FP7-212214

Title	**WP6 Final report:**
	Strengthening the CESSDA RI (D6.1)
Work Package	WP6
Authors	Adrian Duşa (RODA); Jindřich Krejčí (SDA); Janez Štebe (ADP); Zoltán Fábián (TARKI); Peter Hegedus (TARKI); Brigitte Hausstein (GESIS)
Source	WP6 reports on tasks 1-11, co-operation with WP7
Dissemination Level	PU (Public)

Summary/abstract

A strong, upgraded CESSDA Research Infrastructure (RI) should be composed of both experienced, well resourced data archives and smaller institutions with less experience and often less funding. One goal of the proposed cessda-ERIC should be to ensure that participating institutions reach at least a minimal level of compliance with standards which aim to provide a high level of service to, and communication with, the academic community.

The goal of this work package was to audit existing practices and consequently propose standards and action plans to assist less developed data archives to comply with the proposed best practice in these areas. This would be facilitated by a well organised system of staff training and exchange programmes and by assessing aspects of less developed data archives against specific reference models thus enabling relevant advice and support to be offered.

Please note that any referenced CESSDA-PPP documentation should be requested from the relevant author(s)

Table of Contents

1. Introduction and rationale

Workpackage 6 dealt with strengthening the CESSDA network of data archives, basing its work around the fact that extremely large differences exist between individual data archives, both in terms of personnel numbers and (perhaps more importantly) in relation to skills, expertise, knowledge and experience with data archiving activities.

Workpackage 6 therefore had two main objectives:

1. Support capacity-building through developing the skills, knowledge and abilities of less-developed and less-resourced CESSDA organisations;
2. To foster and develop emerging CESSDA organisations through the provision of a complete 'tool kit' of standards, operational tools and expertise

Currently, CESSDA can be seen as a multinational *network* of data archiving institutions, which interact with each other in a more or less standardized way, but with very different operational procedures borne from many years of finding particular solutions to local problems.

This network is about to change dramatically in the future. CESSDA, currently a loose network, is working towards becoming a well organised pan-European infrastructure with a central hub (or a series of hubs) and related nodes. This will be similar to the organisation of a multinational company, with a series of highly standardised global procedures but with the flexibility to accommodate local/regional differences.

The future CESSDA will have to change, moving towards a common, standardised set of operational procedures, thus functioning more like a multinational company.

Obviously, the two broad objectives would not be easy to implement even in a very well organised company, less alone in a pan-European infrastructure that is still in its initial stages. However, there are a series of techniques that can be employed to assist the new RI in reaching these objectives: staff training, staff exchange programmes, employing standard procedures, using common standards, use of a common set of tools, offering the best advice to individual data archives, etc.

With this in mind, both the cessda-ERIC and its members have much to achieve both by providing advice and complying, or working toward compliance, with a common set of procedures and standards.

Workpackage 6 was divided into 11 tasks, which produced a series of focused reports. In this introductory chapter, a brief overview of each task will be presented. Chapter 2 will examine different types of best practice, as currently carried out within the CESSDA network. Chapter 3 covers the training needs for both the current and the future CESSDA RI. Chapter 4 discusses the need for standards in data archiving work, if individual institutions are to communicate effectively in a pan-European network. Chapter 5 presents some hardware and software requirements for all institutions, from minimal to optimal. Finally, chapter 6 presents some alternative solutions to web publishing for data, based on the assumption that a common CESSDA portal will continue to be used extensively by the new organisation.

In chronological order, WP6 has undertaken the following tasks:

1. Current training practices and procedures

This task was designed to survey current training practice and procedures in the CESSDA network. If a broad training plan is to be developed, then information about the current status is essential. For this reason, in the initial master questionnaire a whole section (part 4) was dedicated to training questions. Responses revealed large differences between CESSDA members, both in terms of their relative size and their training activities: some organisations have a regular and systematic programme of training whilst others have none at all. Frequently those institutions without training programmes in place send staff to larger archives to learn specific knowledge or skills.

Perhaps the most interesting finding is related not only to the training itself but to training topics. For those data archives that have training programmes it appears that no institution provides formal training for data archiving activities, instead focusing on training topics in the areas of project management, data analysis, statistical analysis and general computer skills. Whilst these activities and skills are certainly important for the data archiving community, a need for specific training on data archiving emerges as an important issue for the upgraded CESSDA.

2. Self assessment procedures

In order to plan for development in a specific direction, the surveyed data archives were invited to assess themselves comparatively with the Open Archival Information System (OAIS) reference standard.

OAIS tends to be feasible for larger organisations, yet too complex for the smaller ones. For example, it would be superfluous to talk about communication on different levels for a single-employee data archive, or about strict and rigid operational procedures for such an institution. Even though high variation exists between institutions, with respect to following procedures, all data archives are open to comply with common standards.

Whilst it would not be sensible to impose high-level procedures and standards for all institutions, regardless of their size, one key recommendation is that procedures should be staged: data archives start by fulfilling some minimal requirements and then gradually work toward fulfilling other, more complex requirements.

3. Audit of expertise

One key question, given the likely eventuality of a common training plan, is related to expertise, whether it is feasible to have internal-only CESSDA training sessions or whether to get external help as well. Just as in the previous two tasks, information was drawn from the main survey.

The findings show that most employees are social scientists or computer specialists, in many cases both at the same time (due to the nature of data archiving activities), but almost none had a formal specialisation in data archiving (this is only now becoming a profession).

While current expertise can be grouped in several ways, this task concluded with five groups: Standards (DDI, OAIS etc.), Information Technology and Network Infrastructure (NESSTAR, FEDORA, DataVerse, Dspace etc.), Data harmonization tools (DDI tools), Controlled vocabularies (ELSST, HASSET), Data protection and open access (e.g. anonymisation).

4. Recommendations concerning best practices

Less developed data archives can use the expertise of more developed ones in order to quickly develop new skills and assist the implementation of procedures. One option in order to achieve this is to examine various types of existing best practice.

The initial objective of this task was to identify such examples and subsequently find a method of dissemination them within the data archiving community. We found that best practises are highly related to local expertise, and it is not easy to decide which procedure is best, or better than others, given that each organisation presents an almost unique experience.

The preferred solution to this dilemma was to employ external references that could be used to define benchmarking criteria, in order to compare the quality of services they provide. In other words, the best procedure is the one that provides most services according to some defined reference models, e.g. Social Science Data Archives Data Activities Reference Models (SSDA DARM's.)

5 & 7. Staff Exchange and Training programme

These two tasks were dedicated to planning specific training activities for the new cessda-ERIC. These kinds of activity are highly dependent on both the future structure of the ERIC and local (national) financial support for the individual data archives.

There are three possible scenarios: the first is a centralised training programme, where everything is organised in the central hub. The second is a distributed training model, bringing specific expertise exactly to where that expertise is most needed. A third possible scenario is a mixture of both.

There is no reason to limit future training programmes to data archiving professionals only; these training programmes can target external groups: potential CESSDA member organisations; other data archiving organisations in the social science field; data archives, libraries and repositories who are not necessarily specialists in the social sciences; or even data producers and researchers working with data.

6. The CESSDA "Toolkit"

In order to establish the right tools for each particular situation, this task first identified the gaps (technical, organisational) and barriers (technical, administrative), and second identified the resources needed for the upgrade.

The gaps and barriers are: inadequate national funding; inappropriate training, lack of knowledge of standards, lack of appropriate tools for the Data Documentation Initiative (DDI); understaffed data archives, lack of IT specialists; most of the tools developed are local and therefore difficult to share; and a lack of awareness of standards and inadequate control of intra-institutional processes.

Among the resources needed for an upgrade: raising awareness of national funding agencies; use of cessda-ERIC training programmes; the planned Virtual Centre of Competence (VCC) guides; employment and training of more staff (if funding allows); use of VCC tools; assistance from expert groups and forums; employment of the cessda-ERIC minimal and optimal standards criteria; and following the assumed Levels of Service.

8. Necessary standards in data archiving

As data archiving activity is just beginning to become a formal profession, no specific standard exists for social science data archiving, so it makes sense to create a specific reference model (DARM) to serve for benchmarking and comparison purposes.

The key recommendations are: to make adherence to standards part of the new organisation's membership criteria; employ a proactive policy to promote and introduce standards in less developed data archives; and to establish a cessda-ERIC best-practice working group.

9. Recommended and alternative software

As the current situation exhibits an extreme heterogeneity, a special survey was dedicated to this particular area, in order to understand whether common activities can be found. In this respect, we have employed a series of interviews with technical personnel from 10 CESSDA member institutions, plus ICPSR from the United States.

We found that data archives employ a mixture of both commercial and open-source software, with at least one type of open-source software being used by each institution (e.g. for web purposes). Even though CESSDA members seem to prefer open-source software, sometimes there is no other option but to use commercial software (e.g. large databases).

Pre-existing structured initiatives include: DDI Foundation Tools Program; the Data Archive Technologies Alliance; Digital Curation Tools, plans to develop much awaited tools (e.g. a DDI3 editor).

10. Hardware requirements

The same series of interviews were used to harvest information about current hardware utilisation. Just as data archives differ in terms of number of employees or expertise, they also differ in terms of hardware. There are two possible scenarios: minimal standards with minimal software, and an advanced scenario with more developed and in-depth software and hardware requirements.

Proposed minimal standards for servers: at least a single server, with the use of at least four servers by more advanced data archives. Ideally separate servers should operate for: Web front-end; Data storage; Preservation; and Processing (for statistical analysis and software development). These can be installed on different machines or on virtual machines on the same physical server.

11. Publishing on the CESSDA portal

Although a very common process, this is probably the most heterogeneous of all due to the interdependence of particular software solutions which have been adopted over the years. Most of the differentiation comes from Operating Systems, software employed, and DDI production procedures. The Nesstar server is the only common tool for many data archives, but is not easy to integrate with the rest of the website. Consequently, separate parallel servers are maintained for both "traditional" and Nesstar purposes.

We recommend that the cessda-ERIC should encourage the development of different software solutions (both commercial and open-source) under a common framework of standard input-output communication standards. The upgraded CESSDA RI should encourage the development of different software solutions (both commercial and open-source), under a common framework of standard input-output communication standards.

Further, we recommend that data and metadata are kept in separate files, an open standard is used to publish to the catalogue, and that plain XML files for metadata and ASCII format for the data files are used.

2.   Best practices in data archiving across the CESSDA network

2.1. Introduction

The main goal of this task was to assess the degree of compliance with the main elements of OAIS standards and to map those areas where the OAIS model needs to be adapted to current CESSDA member practice. OAIS is the ISO reference model for Open Archival Information System.1 Each of the CESSDA member organisations has a mission to preserve and disseminate digital social science data.

Thus this chapter aims at:

- Building awareness of OAIS recommendations amongst the network members – most importantly the less developed ones;
- Raising issues that needs to be dealt with when preparing a durable future organisation, in this instance the proposed cessda-ERIC;
- Describing the available self-testing and audit tools that may be useful for designing a trustworthy digital archive that meets OAIS minimum criteria.

According to previous compliance assessment exercises the OAIS reference model is useful for social science archives (Beedham et al., Vardigan and Whiteman.) It provides a **coherent terminology**, and helps to design an organisational and functional framework for an institute that accepts responsibility for the long-term preservation of digital and non-digital research data.

The standard puts much emphasis on the separation of several functions within an OAIS archive. In reality, however, smaller archives are currently unable to comply with such organisational complexity. However, OAIS compliance would be beneficial for an upgraded CESSDA beyond the remit of the OAIS minimum requirements. Although it requires considerable resources, CESSDA network members are encouraged to use the available self-assessment tools in order to identify the strengths and weaknesses of their practice from an OAIS perspective.

2.2.   What is the ISO reference model for OAIS?

The OAIS definition of a reference model is a "framework for understanding significant relationships among the entities of some environment, and for the development of consistent standards or specifications supporting that environment. A reference model is based on a small number of unifying concepts and may be used as a basis for education and explaining standards to a non-specialist."

Beedham et al. (2005: 9) stress that the OAIS standard is technologically neutral and platform independent thus enabling OAIS to be implemented in different ways.

2.3.   What is an OAIS archive?

"An OAIS is an archive, consisting of an organization of people and systems, that has

---

1 The OAIS reference model is defined by a recommendation of the Consultative Committee for Space Data Systems (CCSDS). (cf. http://public.ccsds.org/) The recommendation then was approved of as an ISO standard later on. The full ISO standards are available at:
http://public.ccsds.org/publications/archive/650x0b1.pdf).

accepted the responsibility to preserve information and make it available for a Designated Community." (CCSDS, 2002: p. 1-1, emphasis in original) However an OAIS archive should meet a set of criteria defined in ISO reference model document and CCSDS recommendations. Long Term Preservation "is long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community. Long Term may extend indefinitely." A designated community is a group of consumers who should be able to understand the preserved information.

There are three external actors beside OAIS in the reference model: data producer, archive's management and consumer (end user).

Figure 1. Environment model of an OAIS archive
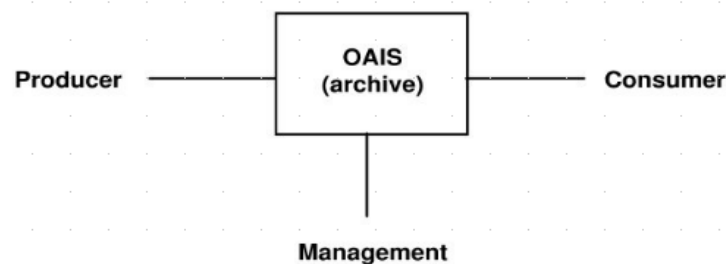
Source: CCSDS, 2002: p2-2

There are six functional entities within an organisation that is an OAIS, as shown in the next figure.

Figure 2: OAIS Functional Entities
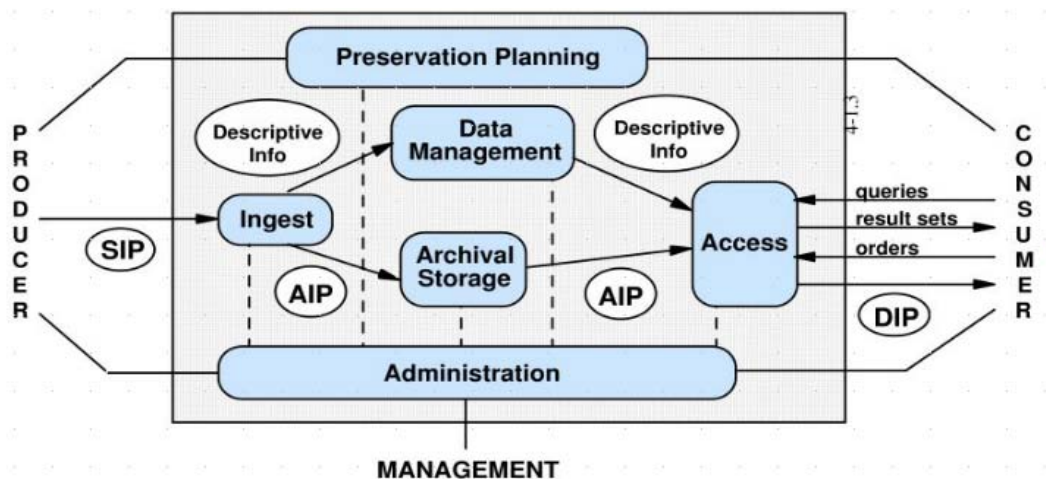
Source: CCSDS, 2002: p4-1

Beside the six functions there are common services which relate to each entity. Common services are IT support services such as operating system services, network services and security services.

