any calculations, interpretations and recommendations must be treated with caution as to their validity across all countries.

6. Identifying organisations according to a core activity, extracting their profile characteristics from generic information sources, and organizing their characteristics for comparative purposes is a very laborious process, during which false interpretations are inevitable; we are aware that this work would reveal very different results if it were taken up by the national representatives in each country and, especially, those involved with the national data production; therefore, our study of the organisations must be accounted as a "third eye" view. On the other hand this limitation has a validity for the purposes of this research because it allows us to identify the difficulties in the flow of information concerning data organisations across Europe.

7. Definitions: an important part in the process of our research was to identify those properties of the organisations which are to be used for defining common categories (typology process); therefore, common definitions were needed, even for identifying commonly used types, such as National Archives, or what we called "working definitions" such as –'parallel actors'. A separate process of acquiring consensus on definitions is certainly needed.