Ingest includes the following operations:

- Receiving submissions in the form of Submission Information Packages (SIPs) from Producers (or from an internal organisation unit under Administration

control);
- Quality Assurance that validates the physical integrity of SIPs (e.g. CRC checks and filters for read-write errors);
- Generating one or more Archival Information Packages (AIPs) from SIPs that conform to the archive's data formatting. This typically means data file conversions and some reorganisation of data file objects;
- Generating Descriptive Information that is extracted from AIPs and other sources. This includes metadata to support searching and retrieving AIPs;
- Co-ordinating updates to *Archival Storage* and *Data Management*.

Archival Storage is responsible for the storage, maintenance and retrieval of Archival Information Packages (AIPSs). "Archival Storage functions include receiving AIPs from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the media on which archive holdings are stored, performing routine and special error checking, providing disaster recovery capabilities, and providing AIPs to Access to fulfil orders."

Data Management supplies the services and functions for populating, maintaining, and accessing both (1) descriptive information which identifies and documents archive holdings and (2) administrative data used to manage the archive. Data Management incorporates directing "the archive database functions (maintaining schema and view definitions, and referential integrity), performing database updates (loading new descriptive information or archive administrative data), performing queries on the data management data to generate result sets, and producing reports from these result sets."

Administration "provides the services and functions for the overall operation of the archive system. Administration functions include soliciting and negotiating submission agreements with Producers, auditing submissions to ensure that they meet archive standards, and maintaining configuration management of system hardware and software. It also provides system engineering functions to monitor and improve archive operations, and to inventory, report on, and migrate/update the contents of the archive. It is also responsible for establishing and maintaining archive standards and policies, providing customer support, and activating stored requests."

Preservation Planning "provides the services and functions for monitoring the environment of OAIS and providing recommendations to ensure that the information stored in the OAIS remains accessible to the Designated User Community over the long term, even if the original computing environment becomes obsolete. Preservation Planning functions include evaluating the contents of the archive and periodically recommending archival information updates to migrate current archive holdings, developing recommendations for archive standards and policies, and monitoring changes in the technology environment and in the Designated Community's service requirements and Knowledge Base. Preservation Planning also designs IP templates and provides design assistance and review to specialize these templates into SIPs and AIPs for specific submissions. Preservation Planning also develops detailed Migration plans, software prototypes and test plans to enable implementation of Administration migration goals."

Access functional entity supports consumers (end-users) in searching and locating the information stored in the OAIS repository. It allows users "to request and receive information products. Access functions include communicating with Consumers to receive requests, applying controls to limit access to specially protected information, coordinating

the execution of requests to successful completion, generating responses (Dissemination Information Packages, result sets, reports) and delivering the responses to Consumers."

OAIS terminology has become widely recognised because it provides a general terminology consisting of terms which have "not already overloaded with meaning so as to reduce conveying unintended meanings" (CRL and OCLC, 2007: p. 8.).

The OAIS recommendation provides the minimum set of criteria, which needs to be fulfilled by an archive to be qualified as an OAIS Archive.2

"The OAIS must:

- Negotiate for and accept appropriate information from information Producers;
- Obtain sufficient control of the information provided to the level needed to ensure Long-Term Preservation;
- Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided;
- Ensure that the information to be preserved is Independently Understandable to the Designated Community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information;
- Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, and which enable the information to be disseminated as authenticated copies of the original, or as traceable to the original;
- Make the preserved information available to the Designated Community." (CCSDS, 2002: p.3.1.)

2.4.   Lessons from UK Data Archive and TNA compliance assessment report

The UK Data Archive and TNA (The National Archive, UK) compared the systems and processes they have with the OAIS reference model (Beedham et al. 2005). The former organisation is a CESSDA member, and was co-ordinator of the CESSDA-PPP. Experiences from their compliance assessment - especially from the UK Data Archive's point of view - are of particular interest for the rest of the CESSDA network and we highlight their main conclusions here.

First of all they concluded that almost any organisation could meet minimum OAIS requirements at a very high, abstract level. However, a deeper investigation gave the opportunity for a critical appraisal of the OAIS model in relation to the conditions of working archives. They also emphasised that they did not go into an extremely detailed level in mapping their operations to OAIS, since such an exercise would become too time and resource intensive. The UK Data Archive and TNA experts were concerned about the lack of OAIS scalability and proposed the development of an 'OAIS Lite' which would be more useful, particularly for smaller archives.

According to their view the OAIS standard places a lot of emphasis on administration and

---

2 For a short and concise description of OAIS see the Wikipedia entry at:
http://en.wikipedia.org/wiki/Open_Archival_Information_System

management function especially on monitoring and decision-making related to preservation planning: "Whilst it is acknowledged that effective administration is essential for good archiving practice, the UKDA in particular considered that this is an area where it may be difficult for small archives to match the reference model and one where it is difficult to see where the model offers scalability for professional but less well resourced organisations. The model makes implicit assumptions about technical infrastructures (by assuming the full or partial automation of processes) when in reality the technical infrastructure of a small organisation may be extremely limited." (p. 83)

2.5. ICPSR from the point OAIS reference model

The Inter-university Consortium for Political and Social Research (ICPSR) is one of the oldest and largest repositories of digital social science data. It was founded in the United States in 1962 with the same mission as the younger European CESSDA archives. By the year of 2006 ICPSR provided access to more than 500,000 discrete data files. Vardigan and Whiteman published an analysis of OAIS compliance with respect to a subset of the model's components.

The self-assessment procedure helped them to reveal the challenges that ICPSR faced in order to full fill the key responsibilities of an OAIS-modelled archive. These two challenges were: lack of written preservation policy and incompleteness of Preservation Description Information (PDI) packages. However ICPSR practice is one of the most transparent, beside their mission statement they have published their relevant policy documents, including Collection Development Policy3, Data Enclave Policy4 and Procedure, Web Privacy Policy5 and Strategic Plan6 and other materials for consumers (Data Use Tutorial, Responsible Use Statement) and for producers (Data Deposit Guidelines) of data.

Although ICPSR employ a rather complex procedure by which they convert Submission Information Packages to Archival Information Packages Vardigan and Whiteman conclude: "Our assessment of the Preservation Description Information that ICPSR maintains is that it needs to be more robust given its centrality to the preservation enterprise. We plan to consult with experts in the digital preservation field about which fields we should add to our existing set of preservation metadata to ensure optimal coverage." (p. 79)

2.6. CESSDA online survey results

During the May and June of 2008 an online survey was conducted among CESSDA member archives regarding their modes of operations. Respondents were the Directors of archives or a designated senior staff member. We analysed responses from 18 CESSDA member organisations with respect to minimum OAIS requirements and OAIS functional entities.

The archives are very diverse with respect to their organisational structure, legal status, funding sources and to the size of their holdings. The surveyed members together preserve and distribute more than 25,000 datasets.

---

3 . http://www.icpsr.umich.edu/ICPSR/org/policies/colldev.html

4 http://www.icpsr.umich.edu/ICPSR/org/policies/enclave.pdf

5 http://www.icpsr.umich.edu/ICPSR/org/policies/privacy.html

6 http://www.icpsr.umich.edu/ICPSR/strategic/

Ingest (acquisition)
CESSDA archives typically preserve and distribute empirical social science data that come from quantitative studies, although some of the archives are also deal with qualitative data as well. For example, the UK Data Archive, the service provider of ESDS (Economic and Social Data Service) is responsible for distributing ESDS qualitative data (cf. http://www.esds.ac.uk/qualidata/). The Finnish Social Science Data Archive (FSD) holds more than 70 qualitative datasets, which can be searched and browsed in a separate catalogue.

Submission Information Packages (SIPs)
Depositors provide their data in standard statistical formats (e.g. SPSS, SAS, NSD-Stat) for most of the surveyed archives. The percentage of non-standard formats is usually below 30% with some notable exceptions. The UK Data Archive ingests: "ESDS datasets which include qualitative data files and the History Data Service files, which are much more diverse in terms of file formats." (direct quote from the UK Data Archive's answer to our online survey question.) Approximately 62 % of data collections were ingested in non-standard statistical formats by the UK Data Archive. The Spanish CESSDA member, ARCES-CIS, receives all files in text format, but they also serve SPSS definition files (variable and value labels) to end users as well.7

Archival storage
The most common format of archival storage is CSV or other row text format. Eight out of the eighteen archives surveyed use this for their storage. We can add here XML (Extensible Markup Language) format as well, which is also a platform independent, non-proprietary, plain text format, but much more suitable for sharing structured and encoded data – including metadata – via the World Wide Web. The DDI8 format, which is commonly used by CESSDA member archives, is also based on XML. However, the ability to store data in the XML has been introduced in DDI version 3.0. Previous versions handled only metadata, i.e. descriptive information on the research datasets. Three archives indicated XML as the standard preservation format at their archive: RODA, the UK Data Archive and NSD.

Nine archives use the SPSS system (SAV) and/or portable format for their archival storage. Probably most of the designated users community are familiar with the SPSS statistical application. SPSS (and other standard tools like Stata) features in the curriculum at many universities that teaches empirical social research. So it is an obvious that SPSS format should be a user option in dissemination information packages.

Access
As information technology has evolved, the Internet became the most common channel for accessing digital social research data. All CESSDA member archives have their own homepage on the World Wide Web, and they offer various data services, including direct access to microdata files or online statistical analysis via Nesstar software system. With few exceptions (CEPS/INSTEAD and CNRS-RQ) all archives have English language webpages as well.

Nesstar (http://www.nesstar.com/) was developed during the Networked Social Science

---

7 See English request form on CIS website: http://www.cis.es/cis/opencms/EN/6_arces/impreso.html.

8 http://www.icpsr.umich.edu/DDI/

Tools and Resources (NESSTAR) project. It was funded during 1998-2000 by DGXIII of the European Commission under the 4th Framework Telematics Applications Program. Currently Nesstar is owned and maintained by the Norwegian CESSDA member organisation, NSD.

Besides the homepages of individual national archives, the CESSDA site (http://www.cessda.org/ offers access to the CESSDA catalogue (http://www.cessda.org/accessing/catalogue/). Although not all the CESSDA archives have joined the integrated catalogue, the importance of this access channel obviously will become increasingly apparent as the construction of the proposed cessda-ERIC gets underway.

2.7. Further possibilities for self-assessment audits

UKDA and TNA OAIS compliance self-testing
Beedham et al. (2005) completed the first systematic assessment of a CESSDA member archive (namely the UK Data Archive) to assess compliance with the OAIS reference model in their institutes. They have suggested an extensive list of questions for OAIS compliance self-testing and we used these when formulating our CESSDA member survey. However there were obvious constraints on the number and type of questions that could be asked in a joint online survey. Moreover, some of our proposed questions have been reformulated or left out during the finalisation of the online survey.

Trustworthy Repositories Audit & Certification (TRAC)
TRAC is a checklist developed from work done by the OCLC/RLG Programs and National Archives and Records Administration (NARA) task force initiative. The checklist is available at: http://www.crl.edu/PDF/trac.pdf.

Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)
The Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE) announced the release of the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) toolkit. This toolkit is intended to facilitate internal audit by providing repository administrators with a means to assess their repository's capabilities, identify their weaknesses, and recognise their strengths. Digital repositories are still in their infancy and this model is designed to be responsive to the rapidly developing landscape. The development of the toolkit follows a concentrated period of repository pilot audits undertaken by the DCC, conducted at a diverse range of organisations including national libraries, scientific data centres and cultural and heritage data archives.

Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE). February 2007, "DCC and DPE Digital Repository Audit Method Based on Risk Assessment, v1.0. retrieved 26.10.2008, from http://www.repositoryaudit.eu/download
Cf. http://www.repositoryaudit.eu/
http://www.dcc.ac.uk/
http://www.digitalpreservationeurope.eu/

NESTOR - Network of Expertise in long-term STORage: Working Group on Trusted Repositories Certification
The referenced NESTOR document "identifies criteria which facilitate the evaluation of digital repository trustworthiness, both at organisational and technical levels. The criteria are defined in close collaboration with a wide range of different memory organisations,

information producers, experts and other interested parties. This open approach is the basis for achieving a high degree of universal validity and practical applicability and facilitates broad-based acceptance of the results of any evaluations conducted on the basis of these criteria. The present criteria catalogue for public comment represents an important milestone on the road towards achieving the working group's goals. The memory organisations should be given a well-constructed, coordinated and practical tool for achieving and demonstrating their trustworthiness.

However, the intention is also to present the opportunity for repository certification within a standardised national or international process as a formal endorsement of an organisation's trustworthiness."

NESTOR Working Group on Trusted Repositories Certification: Catalogue of Criteria for Trusted Digital Repositories, Version 1 (draft for public comment), June 2006, Frankfurt am Main: nestor c/o Deutsche Nationalbibliothek, urn:nbn:de:0008-2006060710
Cf. http://www.nbn-resolving.de?urn:nbn:de:0008-2006060703.
NESTOR website: http://www.langzeitarchivierung.de/.
The NESTOR catalogue of 54 criteria covers three broad areas: (1) organisational framework, (2) object management and (3) infrastructure and security. These should be evaluated not on an absolute basis, but on the basis of the goals of the digital repository concerned. There are five dimensions of evaluation. Four of them about completion: (1) conceptual groundwork, (2) plan or specification, (3) implementation and (4) evaluation. The other, (5) relates to the publication of appropriate documentation relating to an archive's functions, which helps to increase the transparency of operation and therefore generates confidence and trustworthiness.

Data Seal of Approval (DSA)
Based on NESTOR (2006), DRAMBORA (DCC and DPE, 2007), TRAC (CRL and OCLC, 2007) and other initiatives DANS (Data Archiving and Networked Services) prepared a minimum set of criteria for digital research data. DANS is a social science archive and member in CESSDA network.

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

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

The implementation of these criteria is similar to the OAIS reference model and specific responsibilities are as follows:

- "The *data producer* is responsible for the quality of the digital research data.
- The *data repository* is responsible for the quality of storage and availability of the data: data management.
- The *data consumer* is responsible for the quality of use of the digital research data."

2.8. Recommendations

Recommendation	Tool	Effort estimate (in person months = PM)
CESSDA members should review their operations by mapping their practice to the OAIS minimum requirements.	Training of internal and external auditors; Use of the available self-assessment tools.	6 PM per organisation 12 PM of a dedicated Expert group member
One of the most important point from OAIS context the selection of data preservation formats. Platform (OS) independent, open source formats should be preferred such as XML, or the XML-based DDI to proprietary, application- or operation system-dependent formats.	Bench processing conversion tool	Conversion: 3-6 PM per organisation
Submission information packages should be preserved and the changes, conversions that applied during archival processes should be documented ("version history")	Content management system selected	Content management system installed and populated: 3 PM per organisation
Long-term preservation of social research data and distribution of preserved data collections are the two most important goal of a social science data archive. In reality these aims can be conflicting. While the former requires a platform-independent preservation formats, members of the end-user community often require other, proprietary formats. Archives should provide those elements in dissemination information packages that make data collections easily accessible and usable for end-users. From the perspective of the OAIS standard we can emphasise here the difference between Archival Information Packages (AIPs) and Dissemination Information Packages (DIPs).	Identification and versioning issues resolved	Identification, access and versioning of DIP: 6 PM
Social science archives should make their operations more transparent to wide circles of stakeholders: including producers, depositors and consumers	• Standard Policies Tool assistant; • Web	6 PM per organisation

(end-users). Archives should make public their: • mission statements and strategic plans; • targeted designated user communities; • access rules and modes of access; • handling of intellectual property such as copyrights; • practice of ensuring the protection of personal data (i.e., how do they anonymise data collections, how do they handle the personal data of consumers).	publishing content management system.	

2.9. Audit of expertise

The main sources of data and information in this section are the online survey of CESSDA Directors and archive homepages. The Director's survey asked for general organisational information and questions related to the CESSDA-PPP workpackages, with five main topics investigated:

1. Distribution of part and full-time data professionals;
2. Higher education;
3. Specialisation of data professionals employed at CESSDA member archives;
4. Number of the employees with skills in statistical software;
5. Data professionals who have received systematic, job-related training.

The results of this section have been used to estimate the potential of CESSDA member organisations to supply experienced staff for future expert groups. Proposals for training strategies are also dependent on this kind of survey information – the amount and variability of expertise in organisations across the CESSDA network is important for the calculation of supply and demand of professional knowledge.

2.10. The distribution of the part and full-time data professionals

Figure 1 shows the four categories of CESSDA member organisation's staff: full-time data professionals, full-time other staff, part-time data professionals and part-time other staff. Data professionals make up the smallest section of employee type. This indicates that there is room for further professional profile shaping of organisations by increasing the size and status of this category of staff members.

Figure 2. The percentage of the part and full-time data professionals in the CESSDA archives (N=16)
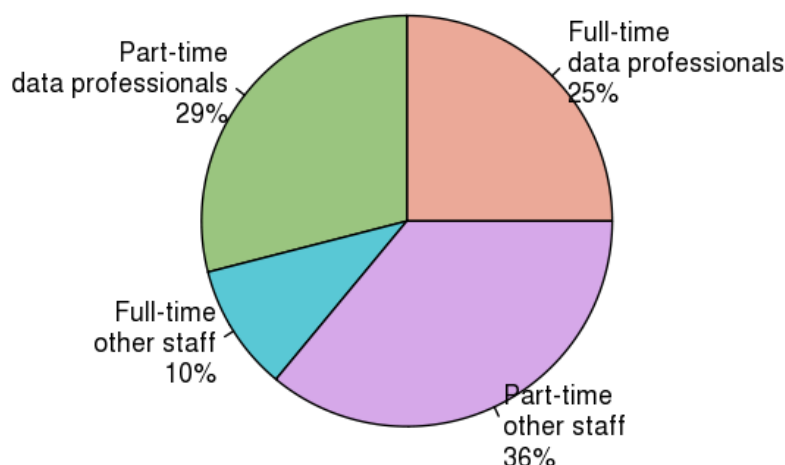
Source: CESSDA Survey, Question 9

2.11. Education level

All archives surveyed have staff with master's degrees and 50% of staff in these organisations have been awarded an MA. Staff with doctorates were the second largest education level category at 17%.

These results support proposal 1) of section 3.9 regarding the self-sustaining capacity of CESSDA member organisations to supply the proposed training programmes.

2.12. The specialisation of data professionals employed in CESSDA archives

The report of Key Perspectives Ltd. in relation to career structure of the data scientists distinguishes four roles: data producer, data scientist, data manager and data librarian.9 The data producers are researchers with domain expertise who produce data, the data scientist enables others to work with digital data, and developments in data base technology. The data manager refers to computer scientists, information technologists or information scientists.

In our questionnaire six roles (i.e. archivists, statisticians, survey research methodologists, sociologists, other social scientists) refer to data scientists. Data managers are described as either information technologists or information scientists.

Most archives employ sociologists or other social scientists, and statistician and survey research methodologist are the next most popular professions. Other important professions include information scientists - almost every archive has information technologists or scientists. On the other hand less than half of the archives employ librarians, archivists or human scientists.

In smaller archives the most frequently occurring specialist is that of sociologist and these organisations tend not to currently employ human or information scientists. In medium and large organisations we found almost every category of specialisation (Figure 4).

---

9 Alma Swan and Sheridan Brown: *The skills, role and career structure of data scientists and curators: an assessment of current practice and future needs*. Key Perspectives Ltd. Truro, 2008.

Figure 3. The specialisation of data professionals employed by CESSDA member archives relative to organisation size (N=17)
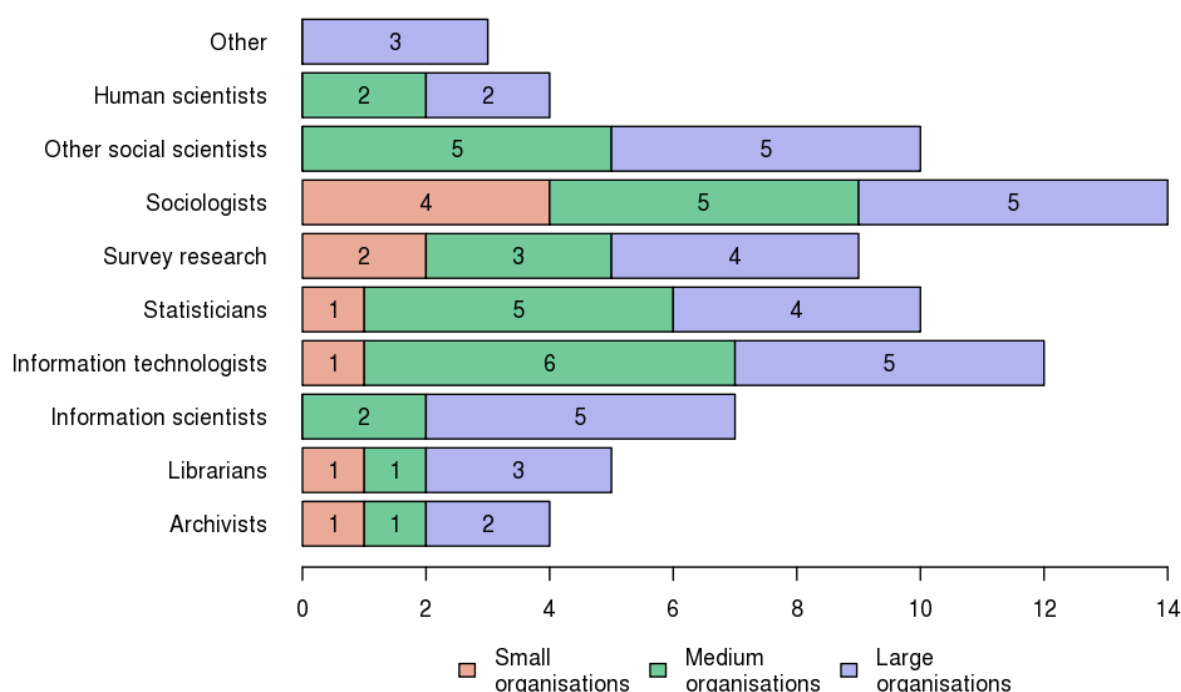
Source: CESSDA Survey, Question 13

Unequal distribution of expertise supports the conclusion that co-operation and staff exchange is needed to balance the professional background structure among organisations as part of working toward more co-ordinated and balanced service provision.

2.13. The data professionals who have received systematic, job-related training

When examining the frequency of job related training, it is obvious that data professionals employed by large organisations often have more training opportunities. This is supported by the results of our survey (see figure below) which showed that whereas just over 50% of data professionals at smaller organisations have received training within the last year, 100% of this type of staff member at organisations classified as "large" have received training within this period. .

Figure 4. The percentage of the data professionals who have received systematic, job-related training (internal, external or both) in the last year (%, N=16)
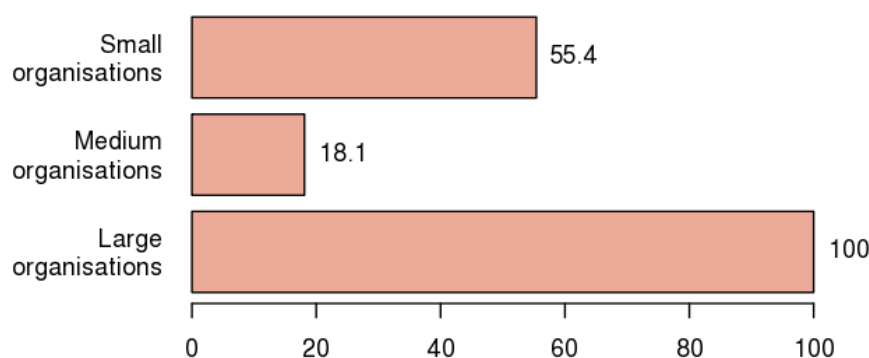
Source: CESSDA Survey, Question 16

These results support Chapter 3's recommendation that a common CESSDA training programme will both make best use of the expertise within the CESSDA membership and work toward ensuring that all member organisations have equal access to training opportunities.

2.14. Proposed expert groups and subgroups

We investigated the variety of professional knowledge and skills of CESSDA member archive employees. These can be categorised in five ways, with each expert group having a number of subgroups:

1. Standards
   DDI
   OAIS
   Other (metadata)
2. Information Technology and Network Infrastructure
   NESSTAR
   FEDORA10
   Dataverse
   DSpace11
   Intranet
3. Data harmonisation tools
   ISCO (International Standard Classification of Occupations)
   ISCED (International Standard Classification of Education)
   International comparative research
   Harmonisation and conversion
   Question Data Bank
4. Controlled vocabularies
   ELSST
   HASSET
5. Data protection and open access
   The Data Protection Act and the anonymisation process
   Anonymisation
   Open Access to Data

We defined the scope of the proposed expert groups as follows:

Expert Group on Standards: Members of this group should have expertise in recognised international standards on data archiving and dissemination, particularly within the social sciences (OAIS, DDI, other including METS). This group should be responsible for assessing the theoretical basis of suggested standards, and make recommendations for their practical implementation, This expert group should also be responsible for assessing data management tools and data description.

Expert Group on Information Technology and Network Infrastructure: This group should comprise of experts with knowledge and experience of Web Content Management Systems

---

10 Fedora Commons is a non-profit organisation providing sustainable technologies to create, manage, publish, share and preserve digital content by bringing two communities together. See also: http://www.fedora-commons.org/

11 DSpace is the software for academic, non-profit, and commercial organisations building open digital repositories. http://www.dspace.org/

(WCMSs) in general and those specifically developed for data archives/digital repositories. WCMS software tools enable data curators and data managers to create metadata descriptions which are compliant to specific standards, manage digital objects, define different access levels and grant access rights to users and groups of users, etc. There are widely known applications like NESSTAR, Dataverse etc. that can act as examples of such tools. However, currently, there is no single application which provides all functionalities required to meet the OAIS compliant digital repository standard.

Expert Group on Data Harmonisation: There are several specific challenges that all large scale cross-national comparative social research organisations face: sampling, questionnaire design and translation, etc. The process of data harmonisation takes place after the completion of survey field work and it is used to standardise key variables so that trans-national comparison between data is possible. Examples for routine data harmonisation tasks are: converting national occupation codes to international standard codes (e.g. ISCO) or to standard occupational prestige score, harmonising income, and educational variables, attitude scales.

Expert Group on Controlled Vocabularies:
"Controlled vocabularies provide a way to organise knowledge for subsequent retrieval." (http://en.wikipedia.org/wiki/Controlled_vocabulary) They are used for indexing subject keywords ("tags"), creating, maintaining thesauri and taxonomies, by which an end-user of a database can find information more easily. There are several controlled vocabularies in the area of social science research, some of the most important include: CESSDA keywords, HASSET (Humanities and Social Science Electronic Thesaurus: the English language source terms for ELSST), and ELSST (European Language Social Science Thesaurus).

Expert group on data protection and legal issues: Data archives have to balance their provision of access to research data whilst taking into account legislation relating to personal information and intellectual property. There are various national legal frameworks and European level regulation for the protection of personal (sensitive) data.[12] It is best practice for data archives to anonymise their data collections after ingest and they should provide contractual conditions for end users in which they set the rules for appropriate, adequate use of data collections. "When data is anonymised, ideally both privacy and information utility levels should be maximised." (Zielinski, 2007) [13] This expert group should also represent expertise on open access issues and initiatives like the Science Commons project.[14]

2.15. Best practice
This section addresses the problem of mapping data archiving activities according to highest quality requirements that the future CESSDA is proposing to fulfil. Proposed full membership criteria include: "To adhere to the OAIS reference model and/or the agreed …seal of approval for archival practices." (CESSDA-PPP Management Board 2009, Annex 1, bullet 1.2.h).

---

12 There is an EU Directive (95/46/EC) on data protection. See the respective European law and practice at: http://ec.europa.eu/justice_home/fsj/privacy/index_en.htm.

13 . Zielinski, Marek. 2007. "Balancing Privacy and Information Utility in Data Anonymisation" (http://www.eife-l.org/publications/eportfolio/proceedings2/ep2007/papers/digital-identity-and-privacy/balancing-privacy-and-information-utility-in-data-anonymisation-1)

14 See http://sciencecommons.org

This issue has wider relevance. By defining best practice member organisations are able not only measure their progress in fulfilling specific criteria (benchmarking, certification), but are also able to effectively facilitate exchange of practices, propose specifically focused training programs, and upgrade associate/affiliate members to become full members of the new organisation. This is the result of a more transparent classification of activities by organisations with clearer definitions, examples, cross references to external standards. It should also result in new models and tools which can be developed to support existing an emerging best practice.

What do we mean by best practice? In the current context our working definition of best practice is:

Written knowledge that is currently agreed upon in a community of practice (e.g. professional community of people working in data archives and others similar institutions) as that which guarantees the highest and most efficient achievement of an organisation's goals (e.g. services provided).

Whilst we can start from existing sources, delivered internally in organisations or published, that contain aspects of current and past practice enacted by Social Science Data Archives, these are not definitive in a sense of the best practice. By comparing them in a clear fashion, that is, by delineating activity by activity, providing consistent terminology for those activities, and relating them to some external, already professionally approved best practice reference model, one can come closer to the definite version of best practice for certain activities. Additionally, agreement of current and past practices can be used as an indication of qualification for best practice. One can spot holes in both current documentation of best practices and criterion reference models. Those reference models that spring from other traditions, or whose aim is other than best practice criteria, might be improved or replaced with new models which synthesise knowledge contained in existing models and current practice source documents.

Highest quality requirements are not currently defined. Whilst mapping of current practices and reference models could provide terminology for, and description of, set activities that are contained in professional organisation practice, this will also require the additional grading of certain activities to a level of performance reflective of the organisation concerned.   Some of the requirements could be necessary, others optional, dependent on the specific profile of each members organisation. These demarcations should be delineated carefully.

Additionally there is a need to take into consideration the size of the organisation, e.g. number of staff, organisational complexity, number of studies processed, type of material and volume, etc.)

2.16.   Mapping of data archiving functions to Reference models
2.16.1.   General purpose mapping
The above mentioned Social Science Data Archive current practice source documents need to be mapped individually to each of the examined Data Activity Reference Models (DARM's).

By doing that one can identify:
- typical examples of each principle (For each activity one can find out a variety

of source examples; one can select those that are typical in some aspect, e.g. in relation to how to deal with a specific type of material, qualitative, quantitative etc.)

- a dictionary for traditional designation of processes and tasks in a data archive;

- holes in both reference models (for processes that are contained in sources, but have no relation to reference models) and for current practice sources;

- ways of synthesising knowledge that is contained in sources, that is otherwise difficult to compare, one by one related to particular reference activity in DARM

- methods of mapping knowledge developed in the 50 years existence of Social Science Data Archives (see, e.g. past guidelines and best practice documents) and current DARMs, can be preserved and made accessible to future professionals

- Communication needed with other professional communities for exchange and review the best practice DARM. The results of this could also serve as input to more formal professional training, curriculum building and planning, etc.

One outcome of such an exercise would be a detailed level mapping between reference models. This would show where there is duplication and where are unique assignments of activities. The end result would be a collection of detailed level activities, from which a comprehensive Social Science Data Archive DARM could be extracted. This should include best practice examples taken from Social Science Data Archives and supplemented with additional considerations.

2.16.2.   A Social Science Data Archives DARM as dictionary for specific purpose mapping

A comprehensive Social Science Data Archive DARM dictionary could be used for specific purposes planning:

- Assignment of roles to activities for human resource planning;
- Consequently, required specific expertise could be accessed, which would assist in planning training programmes;
- Tools for support could be mapped to specific activities. Therefore both existing and coordinated planning of future tools could be effectively achieved;
- Cost and human resource estimates could be made for upgrades, when implementing a model compatible to CET DARM;
- Level of service reporting and benchmarking.

2.16.3.   Conclusions

The aim of this report was to explore the possibility of having a general ontology (dictionary, terminology, reference model) that the current and best practice of existing Social Science Data Archives could be mapped to. One obvious use of such a model would be for benchmarking and certification purposes. Furthermore, such a model could facilitate the planning of changes and adaptation of current practices to better accommodate users needs and simultaneously fulfil its overall aims.

We firstly addressed whether existing models were suitable for our requirements. This was done by examining two established traditions of describing Data Archiving activities: Digital Preservation and an analogy of a Data Life Cycle. We noted that there was much overlap between these two approaches, alongside strengths specific to each model. Additionally, we also noted that additional testing of these approaches was necessary.

Neither the OAIS model per se, e.g. its derivations in TRACK, DSA, DRAMBORA for the tradition of digital preservation, or DARM's taken from data life cycle approaches such as LIFE 2, CET DARM etc. fully suit our requirements. Instead, in order to assess the relevance of traditional reference models, these should be compared individually with existing guidelines and public documents on the routine daily activities of Social Science Data Archives. The outcome of such a comparison will highlight remaining and specific processes that are not part of the general model (e.g. pre-ingest activities poorly elaborated in OAIS). It is our assumption, based on available descriptions of traditional reference models, that as a result a unified Social Science Data Archive reference model use could be extracted, with explicit relations to both traditional DARMs and current documents describing Social Science Data Archiving activities. This is both the basis and approach for our proposal.

This approach would result in a unified terminology, based on, and with clear reference to, equivalent meanings in other traditional models. It would facilitate internal organisation of activities according to the requirements contained in a model, and external exchange of knowledge between similarly oriented organisations. By choosing a reference model for certification and mapping one can assign the general requirements of a model to a specific organisation and suggest improvements or measure its adherence to wider principles beyond just those used as a basis for the reference model.

Additionally, specific mappings could be used by the Social Science Data Archive (SSDA) DARM, e.g. for activities relating to an organisation's personnel and human resource needs or training requirements; the level of service and reporting; the methods and mechanisms for assessing service quality and end products in relation to user need. Costs and human resources estimates could be extracted based on comparison with some traditional DARM models.

This could also be used as a conceptual source for planning purpose. Blue guides (it is not currently used for this purpose, but there is potential), planning tools etc. that follow those principles. Tools, legal documents, relations between internal and external stakeholders and mutual obligations could be extracted, and informed by the knowledge base of a SSDA DARM, that could support planning and introduction of best practice procedures into the daily activities of related organisations. Harmonisation of activities could be achieved, and an informed decisions taken regarding the potential efficiency of having some common services, e.g. a common preservation service. One needs to take into consideration aspects of harmonisation required at a CESSDA level to enable common services and increase overall efficiency. This type of DARM could also be used as a criterion for classification and for the selection of best practice examples from current practice descriptions. Both would be nearly impossible without unified terminology. These examples could then feed into the best practice guides.

An SSDA DARM and its associated guides, tools and documents needs careful consideration in order to avoid being too prescriptive and to enable flexibility in accommodating current conditions in associated/related organisations. Size, specialisation

and the specific profile of existing organisations should be allowed to vary, thus retaining relevance to local clientele and local service conditions. What level of adherence to "best practice" solutions should be demanded for full members: a detailed adherence to 'data pipeline' models or a set of principles formulated in abstract terms. A detailed description of local processes is needed anyway (as a requirement) to increase transparency, and evaluation, and this is useful also from the interests of knowledge sharing (previously CESSDA Expert Seminars served this need). Examples include: "Preservation policy", Strategy, Guide, etc.; Users, producer agreements; staff and data user agreements in handling data in a restricted environment. The existence and adherence to such guidelines helps facilitate procedural transparency and thus may help build trust in these processes. Revisions of procedures could be better managed; versions of procedures could be stored and linked to data objects, so as to add to the consistency of AIP. Rights management could also be organised with reference to existing legal and supporting standards requirements (Intellectual Property Rights and confidentiality).

We conclude that it is preferable to support the introduction of best practice and procedural change rather than police organisations with certification set-ups. By taking this proactive approach both long standing, highly regarded organisations and those that are currently poorly resourced, can benefit.

Recommendation	Tool, stakeholder	Effort estimate (PM = person months)
1. Engage in designing a first version of a Social Science Data Archives DARM and sustain its further development	Establish dedicated cessda-ERIC best practice working group consisting of experts on digital preservation and data activities planning from a varied set of organisations.	24 PM
2. Make a proposal of quality requirements, mapped to the proposed DARM, which contains graded levels of service. This should set minimum requirements for certain type of organisation, balancing the needs and conditions of a variety of organisations in order to maximise the general quality of service provision to users in line with cost and funding opportunities.	-- " --	12 PM
3. General requirements should be clear and simple, and adaptable to a variety of settings. They should, however, still aim to deliver a high level of service and quality product.	-- " --	12 PM
4. Instead of lengthy and resource intensive digital preservation certification the digital preservation expert group should prepare an advisory procedure that assists in realistic self-assessment and facilitates improvement in activities where weaknesses are found. This process should also estimate the cost of improvements.	-- " --	12 PM

5. Preparation of a harmonised set of measuring methods and indicators of LS service activities related to the above mapping.	-- " --	12 PM
6. Use LS for benchmarking purposes.	LS tool	6 PM per organisation
7. Monitor user demand and enable responsiveness both through a variety of users groups and by the introduction of specific services and products.	DARM based focus group exercises	6 PM
8. Maximise data access and minimise barriers by introducing and promoting open access arrangements, e.g. by introducing data specific Creative Commons licences, and shared responsibility models.	Standard licence and access arrangements	6 PM per organisation
9. Introduce an additional layer of best practice principles to accommodate the federated services. Divide the requirements between the central facility and nodes.	Establish dedicated best practice working group consisting of experts on digital preservation and data activity planning from a variety of organisations.	6 PM
10. Facilitate the adoption of minimum requirements by producing a CESSDA toolkit, that is set of templates of strategic documents, contractual forms, etc. and generic guidelines, use cases, best practice solution scenarios, automated tools, suggested metadata models (e.g. based on PREMIS), etc. This toolkit should both emulate and collaborate with related DARMs from other traditions and fields of practice (See Green 2008). For example, a specific SSDA purpose Standard Policies Tool Assistant for defining organisational policy in access, etc. that is ready for tailoring specific requirements, should be included in list of general requirements.	Dedicated CESSDA ERI best practice working group consisting of experts on digital preservation and data activities planning from a varied set of organisations.	12 PM
11. Design a generic list of template documents suited to the general requirements of a SSDA.	--- " ---	3 PM
12. Support organisations in preparing, documenting, publishing and implementing changes in data activities.	Mentoring service	24 PM
13. Specifically consider the needs of medium and small organisations, who often have less complex organisational structures and governance models, and smaller throughput. Most traditional DARMs are based on big volume. Suggest alternative tools (e.g. explore the potential and facilitate use of open source software such as Fedora), ready made contractual models, etc. and organisational structures to accommodate this type of organisation to spare resources while still achieving minimum requirements. Local federation of nodes could be sought to facilitate the organisation of specialist and	Tools alliance contribution	24 PM

common services between medium and small organisations (e.g. LOCKSS).		
14. Assist in the professional development of archive employees by providing training and online educational materials (e.g. wikis, news, source references, legislation, standards, tools, etc.) whilst balancing the need for general expertise in data service and preservation with specific subject expertise, which assists both in understanding user needs and translating that demand into future product/resource development.	Expert groups	6 PM per organisation

3. Towards a common CESSDA training strategy

3.1. Introduction

Although planning training issues is primarily focused on dealing with any skills gap between staff at well-resourced archives and those at developing institutions, there are wider issues at stake. The upgrade of the CESSDA RI aims at a continuing high level of professionalism across all its member organisations, involving the systematic development of human resources including the continual updating and improvement of professional knowledge and competence.

There were three tasks devoted to training issues within WP6:

- Task 1. Audit of training opportunities;
- Task 5. CESSDA programme of staff exchanges;
- Task 8. Recommendations for a Common CESSDA strategy for training issues.

Chapter three summarises the major findings and recommendations of these tasks and describes the subsequent strategic plan for the establishment of the proposed cessda-ERIC Training Programme. More detailed information related to these summaries can be found in the following two reports:

- Current training practices and training opportunities for CESSDA; CESSDA-PPP interim report from task 1 (Krejci and Cizek 2008);
- Establishing a CESSDA Training Programme: Challenges and possible strategies; CESSDA-PPP interim report from tasks 5 and 7 (Krejci, Hausstein and Dusa 2009).

3.2. Current training practices and training opportunities
3.2.1. Method and data

An audit of existing training practices and training opportunities was carried out as part of task 1 and based on data obtained from: (1) the CESSDA Web based questionnaire survey of data organisations; (2) an extensive survey of CESSDA member archive websites and other data organisations with relevant missions. The full results of this audit are available in the interim report from task 1 (Krejci and Cizek 2008) and its accompanying, but separate, Appendix 3 which includes a detailed overview of existing training opportunities. For the purpose of this analysis all current CESSDA organisations were divided into four groups according to number of data professionals employed, and an organisation's categorisation within the current CESSDA membership fee structure (see Table 1).

Table 1.Division of CESSDA organisations into groups according to number of data professional staff and current CESSDA membership fee structure

Group	Archive	No of data professionals (F/T = Full Time; P/T = Part Time)
XL Very large organisations	NSD	64 = 61 F/T and 3 P/T
	UKDA	55 = 28 F/T and 27 P/T
	GESIS	29 data professionals*
L Large organisations	DDA	20 = 17 F/T and 3 P/T
	FSD	17 = 12 F/T and 5 P/T
	DANS	22 employees + 3 management + 4 external*
M Medium size organisations	CNRS-RQ	9 = 6 F/T and 3 P/T
	FORS	7 = 3 F/T and 4 P/T
	SND	7 = 6 F/T and 1 P/T
	EKKE-GSDB	6 = 4 F/T and 2 P/T
	CEPS	5 F/T
	ARCES-CIS	N/A
S Small organisations	TARKI	4 = 2 F/T and 2 P/T
	ADPSS	3 = 2 F/T and 1 P/T
	SDA	5 = 2 F/T and 3 P/T
	ADP	2 = 1 F/T and 1 P/T
	ISSDA	1 F/T
	RODA	1 F/T

Source: CESSDA Survey (May-June 2008), GESIS, DANS.
Note: * = information taken from the Internet

The priority target group for intended training and exchange programmes are data professionals at CESSDA member archives. We define 'data professional(s)' as staff participating in the management of an archive, or engaged in data management, archiving, or data-related information systems.

3.2.2. Current training practices at CESSDA organisations

Survey results show that CESSDA member archives have strong connections to social research practice through the particular specialisation of their employees. Most archives employ social scientists, statisticians and survey research methodologists, while specialists in archiving and librarians were found infrequently. Other important professionals, such as internet technology specialists, were often missing from small archives. More than half of archives surveyed ensured expertise through outsourcing/contracting out work.

There are large differences in training practices throughout CESSDA member organisations. Several organisations referred to regular and systematic training; at the same time no job related training took place during 2006-7 for at least four CESSDA member archives. Often the content of proposed training does not correspond to the supposed structure of knowledge management at social science data archives. A large part of available training does not directly relate to archival skills, but focuses on data analysis

methods, statistical software and more general computer skills. This kind of training, though very valuable for archival work, cannot fully compensate for the current lack of training in professional archival skills.

3.2.3. Existing training opportunities

Available training opportunities in data management and archiving skills are unsatisfactory in terms of building the high level of professionalism intended for an upgraded CESSDA RI. There are only four continual programmes providing at least some training targeted directly to social science data management and archiving. Their training methods scope, areas of training, timing and capacity is limited. Three different areas of specialised training relevant to CESSDA's needs can be identified:

Training in skills related to social science research based on data analysis, e.g. methods of data analysis, statistics, statistical and other relevant software, etc. Training in this area is widely available in various forms and types, including courses within university programmes. Relevant courses are often also provided by CESSDA member institutions.

Training in general digital preservation skills, e.g. courses in long-term digital data preservation, digital repository standards, assessment methods, data digitalisation, etc.: The availability of training in this area is limited with regard to the number of opportunities, capacity, and visibility. At the same time there are important activities geared toward changing this situation and rapid developments are expected.

Training in specific social science data archiving skills, e.g. specialised courses of metadata production, DDI, NESSTAR, specialised controlled vocabularies and thesauri, data harmonisation, etc.: There are only four established programmes15 continually providing at least some training targeted directly to social data management and archiving. However, their scope regarding areas of training, timing, and capacity is limited.

3.2.4. Requirements and priorities at CESSDA organisations

Most CESSDA organisations expect that reaching and maintaining appropriate standards and implementing new technologies in connection with the CESSDA RI upgrade would require new skills to be developed beyond their existing training capacities. In this respect six of the seventeen responding organisations gave highest priority to improvement of professional knowledge and training practices.

Priority areas of required training as listed by survey respondents significantly differ from the focus of current training practices and existing training opportunities. The most frequently mentioned topics relate to metadata production and standards, long-term digital data preservation and specialised tools for presentation and data dissemination.

Analysis of usability of methods of training shows that all methods of training suggested for the Training Programme meet proposed user requirements. At the same time the idea of organising more CESSDA expert workshops proved less popular than other suggestions. Some small archives are not sure if they would be able to use staff exchange programmes. Individual visits appeared to be problematic for larger archives, which often act as host institutions to non-CESSDA member visitors. However, a preference was expressed for structured and systematic programs instead of individual visits.

---

15 CESSDA Expert Seminars, IASSIST workshops organised within annual conferences, GESIS-ZA Eurolab programme, ICPSR training programme in digital preservation.

3.2.5. Capacities for organisation of training

Most of the current CESSDA member organisations (including smaller archives/institutions) have the capacity to be not only recipients, but also contributors to organising the proposed Training Programme. Nearly all CESSDA member organisations have experience of organising training activities. Some would also be able to contribute by provision of some technical background and access to their training facilities. An upgraded CESSDA RI could also take advantage of established extensive training programmes and large training centres at several member archives.

Co-operation with organisations and projects with similar objectives to CESSDA has the potential to make CESSDA training activities more efficient and help in filling identified gaps within the CESSDA network.

3.3. Background to establishing the cessda-ERIC Training Programme

Current training practices at many CESSDA organisations are insufficient with respect to current and proposed levels of demand for the systematic and continuous development of professional competence. There are relatively few CESSDA member organisations which already have existing external training opportunities specific to social science data archives' skills and knowledge. At the same time organising separate, extensive continuous training programmes for a limited number of people at each organisation would be inefficient and often impossible. That is why we propose establishing a structured and centrally driven common Training Programme, with participation a crucial component in the successful building, development and operation of the proposed cessda-ERIC.

3.3.1. Aims

The aims behind the establishment of the proposed CESSDA Training Programme are:

Specific ERIC related aims:
   a) Enable current CESSDA member organisations to fulfil the membership criteria of the proposed CESSDA;
   b) Develop additional skills and competences in connection with enhanced capacities and fresh activities within the new ERIC;
   c) Establish a means for the continuous development of professional competence within the new ERIC;
   d) Establish appropriate channels for the exchange of knowledge and harmonisation of professional competencies among CESSDA member organisations;
   e) Establish appropriate channels for knowledge transfers to potential members and emerging archival activities;
   f) Promote the use of CESSDA's services among the research community.

Development of the field of data preservation and management:
   g) Contribute to the institutionalisation of data preservation and management;
   h) Develop increased awareness of data management and data archival work among social scientists.

3.3.2. Selected arguments

Membership criteria: Full membership of the proposed ERIC will require compliance with common standards related to: data management and preservation (adherence to the OAIS reference model or CESSDA Seal of Approval and approved certification of operational procedures and processes); metadata production and presentation, a common authentication system; 'local' language thesaurus; sharing of data archiving/preservation

tools; contribution to cross-national data harmonisation infrastructure and 'question bank' (see CESSDA-PPP 2009a and 2009b). A common Training Programme will:
(1) Provide essential knowledge regarding the requisite standards for data management and preservation, and;
(2) Assist in the acquisition of skills necessary for the implementation and achievement of these standards.

Continuous development of professional competence within CESSDA: A key aspect of the new organisation will be the creation and maintenance of professional excellence in the fields of data archiving, access and preservation (e.g. see the CESSDA-PPP 2009a: 3.2 and 3.3). It requires the systematic development of human resources including the continual updating of professional knowledge and competence. At the same time the possibility of systematic and efficient knowledge management based on an assessment of needs, long-term planning and the systematic realisation of training plans is limited depending on the size of the organisation. A common CESSDA training programme could provide:

(1) Professional training plans and recommendations;
(2) Appropriate external training opportunities.

Both should be developed centrally to ensure efficiency.

Establishment of channels for the exchange of knowledge and harmonisation of professional competence among CESSDA members: There are differences in both the kinds and level of professional competences among current CESSDA member organisations. The establishment of channels for the exchange of knowledge will provide an important tool for the efficient development of professionalism, and this will also assist in levelling the disparity in skills across the current CESSDA network. The harmonisation of professional competence is an important aspect of an enhanced RI.

Establishment of channels for knowledge transfer to potential members and emerging archival activities: Support for new and emerging archival activities will have two main aims:
(1) Help new archives/members to attain the proposed membership requirements;
(2) Contribute to CESSDA's mission of supporting new archives and widening participation.

Dissemination of knowledge on usage of data services among researchers: Ensuring accessibility, and knowledge of, CESSDA's services and the promotion of skills to utilise these services is an important precondition for the success of the upgraded RI. It will also contribute to CESSDA's strategic goals through the promotion of data sharing and secondary data analysis.

Contribution to the institutionalisation of the Social Science data field and the opening up of opportunities for Data Scientist's career development: A significant issue associated with the organisation and funding of social science data archiving is related to the fact that the status of certain areas of data archival work is often unclear. Currently data archiving is not a part of any clearly defined scientific field. That is also why archiving projects are often seen to "interfere" with systems dedicated to the functioning and support of scientific work. The status of data archivists is also unclear because their position within the evaluation of scientific work and conditions for their career development are often highly variable. The JISC report on the role of data scientists (Swan and Brown 2008) argues for the formalisation of the role of "data scientist" including the establishment of channels for

professional career development. Organisation of a targeted Social Science data training programme is an important precondition for such process.

General cultivation of data management and data work in the Social Sciences: Interest among researchers in data management skills is steadily increasing. At the same time data archiving is dependent on their outputs. Thus any contribution to the upgrading of researchers' data management skills helps not only the quality of research, but also improves:
(1) Efficiency of data services;
(2) Visibility of data services and establishment of the field of Data Science. Moreover, the training of researchers in data work also complies with CESSDA's objective of promoting secondary data analysis.

3.4.   Target training groups

We recommend opening up access to proposed training activities to the following target groups:

1) Data professionals from CESSDA member archives and associated organisations;
2) Data professionals from potential CESSDA member organisations;
3) Data professionals from other data organisations in the social science field;
4) Data professionals from data archives, libraries and repositories who are not specialists in the social sciences field;
5) Data producers and researchers working with social science data.

Groups 1 and 2 are priority target groups. Their needs will constitute top priority for the proposed Training Programme. Group 5 is specific, because it represents users of data services, while the other groups primarily include data providers. Targeting this 'data user' group is essential in respect of an upgraded RI's aim of disseminating knowledge through its services and promotion of more extensive usage of these services. It will also allow CESSDA to contribute to the general cultivation of data management and quality of data work in the social sciences. We recommend: (1) Opening access to the Training Programme to data producers and researchers, and (2) Establishing a specific strategy including the organisation and promotion of training activities targeted specially to users of CESSDA services.

Beside others there are the following important benefits from targeting of groups beyond CESSDA and its associated organisations:

a) Inclusion of groups 3 and 4 will open new channels for collaboration beyond the current membership network, including more opportunities for the involvement of external experts in organising training - thus establishing an important conduit for quality enhancement of the programme;
b) Targeting of different groups will facilitate combining a greater range of sources for financing the programme; diversity in financing could be invaluable to the long-term sustainability of the programme;
c) Achieve an economy of scale through a more extensive training programme;
d) Extend the visibility of the upgraded RI and propagate social science data archiving and secondary data analysis;
e) Create a potential pool of new employees and a better means of recruiting new staff.

The targeting of non-member organisations and groups will probably require different approaches. We recommend a deliberate strategy based on following principles:

f) A Training Programme should be prepared primarily in respect of (1) the needs of the new RI, (2) co-operation with non-member institutions should be established primarily in terms of the upgraded RI's needs;

g) Open access to CESSDA's training activities to all target groups. Members of groups 1, 2 and co-operating institutions should have priority in cases of limited availability;

h) Programme topics should be extended beyond CESSDA's needs only on the basis of availability of external funding for this purpose and pre-establishment of demand for the proposed courses.

3.5. External co-operation

An upgraded RI **should** endeavour to cooperate with selected non-CESSDA member organisations regarding appropriate coordination of respective training activities and sharing available capacities. Co-operation with non-members can bring significant advantages:

- Easier access to external resources (expertise) and training opportunities;
- Co-operation to actualise specific training activities;
- Co-ordination of activities addressing common aims (e.g. professionalisation).

Potential partners for co-operation include:
- International organisations in social science data archiving (IASSIST[16], ICPSR[17], IFDO[18]);
- International organisations and projects in the field of digital data management and preservation in general.[19]

We recommend contacting relevant projects and organisations and establishing channels for the exchange of information as soon as possible. Co-operation with IASSIST and ICPSR should be a priority.

3.6. Training Plan

We recommend basing future training activities on a continuous planning model and systematic achievement approach. In this case the Training Programme will consist of two tasks:
- Development of training plans;

---

16 . International Association for the Social Science Information Service and Technology (IASSIST) is an international organization of professionals working in and with information technology and data services to support research and teaching in the social sciences. http://www.iassist.org

17 Inter-university Consortium for Political and Social Research, http://www.icpsr.umich.edu/

18 International Federation of Data Organisations for Social Sciences (IFDO), http://www.ifdo.org/

19 For example: DPE: Digital Preservation Europe, http://www.digitalpreservationeurope.eu/; PLANETS: Preservation and Long-term Access through Networked Services, http://www.planets-project.eu; CASPAR: Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval etc.

- Realisation of CESSDA specific training activities in accordance with the training plan.

The CESSDA training plan should comprise of a long-term strategic plan and supplements including short-term (annual) plans of training activities. The scheme for the development of the CESSDA training plan is shown in Figure 6.

Figure 6.Scheme for the development of the CESSDA training plan
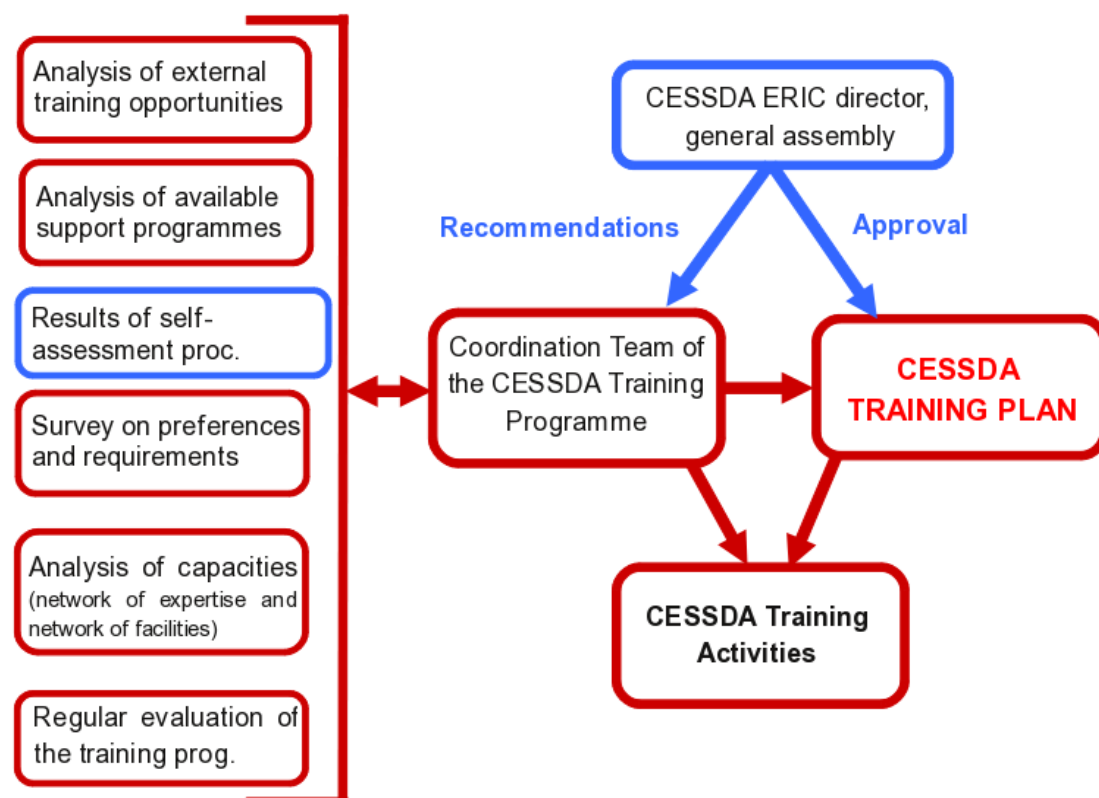
Source: CESSDA-PPP, WP6, Task 7 report.

The plan should be regularly updated. Planning should be based on the requirements of new organisation's governing bodies and the following analysis:

- A survey on the preferences and requirements of cessda-ERIC member and associated organisations and proposed Working Groups: we envisage the development of a simple on-line accessible questionnaire for the submission of requests;
- Self-assessment procedures (see Chapter 2) measuring the compliance of all members and associated organisations with the upgraded RI's standards and requirements: we expect to analyse the data from the self-assessment procedures developed by the new organisation for a variety of purposes;
- Overview of external training opportunities: relevant external training activities should be continually monitored;
- Overview of capacities: A network of expertise and a network of capacities will be organised. This will inform the proposed Training Programme and establish sources of external support and collaboration;
- Overview of support programmes: Evaluation of the strengths/weaknesses of

previous phases of the programme provided regularly by the proposed Co-ordination Team.

The programme Co-ordination Team should report regularly on the progress of the training plans in achieving their prescribed outcomes. The cessda-ERIC governing body should approve training plans before their implementation.

3.7. Topics for training and types of knowledge transfers
3.7.1. Topics for training

The content of the programme should not be fixed, but flexible according to the needs of the developing upgraded RI. These needs will be systematically planned according to the results of need assessments, the evaluation of priorities and the ongoing monitoring of current training.

The Training Programme should not concentrate on training in skills related to social science research based on data analysis (e.g. methods of data analysis, statistics, statistical and other relevant software, etc.), because the training needs in this area are covered by existing training opportunities.

Instead, new training activities should concentrate on two priority tasks:

- Development of professional skills and competences specifically related to data archiving;
- Providing knowledge on the infrastructure and its data services to users.

Activities in the first instance should focus on establishing a comprehensive set of training opportunities, which would cover perceived gaps in content and the capacity of existing external opportunities. It would also enable systematic knowledge management in relation to the requirements for membership of the new RI, development and implementation of new technologies, the enlargement and building of new capacities and open possibilities for the development of systematic career opportunities for data scientists. It should therefore include:

- Systematic training in basic archiving skills, e.g. metadata production, techniques in digital data preservation, controlled vocabularies and thesauri, standards and best practices in archiving, specialised tools and systems - NESSTAR, etc;
- Skills necessary for data archive management, e.g. data services management, open access and open source issues, legal issues, etc;
- Skills related to specific ERIC services, this should include integration of international data, data harmonisation, data anonymisation, etc;
- Specific skills related to the development and implementation of technology, e.g. courses in specific tools and software essential to, or associated with, the new organisation.

Activities in the second instance should:
- Provide easily accessible and clear practical information to a wide range of users of various CESSDA and associated services;
- Contribute to the dissemination of data services information to a wider audience: promoting the wider usage of CESSDA services and secondary data analysis in general.

3.7.2. Types of knowledge transfers

Based on aims and potential topics for training the following five different types of knowledge transfer channels may be identified. The methods of training offered by the programme should correspond to specific requirements within each of these types:

- Training in (1) basic archiving skills targeted especially to new and inexperienced staff, (2) advanced skills reflecting the aims of continual professional development and career building;
- Training in specialised skills and new technologies reflecting the development of the upgraded RI;
- Providing an expert knowledge base which is easily accessible and functions as a comprehensive source of day-to-day advice;
- Disseminating personal experience and expert advice in data archiving tasks;
- Addressing data users.

3.8. CESSDA training activities
3.8.1. Structure of the CESSDA Training Programme

We recommend establishing a structured Training Programme to address different target groups and different types of knowledge transfer through appropriate training methods. The resulting proposed structure of the programme is summarised in the figure below.

Figure 7. CESSDA training activities and types of knowledge transfers
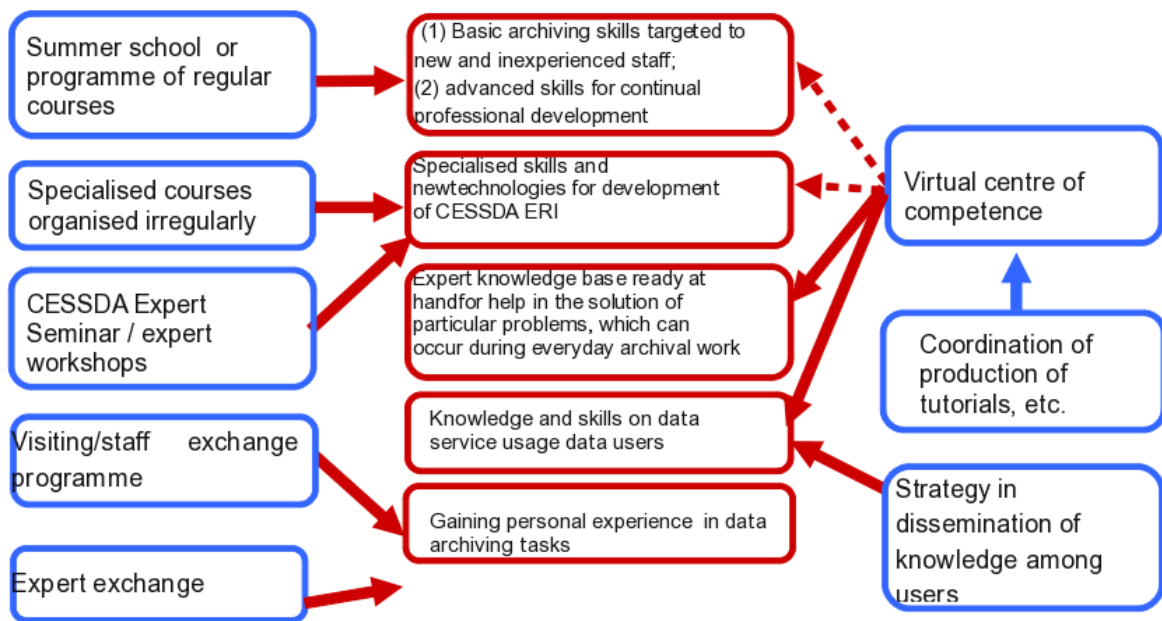
Source: CESSDA-PPP, WP6 Tasks 5 and 7

3.8.2. Summer school and/or continual programme of regular courses

A summer school should be organised to provide opportunities of training in (1) basic archiving skills targeted especially to new and inexperienced staff, (2) advanced skills reflecting the need for continual professional development and career building.

Emphasis should be on orderliness, intensity and regularity of training. Associated activities should be organised annually. Different topics could be repeated in different periods according to the needs of the target community/members of the new RI/external

demand. The programme should be structured and allow participation only in selected courses. Following completion of a course, participants' competency could be tested and confirmed by certification.

A summer school could be hosted and organised either by a specific organisation/location, or rotate between the new organisation's members, according to available resources and organisational/technical frameworks/competency in running such an event.

3.8.3. Adhoc specialist courses

Smaller ad hoc courses could be organised to provide training in specialised skills and new technologies, thus ensuring that personnel of the new RI's member organisations are at the cutting edge of emerging and developing technologies. Such courses should be targeted primarily toward the needs of experienced data professionals within the CESSDA network and its associated organisations.

We recommend and emphasise the need for flexible arrangements and timing with regard to the ongoing development of the new RI and the availability of expertise to participate in any proposed courses/training. These events should therefore be carefully scheduled to fit with the availability of trainers and resources.

The location of specialist courses should not be fixed, but selected according to demand and expertise, and any existing offers from member organisations. Online video conferencing could help facilitate wider participation, and courses could also take place in conjunction with other proposed activities, e.g. expert seminars, summer schools, etc.

3.8.4. CESSDA Expert Seminars and adhoc expert workshops

The Training plan and identified training needs should be taken into consideration in the planning of Expert Seminars. This type of event should be used for the exchange of specialist skills, new technologies and dissemination of information regarding the future development of the new organisation. They should also be used as a forum for feedback and planning. As it is not proposed that such events are "fixed" in the calendar, they should be organised according to participant availability. As above, online video conferencing provides a potentially very useful way tool for increasing participation in such events.

3.8.5. Visiting and staff exchange programmes: Background

Visiting programme (VP): Short-term stay(s) of employee(s) from one archive with the specific goal of improving their expert knowledge and/or understanding of certain archival practices;
Staff exchange programme (SEP): Establishing of bilateral partnerships and staff exchange between experienced and less developed organisations. Such an exchange allows less developed archives to benefit from the experiences and knowledge of staff at more well established partner archives/organisations.

Gaining personal experience in dealing with data archiving tasks is very important especially for new and underdeveloped archives and for those launching new activities. Such visits and exchanges should allow less experienced partners and potential members of the new RI to take full advantage of existing expertise within the network. Individual visits organised through the work of CESSDA and EDAN have already proved successful (see Hausstein and Guchteneire 2002). Additionally, specialised EU programmes, e.g. Marie Curie Fellowships, might also be able to support these exchanges and visits.

On the other hand, hosting institutions can experience serious disruption through frequent visits. Larger archives, which often act as hosting organisations, prefer structured and systematic programmes rather than individual visits. Additionally, the prolonged absence of an employee from smaller archives could cause problems in the organisation of work. Therefore, the establishment of staff exchange and/or a visiting programme **should** be undertaken in a deliberate and considered manner. Such a programme must be structured, and backed by financial and organisational conditions favourable for both guest and host, and participation distributed among a wide number of organisations.

3.8.6. Staff exchange and/or visiting programme (SEP) for less developed archives

An SEP should be an essential part of the proposed Training Programme, especially within the context of the widening issues explored by the CESSDA-PPP Workpackage 7 team. It is one way to disseminate the skills of staff employed at member organisations and foster knowledge transfer between partners/associates. The concept of exchanging staff with partners (at various levels of the new organisation's membership, but most especially at the associate level) is a way of sharing knowledge and building new perspectives to help spur on the development of new initiatives.

Proposed objectives for an SEP:
1. Increase the upgraded RI and its partners' capacity to foster development;
2. Strengthen relationships among developing partner organisations;
3. Enhance the skills and knowledge of programme participants;
4. Foster cultural diversity and awareness for the new organisation and its member organisations.

We recommend establishing a structured and coordinated SEP which should be internally and externally financed. Each year there should be an announcement offering staff exchange opportunities whose duration varies between one week and a month. The provision of SEP's will of course depend on securing available resources and offering interesting topics. The SEP should be organised by the Coordinating Team. cessda-ERIC associate members and potential members are eligible to apply for an SEP. The programme should consist of placements in one of the cessda-ERIC full member organisations.

3.8.7. Expert exchange within the cessda-ERIC

Short term stays to gain a certain kind of expertise: CESSDA visiting programmes should be made available to all levels of ERIC membership and associate organisations. Short term stays should be for the purpose of gaining personal experience with new technologies or rectifying particular skills gaps. The aim is not to establish extensive bilateral partnerships and complex transfers of knowledge, but to facilitate the short term stay of an expert from a cessda-ERIC member organisation seeking some specific kind of expertise from another member institution.

In-house training based on expert exchange: Establish a system of expert exchange that could also be used for the organisation of in-house training for cessda-ERIC member organisations and associate members.

We recommend establishing an expert exchange system as a common platform allowing for a centralised gathering and announcement of requests, opportunities and co-ordination of activities in this area.

3.8.8. CESSDA Virtual Centre of Competence (VCC)

The Training Programme will deal with many types of participant and organisations. For this situation, **we recommend** the provision of a Virtual Centre of Competence (VCC): a place where both specialists and newcomers can meet virtually in order to solve difficulties step-by-step, as they appear. Task 11 of work package 7 is devoted specifically to planning on building the CESSDA VCC and provides further elaboration of this part of the training programme.

The VCC will have a similar structure to the current discussion forums, but will also take advantage of recent developments in the field of communication. As internet bandwidth is getting bigger and more affordable, it is anticipated that live video tutorials will soon be disseminated from this site. Additionally, video-conferencing is being considered as a practical solution to travel/transportation issues. A wide variety of e-tutorials, e-books, e-papers, reference guides and manuals will be available from the VCC website, where interested persons will be able to easily browse for information.

Coordination of the production of tutorials and other learning materials for users: All ERIC services should be backed-up by appropriate user guides and tutorials. We anticipate that these materials will be developed by providers of particular services and organisers of various CESSDA-based activities. For most of them these will also be relevant for inclusion into the knowledge base of the CESSDA VCC. Therefore, **we recommend** a centralised co-ordination approach in relation to the production of learning materials for users. This will ensure the appropriate cross-referencing of information on different CESSDA services in the VCC.

3.8.9. A common strategy for the dissemination of knowledge on data services among users

The general promotion of data sharing, open access and secondary analysis is one of CESSDA's strategic goals. In this respect **we recommend** the following:

- Wide dissemination of learning materials on how to use CESSDA services;
- Development of special materials useful for the teaching and studying of data analysis at universities, e.g. training data packages, training kits on research infrastructures, data management, etc. These should be readily available both to teachers and students at universities;
- Providing support for the activities of CESSDA member organisations disseminating knowledge at a national level;
- Coordination of selected training activities with regard to wider CESSDA strategies regarding promotion of, visibility and propagation of the ERIC.

3.9. Resources for organisation of the training programme

CESSDA training activities will require appropriate institutional resources including requisite levels of knowledge, training facilities and equipment. There are four possible channels for such resources:

- Existing expertise and facilities at CESSDA member archives: The CESSDA network should be a primary source of expertise. Most CESSDA member institutions have experience of organising training activities. Many of them should be able to contribute to the programme through the provision of access to training facilities;
- Co-operation with other organisations active in data archiving: Collaborate with

organisations with similar objectives - potentially such organisations could become an important source of additional expertise (see above point 4);

- Co-operation with universities and training centres: The necessary technical background and training facilities may be hired from specialist institutions. In the case of an extensive CESSDA training programme, including the provision of courses for universities and research institutions, access to facilities and equipment could also result from bilateral cooperation;
- Building of new capacities: New training facilities can be newly constructed and new experts can be hired during the construction phase of the cessda-ERIC. Such activities can be financed from specialised projects.

In essence, organisation of the programme could be facilitated by using (a) existing sources within the CESSDA network, (b) cooperation with non-CESSDA organisations, and (c) individual contracts for particular events. The construction of new facilities may result from decisions on the range of the proposed training programme, or could arise from the availability of specific opportunities, e.g. support programmes oriented to the construction of new capacities.

For purposes of grounding training plans on the up-to-date availability of expertise and training facilities, with respect to concrete and detailed training needs, **we recommend** organising a survey of CESSDA member capacities at the beginning of the Construction Phase of the cessda-ERIC.

3.10. A distributed or centralised approach?

Three different possible infrastructures could be employed to ensure appropriate facilities for the proposed Training Programme:

- Distributed training infrastructure: Establishment of a small coordination centre (which may be virtual), network of expertise and network of multiple training facilities. Responsibility for the organisation of training activities would be distributed among CESSDA members;
- Centralised training infrastructure: Construction of common CESSDA training centre, drawing together expertise from across the CESSDA network and providing a convenient framework for the organisation of training activities;
- Combination of distributed and centralised approach: Construction of a CESSDA training centre for some types of activities and the distribution of other activities among CESSDA members.

Advantages of a distributed training infrastructure:

- Corresponds to a distribution of expertise and existing facilities within the CESSDA network;
- Allows more flexibility and operability to fulfil different types of training needs;
- Allows more flexibility and operability to react to frequent changes in funding policies;
- It would not require significant investments re:the building of a new central facility, as it would be possible to invest in improving existing facilities among a large number of CESSDA member organisations;
- It would ensure wider participation in common activities and could stimulate the development of CESSDA archives, including less resourced organisations.

Advantages of a centralised training infrastructure:

- Provides a significant contribution toward the visibility and propagation of the ERIC;
- Under some circumstances it could provide better conditions for financing the training programme within the ESFRI process (e.g. a significant contribution from one country and contribution from the EC for international ERIC work);
- It would provide an opportunity for securing other sources of financing, For example, structural funds if the training centre were built in an underdeveloped (<100% GDP) region of the EU (this could be employed in relation to ESFRI or independently of this process).

Broadly speaking, the distributed model corresponds best to CESSDA's needs, but the availability of financial support for the proposed training programme under a centralised model may prove a decisive factor. A CESSDA Training Programme **should not** divide CESSDA archives into two groups, the first comprising of well developed archives who would be pure contributors to the programme, and the other composed of less developed recipients. On one side there are benefits for continuous professional development relevant to all organisations, on the other side there is the need to establish a structured, but flexible, training programme based on common aims, various training methods and the need to secure and share practice based knowledge from multiple sources. Therefore **we recommend** opting for a distributed or mixed model of training infrastructure.

3.11.   Organisational structure of the programme

Figure 8..Organisational structure of the CESSDA Training Programme
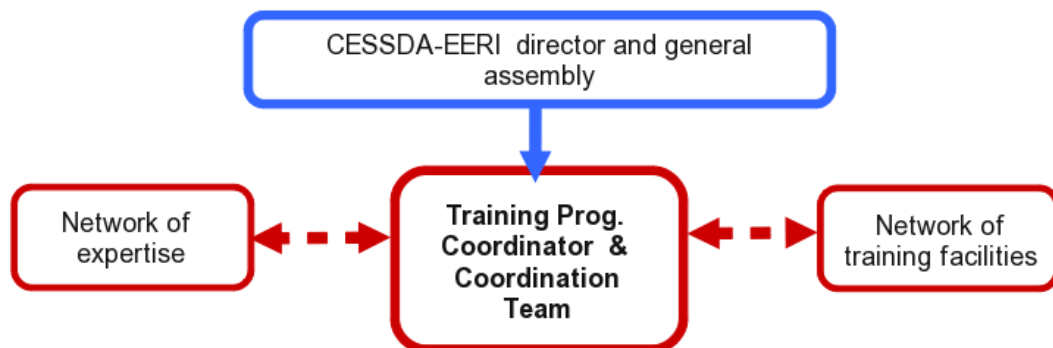
Source: CESSDA-PPP, WP6 tasks 5 and 7.

The Coordination Team (CT) should comprise the following personnel:

1. A Training Programme Co-ordinator: a permanent CESSDA-ERIC employee who will lead the team;
2. A part time IT expert, also an ERIC employee;
3. A small group of experts from different CESSDA member archives (residing at their respective institutions, but regularly participating in meetings)

The CT should:

1. Co-ordinate and take overall responsibility for the programme;
2. Develop a training plan;

3. Develop a curriculum for the Summer School;
4. Organise networks of expertise and facilities;
5. Co-ordinate training activities and assist with the organisation of such events;
6. Regularly evaluate the training programme and report to the ERIC governing bodies;
7. Actively seek support for training activities and stimulate and co-ordinate project activity among CESSDA member organisations;
8. Actively seek out co-operation opportunities for the CESSDA network and formulate proposals based on these findings.

Network of Expertise: This should consist of a database of contacts and information on available expertise. The Network of Expertise should be used for the organisation of training activities and the establishment of communication channels.

Network of Facilities: This should be a database of information on available and appropriate training facilities and communication channels suitable for the organisation and facilitation of training activities.

Establishing the Network of Expertise and the Network of Facilities for training purposes **should** be co-ordinated with other activities addressing expertise both within and outside the current network.

3.12. Financing the programme
Funding of the programme should be derived from the following sources:

- The cessda-ERIC central budget

Expenditure for co-ordination and centrally located activities, including the salary of the coordinator and remuneration for other CT members (in minimum 12 PMs/year distributed among CT members), a portion of an IT expert's salary (2-5 PMs/year), travel expenses for co-ordination purposes, part of any expenditure related to the cessda-ERIC central office and its equipment, expenditure for survey and analysis work, building and maintenance of the CESSDA VCC and the Expert Mailing List, etc; should be covered by the international budget of the cessda-ERIC.

The cessda-ERIC **should** also establish a special fund to provide loans for organisers of training activities, in order to allow any advance costs to be covered prior to the receipt of participant fees.

Outreach fund: In accordance with the project objectives and the proposed training programme **we recommend** establishing a CESSDA Outreach Fund to assist selected participants from less resourced archives and emerging archival activities.

- Participation fees

Organisation of CESSDA training activities should be covered by participant fee costs, if no external sources of funding (see below) are available. This may mean that expenses associated with the participation of employees of CESSDA members will need to be covered by national sources. In some cases there may be different fees for CESSDA member organisations and non-members, or reduced rates for some type of participants. Loans from a central budget **should** be available to organisers to cover costs before participant fees arrive.

- External support

There are many opportunities to obtain external support for organising training activities, participation in training programmes, and the building of training facilities. The Co-ordination Team **should** be active in seeking out the widest possible number of funding sources when organising common CESSDA projects.

3.13. Phases of the CESSDA Training Programme

Preparatory phase
The following tasks should be performed before launching of the Training Programme:
- Adopt a resolution on the ERIC Training Programme and make a decision on the proposed possible training infrastructure models (i.e. distributed, centralised or mixed), location of the training centre and budget size for centralised activities;
- Start contacting potential external partners;
- Recruit the ERIC Training Programme Co-ordination Team;
- Start fundraising.

Construction phase (years 1-2)
The construction phase will require extended effort and resources from the upgraded RI, and in particular input from the proposed Co-ordination Team.

Preliminary training activities (year 1): Time and capacity requirements for developing and launching the CESSDA standards, assessment processes and development of the Training Plan means it will not be possible to have the full range of Training Programmes available immediately. At the same time, we should anticipate a greater demand for training during the construction phase of the ERIC. For these reasons, **we recommend** starting with the following preliminary training activities before the first standard training plan is developed:

- A CESSDA expert workshop on training issues to identify basic and current training needs and how these relate to the construction process of the new organisation;
- Set of training courses addressing basic archiving skills and the current training needs of the ERIC;
- Establishment of a platform for expert exchange and creating the basis for a future Visiting and Expert Exchange Programme.

Construction of the CESSDA Training Programme (year 1-2):

- Establish the Co-ordinator central office;
- Implement a survey on preferences and needs of CESSDA members and associated organisations;
- Conduct a survey on available capacities within the CESSDA network;
- Establish contacts and co-operation with associate and affiliate partners;
- Establish the Network of Capacities and Network of Expertise;
- Develop the CESSDA Training Plan;
- Develop the Summer School curriculum;
- Begin to develop a programme of courses;
- Launch the CESSDA visiting and staff exchange programmes;
- Build the VCC;
- Construct/formalise the training centre(s) if it's agreed to use a centralised or mixed

training infrastructure model at the beginning of the programme.

Operational phase (year 3+)

It is envisaged that there will be regular updating of the Training Plan, continuous refinement of training activities, combined with maintenance and further development of established systems.

4.   Meeting the standards in data archiving[20]

Based on the conclusion of the WP6 task 6 interim report, a number of documents and software tools are recommended for inclusion in the CESSDA toolkit. These would assist less-resourced, smaller archives to work towards upgrading their capabilities to meet the proposed cessda-ERIC Full membership criteria. This recommendation is based on an analysis/identification of technological, organisational and financial barriers that hold back the development of emerging social science data archives. The report also suggests resolutions/strategies for overcoming the identified obstacles.

> Proposed elements of a CESSDA toolkit:
>
> *Marketing materials and public documents on activities*: Flyers, newsletters, annual reports;
>
> *Legal documents*: Sample contracts and end-user licenses, Guides to EU legislation / directives on research data, Best practice of sharing public use micro-data;
>
> *Technical tools and best practice guides*: DDI best practice guides, cessda-ERIC archival standards guide, DDI software tools, Guide to open source web content management tools and data publishing tools, Step-by-step guide to sharing metadata with the CESSDA network, ELSST thesaurus localisation and maintenance guide.

There are numerous standards and reference models for digital repositories. However none of the existing models fully conform to the aim of establishing a general ontology (dictionary, terminology, reference model) that would facilitate the mapping of current best practice used in social science data archives. It makes sense to compare models with existing guidelines and documents describing social science data archives routine activities, and extract a new specific social science data archival reference model. The Data Seal of Approval (DSA) could be used as a starting point for the development of a seal of approval. This could be more specific concerning its aims and target groups; and thus better reflect more particular/relevant requirements of the consortium. Its development requires a revision of the current data seal guidelines and opens up the possibility for greater elaboration and/or new additions. In the medium-term, **we recommend** developing a comprehensive framework of best practices.

4.1.   Summary
Section 4.2 highlighted potential shortages and barriers, section 4.3 identified resources requiring upgrade. Section 4.4 recommends possible solutions for inadequate organisational, technical settings, and finally section 5 lists recommended elements of the proposed CESSDA Toolkit. Operating the upgraded RI should be based on a set of agreed quality criteria. We have made recommendations concerning the necessary standards in social science data archiving required for establishing the proposed ERIC.

As argued in section 4.6 there are two levels/types of standards: (1) minimum, and (2) best practice. The DSA could be used as a starting point for a seal of approval, providing minimal standard definitions (section 4.8.1). Finally, the last section (4.8.2) argued for the provision of a best practice working group, which would be responsible for further elaborating standards and recommendations for operating a quality guaranteed, digital social science data archive. These standards and recommendations would need to reflect

---

20 This chapter is based on interim reports from tasks 6 and 8.

the aim of preserving data collections long-term whilst providing controlled, open access to a designated user community.

4.2. Identification of the shortages (technical, organisational) and barriers (technical, administrative)

The minimal set of requirements for Full and Associate membership of the proposed ERIC have been intensively discussed during the project. According to the statutes of the ERIC, all members will be obligated to:

- Make a full contribution to the cessda-ERIC budget;
- Be fully compliant with the DDI metadata standard;
- Adopt and apply the proposed cessda-ERIC common Single Sign-on user authentication system;
- Enable the harvesting of all their catalogue records for inclusion in the cessda-ERIC data portal;
- Make their data holdings downloadable through common data gateways;
- Maintain their local language(s) within the multi-lingual thesaurus;
- Share their data archiving tools (under the IP conditions set in the Intellectual Property Rights annex to these statues);
- Adhere to the OAIS reference model and/or the agreed seal of approval for archival practices;
- Contribute to the ERIC's cross national data harmonisation activities;
- Contribute material and/or expertise to the cross-national question bank;
- Help and support designate members of the new organisation to achieve Full or Associate membership (as applicable);
- Where possible, facilitate access to national government (and research-funded) microdata.

As noted above, it would be required that service provider archives should "adhere to the OAIS reference model and/or the agreed seal of approval". At the time of writing there are no agreed quality standards for the planned ERIC. However, intensive discussion is taking place and existing international initiatives are being carefully considered, e.g. NESTOR[21], Data Seal of Approval (DSA)[22] etc.[23] Starting with previous CESSDA-PPP interim reports re: best practice and standards, we tried to map some of the most relevant elements of *Catalogue of Criteria for Trusted Digital Repositories* and the DSA to the proposed cessda-ERIC Full-member criteria. The results are shown on the following pages:

---

[21] See the homepage of the German NESTOR project at http://www.langzeitarchivierung.de/. For the English version, see: http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf Cf. NESTOR Working Group on Trusted Repositories Certification: Catalogue of Criteria for Trusted Digital Repositories, Version 1 (draft for public comment), June 2006, Frankfurt am Main: nestor c/o Deutsche Nationalbibliothek, urn:nbn:de:0008-2006060710
http://www.nbn-resolving.de?urn:nbn:de:0008-2006060703.

[22] Sesink, Laurents, René van Horik, and Henk Harmsen. sa. Data Seal of Approval.
http://www.datasealofapproval.org/sites/default/files/Data_Seal_of_Approval_1-4.pdf

[23] For example Task 9 within WP6 will deal with this issue. See also Dusa, Adrian and Janez Stebe. 2009. "Task 6.4 Report: Recommendations concerning best practices ". Draft v0.4.
http://www.cessda.org/ppp/wp06/WP6_T6.4.pdf

Organisational barriers

cessda-ERIC requirements	Criteria (Number in NESTOR / DSA catalogue)	Explanation from NESTOR catalogue (in quotation marks)
To make a full contribution to the -ERIC budget	Adequate financing of the digital repository is secured. (NESTOR 4.1)	"The digital repository should be able to demonstrate that the proposed services can be financed, both in the short and long term. The financing of the digital repository should have a legally secured basis. In the case of state-financed digital repositories, the financing should be included in the formal planning documents (at least medium-term). A private digital repository should be able to guarantee its financial sustainability on the basis of charged use of its services and on a long- term business plan."
General requirement	Sufficient numbers of appropriately qualified staff are available. (NESTOR 4.2)	"The qualifications and training of the staff should be adequate for the goals, tasks and processes of the DR. Suitable schemes should be in place to ensure adequate training and further training in the long term. Staff numbers should be sufficient to allow all necessary processes to be fully completed. The long-term planning of the DR should consider staffing resources."
General	The digital repository has defined its designated community(ies). (NESTOR 1.3) A data repository has an explicit mission in the area of digital archiving and propagates it. (DSA 4.)	"The general definition of the framework for a DR involves defining the designated community(ies)/designated community. This includes knowledge of the specific requirements of the designated community(ies) influencing the selection of the services to be provided. If the designated community or its requirements change over time, the DR should respond by adapting its services."

Technological barriers

cessda-ERIC requirements	Criteria (Number in NESTOR / DSA catalogue)	Explanation from NESTOR catalogue (in quotation marks)
Make their data holdings downloadable through common data gateways; Enable the harvesting of all their catalogue records for inclusion in the -ERIC data portal; Adopt and apply the ERIC common single sign-on user authentication system.	The digital repository ensures its designated community can access the digital objects. (NESTOR 2.1) "The research data can be found on the internet." (DSA 1st quality criterion)	"The DR should ensure that authorised users have access to the digital objects. This includes the provision of adequate research opportunities. When determining its service portfolio, the DR takes considers the needs of its designated community into account. The DR announces in advance its conditions of use ..."
Maintain their local language(s) within the multi-lingual thesaurus; Be fully compliant with the DDI metadata standard.	The digital repository ensures that the designated community can interpret the digital objects. (NESTOR 2.2)	"The DR should take appropriate measures to ensure that the digital objects can be interpreted on a long-term basis, thereby creating the basic requisites for adequate usage. This includes the ability to interpret both content and metadata. In ensuring this, the DR should consider the needs of its designated."

cessda-ERIC requirements	Criteria (Number in NESTOR / DSA catalogue)	Explanation from NESTOR catalogue (in quotation marks)
Make their data holdings downloadable through common data gateways; Enable the harvesting of all their catalogue records for inclusion in the -ERIC data portal; Adopt and apply the ERIC common single sign-on user authentication system.	The digital repository ensures its designated community can access the digital objects. (NESTOR 2.1) "The research data can be found on the internet." (DSA 1st quality criterion)	"The DR should ensure that authorised users have access to the digital objects. This includes the provision of adequate research opportunities. When determining its service portfolio, the DR takes considers the needs of its designated community into account. The DR announces in advance its conditions of use ..."
Be fully compliant with the DDI metadata standard. Adhere to the OAIS reference model and/or the agreed ERIC seal of approval for archival practices.	The digital repository acquires adequate metadata to record the changes made by the digital repository to the digital objects.(NESTOR 12.4)	"The DR should document all changes made to the digital objects. This also includes recording the people, systems and rights involved (...). This documents authenticity (...) and also ensures technical preservation of the digital objects."

After the above mapping exercise we collected possible barriers to compliance with the proposed cessda-ERIC membership requirements. (Table 2)

Table 2. Possible barriers to Full-membership of the new organisation

Requirement	Barrier(s)
Make a full contribution to the ERIC budget	Inadequate national funding
Be fully compliant with the DDI metadata standard	Inappropriate training; Lack of knowledge of standards; Lack of appropriate tools for DDI compliance.
Adopt and apply the ERIC common Single Sign-on user authentication system	Lack of IT specialists who can configure and maintain a Single Sign-on user authentication system
Enable the harvesting of all their catalogue records for inclusion in the Data portal;	Lack of IT specialists who can configure and maintain an OAI-PMH server and IT infrastructures
Make their data holdings downloadable through common data gateways	Legal and technical barriers: Lack of IP rights management; No contracts for distributions; Same barriers as highlighted in latter two points.
Maintain their local language(s) within the multi-lingual thesaurus	Insufficient number of staff; Inadequate training for maintaining localised ELSST
Share their data archiving tools (under the IP conditions set in Intellectual Property Rights annex to these statues)	No tools to share; Tools available only local language; Unclear IP rights.
Adhere to the OAIS reference model and/or the agreed seal of approval for archival practices	No clear standards are set; Lack of awareness; Inadequate control of intra-institutional processes.
Contribute to the ERIC's cross national data harmonisation activities	Insufficient number of staff; Inadequate professional background and experience.
Contribute material and/or expertise to the cross-national question bank	Lack of tools to produce in-depth data descriptions; Lack of human resources: insufficient number of properly trained staff.

Requirement	Barrier(s)
Help and support designate members of the ERIC to achieve full membership or associate membership	Insufficient number of properly trained staff / inadequate funding
Where possible to facilitate access to national government (and research-funded) microdata	Inappropriate national legal system; Lack of awareness of or unfavourable attitude toward open access policy by legislators and governmental officials.

4.3. Identification of resources required for an upgrade

Based on the barriers listed in table 1, we conducted a survey among institutes participating in this workpackage. Archive representatives were asked to indicate on a four-point scale which of the prescribed barriers was perceived as being applicable to their organisation/archive. A higher score indicated a more detrimental barrier. The answers are presented in Appendix II. Mean scores ranged between 1,83 (ADP) and 3,53 (RODA). SDA and ADP could not, at the time of the survey, evaluate the adequacy of their national funding levels.

GESIS is one of the oldest and most developed social science archive in Europe. For GESIS the most detrimental barrier to participating in the ERIC is the "inappropriate national legal system" to "facilitate access to national government (and research-funded) microdata". There are more areas where they also saw reason for concern.

Generally speaking less-developed archives scored higher in the survey. An exception to this general trend was ADP, who expressed less concern about barriers to membership than other smaller archives. Emerging archives generally take issue with their lack of adequate funding, and this factor seems to be most critical as it informs all other perceived barriers.

Figure 9. Mean scores of the Full-member check list for WP6 partners
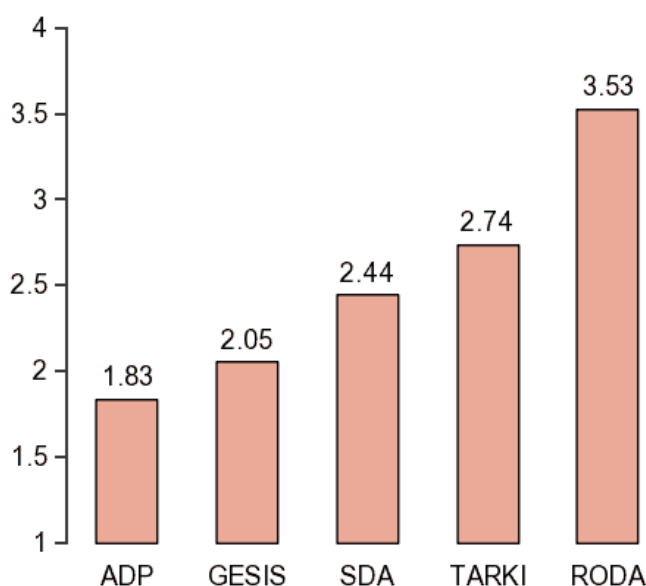
Source: WP6 Partner survey

4.4. Identification of possible solutions

No	Requirements	Barriers	Possible solutions
1	Make a full contribution to the ERIC budget	Inadequate national funding	Raise the awareness of national funding agencies; demonstrate the importance and quality of services you deliver to the user community. Create alliances among users and depositors. Ask for official support/backing from the ERIC board to convince national funding bodies.
2.1	Be fully compliant with the DDI metadata standard	Inappropriate training, lack of knowledge of standards	Use cessda-ERIC training programmes, VCC guides.
2.2		Lack of appropriate tools for DDI	Use VCC tools, ask assistance from expert groups, and members' forums.
3	Adopt and apply the ERIC common Single Sign-on user authentication system	Lack of IT specialists who can configure and maintain a Single Sign-on user authentication system	Get help from expert groups. Select appropriate software tools (open source or proprietary Web CMS). Hire competent IT specialist.24
4	Enable the harvesting of all catalogue records for inclusion in the ERIC data portal	Lack of IT specialists who can configure and maintain an OAI-PMH server	Same as previous. Note: some software has a built-in capacity to communicate with OAI-PMH servers (e.g. Dataverse).
5.1	Make their data holdings downloadable through common data gateways	Lack of IP rights management, no contracts for distribution networks	Use VCC sample contracts, policy guides; Outsource some of the tasks to legal experts/ lawyers.25
5.2		Same barriers stated for points 3 and 4.	Apply related solutions shown above

---

24 The CESSDA-PPP is currently working on a prototype for single sign on based on Shibboleth (http://shibboleth.internet2.edu/). The expectation is that managing and maintaining the system will be done centrally on behalf of all members. Where individual members need technical support to implement it, this may simply need be provided by the hub's technical support team (wherever they are located).

25 Work Package 10 is working on common licence/licences for the ERIC. This will include clauses on IP and distribution. Members will be encouraged (expected even!) to promote and use this locally to facilitate shared access.

No	Requirements	Barriers	Possible solutions
6	Maintain their local language(s) within the multi-lingual thesaurus	Insufficient number of staff; inadequate training for maintaining localised ELSST.	Employ a sufficient number of staff and train them using CESSDA resources and other available tutorials.
7.1	Share data archiving tools (under the IP conditions set in Intellectual Property Rights annex to the ERIC statues).	No tools to share.	Localise available tools and share translations with the community.
7.2		Tools available only local language.	Develop bilingual interface (English and local) by default, when tools upgraded. Ideally multilingual support will extend to other languages as well. Keep in mind possible character encoding issues and consider use of Unicode.
7.3		Unclear IP rights.	IP rights should be included in any contracts with developers
8.1	Adhere to the OAIS reference model and/or the agreed ERIC seal of approval for archival practices	No clear standards are set by the ERIC at the time of writing but DSA is expected to be adopted.	Minimum and optimal criteria should be agreed among members. See WP6 interim report, task 9
8.2		Lack of awareness among management.	Raise awareness by highlighting the advantages of membership and need for quality management of the required institutional processes and reputation of an organisation/archive. (Trust is necessary for long-term preservation)
8.3		Inadequate control of intra-institutional processes.	Define clearly in writing the tasks and responsibilities of each unit, staff member and any external contractors.

No	Requirements	Barriers	Possible solutions
9	Contribute to the ERIC's cross national data harmonisation activities	Insufficient number of staff, staff with inadequate professional background and experience.	Employ a sufficient number of qualified/experienced staff for cross-national empirical research.
10.1	Contribute material and/or expertise to the cross-national question bank	Lack of tools to produce in-depth data descriptions	Use ERIC tool resources (such as DDI tools26).
10.2		Lack of human resources: insufficient number of properly trained staff.	See point 6.
11	Help and support designate members of the cessda-ERIC to achieve Full or Associate membership,	Insufficient number of properly trained staff / inadequate funding	See point 6.
12.1	Where possible, to facilitate access to national government (and research-funded) microdata	Inappropriate national legal systems	Refer to EU level directives and best-practice; Consult with national legal experts and legislators. Ask for assistance from the ERIC, whose central hub will negotiate with EU-level organisations (such as EUROSTAT).
12.2		Lack of awareness of and/or unfavourable attitude toward open access policy among legislators and governmental officials.	Demonstrate international best practice in sharing public use micro-data (e.g. UK, Canada)

4.5.   Proposed elements of CESSDA strategic toolkit

Marketing materials and publically available documents:

- Flyer;
- Newsletter (online and e-mail subscription). This should contain usage statistics, new major studies acquired and archived, training programme and workshop announcements;
- cessda-ERIC annual reports: financial report, summary of development and activities
- Legal documents
- Sample contracts for dataset deposition and distribution

---

26 See http://tools.ddialliance.org/

- Sample end-user license
- Guide to EU legislation / directives on research data
- Best practice of sharing public use micro-data
- Technical tools and best practice guides
- DDI best practice guides
- CESSDA archival standards guide
- DDI software tools
- Guide to open source web content management tools and data publishing tools
- Step-by-step guide to sharing your metadata with the CESSDA network
- ELSST thesaurus localisation and maintenance guide.

4.6. Purposes and structure of standards

Developing and maintaining common standards is a prerequisite for ensuring the compatibility of resources and that processes function correctly across the future cessda-ERIC network. Furthermore, common standards are central to the aims of the ERIC (see CESSDA-PPP 2009b: 3.2 and 3.3) with respect to: (a) professionalism and demonstrating the highest possible standards in all activities and enhancing professional development amongst its member organisations, and (b) quality and facilitation of high quality research, teaching and learning.

It is therefore anticipated that the ERIC will set up a system of standards and propose an associated implementation plan. This should:

1. Guarantee that at each member organisation, settings of systems and procedures including their agreed quality profiles allow integration of its services and resources into a common network;
2. Provide a complex framework for the continuous quality improvement and evaluation of member and candidate member organisations achievements. This type of monitoring should be compliant with the aim of developing integrated services.

One of the most general and widely used definitions describes **quality** simply as "fitness for use" (Juran and Gryna, 1980). In practice this definition leads to multidimensional and complex quality criteria originating from different types of product or service use (as the importance of different product and service characteristics varies between users) and varying conditions for the production process and operation of services.

This is also true for social science data services. There are different types of data and requirements concerning processing, documentation and establishing access conditions. Data services are used by researchers from different research areas where specific requirements must be met in order to provide an extensive service. Preservation services are provided by a wide range of organisations. Cost efficiency is also an important factor and procedural flexibility is important in mitigating this issue.

In summary, the standards adopted by the ERIC **should** (1) set up clear and strictly obligatory conditions for all data services provided within the network, and (2) produce a flexible and sensitive framework of guidelines allowing high variance in the type and level of services, thus accommodating different organisational environments. This balancing of rigour and flexibility may be overcome by creating (at least) two different types of standards:

1) A minimum set of standards, which would clearly define the boundaries of acceptable professional and useful practice within the CESSDA network;
2) Best practice guidelines which describe the best methods and standard operating procedures whilst respecting differences in organisational environments. These guidelines will provide a framework for working toward outlined standards for high quality services and professional conduct.

In addition there could be a system of quality certification for different types and levels of service. This would facilitate: (a) more transparency for clients and evaluators; and (b) greater benefits for organisations making procedural and professional improvements.

4.7. Reference models

The quality of data services, as is true for any other kind of product, is achieved through ensuring the quality of production systems and processes. Quality may also be controlled through the regular evaluation of outputs, e.g. quality of services from a user viewpoint. However, in the latter case, such an approach would not directly address the problem of process compatibility within the network. Therefore an assessment of production systems and procedures should take priority in the new organisation.

The concept of a Reference Model is a useful tool for the description and implementation of process and system standards. It is based on setting up a complex framework of entities and their relationships, including a unified terminology for actors and their roles, functions, processes, mutual relationships and related standards and guidelines. Data organisations may then map their existing practices onto this reference model in order to evaluate their current standard, identify strengths and weaknesses and make plans for future improvements.

In their overview of existing models of best practice relevant to social data archiving, Štebe and Duşa [ibid.] differentiated between two approaches associated with data organisation activities: (1) an approach based on centrality of digital preservation (e.g. OAIS, TRAC, DRAMBORA, NESTOR, Data Seal of Approval) and; (2) a knowledge management perspective based on data life cycle models (e.g. intellectual background of DDI 3, the data life cycle model at ICPSR, DCC Curation Lifecycle Model).

Descriptions of data archiving activities based on reference models have been developed using both approaches. Fábian [2009] provides an in-depth examination of an ISO reference model of Open Archival Information System (OAIS)27 and an overview of related models (TRAC, DRAMBORA, NESTOR, Data Seal of Approval),28 which are

---

27 The Reference Model for an Open Archival Information System (OAIS) was originally produced by a Consultative Committee for Space Data Systems (CCSDS) as recommendation for space data system standards, and was later approved as norm ISO 14721:2003 applicable to any digital archive specifically applicable to organizations making information available for the long term. http://public.ccsds.org/publications/archive/650x0b1.pdf

28 TRAC = Trustworthy Repositories Audit & Certification, developed by the Office of Programs and Research (OCLC) and National Archives and Records Administration (NARA), see http://www.crl.edu/PDF/trac.pdf; DRAMBORA = Digital Repository Audit Method Based on Risk Assessment, developed by The Digital Curation Centre (DCC) and Digital Preservation Europe (DPE), see http://www.repositoryaudit.eu/download/; NESTOR = certification developed by Working Group on Trusted Repositories Certification organised at the NESTOR project (Network of Expertise in Long-term STOrage of Digital Resources – A Digital Preservation Initiative for Germany), see http://nestor.cms.hu-

based on a digital preservation approach. Štebe and Duşa [ibid.] compared selected digital preservation reference models (OAIS related models, Data Seal of Approval) and data lifecycle models (DCC Curation Lifecycle Model LIFE 2 and Cost Estimation Toolkit (CET)).29

The main advantage of an OAIS model may be summarised in terms of the greater maturity of tools and materials, and that it has already been successfully used for assessment of social science data archival activities in the UK (at the UK Data Archive: see Beedham et al. 2005) and in the USA (at ICPSR, see Vardigan and Whiteman 2007).30 However, the OAIS standard has been assessed as unsuitable (particularly with regard to its lack of scalability) for CESSDA purposes (see Beedham et al. 2005; Fábian 2009; Štebe and Duşa 2009). The OAIS model assumes a complex organisational structure and strongly emphasises technical infrastructures and the automation of processes. These requirements and emphasis is incompatible with the current status of some smaller data organisations.

The DSA formulates a minimum set of criteria for archiving digital research data. It adopts basic principles of digital preservation (distilled from NESTOR, DRAMBORA and TRAC) and combines them with data producers, sponsors and researcher perspectives. On the one hand it reduces the complexity of the OAIS structure, and the provision of detailed standards concerning data preservation processes. On the other, it goes beyond OAIS in bringing in new perspectives relevant to the provision of data services and allows greater flexibility and scalability.

In comparison with OAIS both of the examined data life cycle models (LIFE 2, CET) have certain weaknesses concerning their maturity and do not as yet have fixed standards for some important archival activities (e.g. CET does not address long-term preservation). However, there are also advantages as LIFE 2 and CET address different types of functions and can thus be used to supplement standards derived from the OAIS model. LIFE2 exceeds the OAIS model perspective by mapping the pre-archive stage and provides separate independent stages for data acquisition, elements of the cost model, and establishes career structures and appropriate professional roles. In contrast, CET provides realistic workload measures and includes development planning. The strength of both the LIFE 2 and CET models is also their higher flexibility and scalability.

Both approaches, i.e. digital preservation and data life cycle, are widely referenced within existing strategic documents, metadata standards, and the guidelines used by CESSDA and its member organisations. These materials including general standards (e.g. DDI), legislation relating to archives, various software tools such as NESSTAR have often been directly influenced by these models and related documentation.

---

berlin.de/moinwiki/WG_Trusted_Repositories_-_Certification: Data Seal of Approval has been created by DANS (Data Archiving and Networked Services) and it is available on-line: www.datasealofapproval.org

29 Digital Curation Centre (DCC) Curation Lifecycle Model, see http://www.dcc.ac.uk/lifecycle-model/ ; CET = The Cost Estimation Toolkit (CET) of the NASA: http://opensource.gsfc.nasa.gov/projects/CET/CET.php

30 There is also an on-line tutorial referring to the OAIS available on the ICPSR Web: Digital Preservation Management Tutorial: http://www.icpsr.umich.edu/dpm/

4.8. Establishing the cessda-ERIC minimal requirements and best practice standards

In their report Štebe and Duşa [ibid] concluded that:

No existing models fully conform to the aims of establishing a general ontology (dictionary, terminology, reference model) that would facilitate the mapping of current best practice used in social science data archives;

It makes sense to compare these models with existing guidelines and documents describing social science data archives routine activities, and extract a new specifically social science data archival reference model from this comparison.

4.8.1. Launching the cessda-ERIC and standards adherence as membership criteria

Annex 1 of an early version of the proposed Statutes for the new ERIC, set out the following obligation for Full members:

"To adhere to the OAIS reference model and/or the agreed CESSDA-ERIC seal of approval for archival practices" (CESSDA-PPP 2009b: Annex 1, point 1.2 h.)

In current situation this combination of two models allows meeting CESSDA priority aims concerning standards in the best possible way using existing tools (OAIS) or slightly modified versions of them (Data Seal of Approval):

The CESSDA-ERIC seal of approval should ensure minimal requirements concerning archival practices in the network and at the same time facilitate overcoming problems associated with the lack of scalability of the OAIS model.

The OAIS model offers a more complex framework for archival practices. It provides a meaningful set of requirements that are only fully applicable within large organisations. At the same time, a large part of the OAIS model (even that which lie beyond the minimum requirements of the OAIS model) is also relevant to organisations, which can not comply with requirements assuming of high organisational complexity and would not meet all of the criteria for certification, but which could use segments of the OAIS model for partial testing and making improvements.

At the same time the Service Level Agreement (a preliminary draft was distributed at CESSDA General Assembly in 2009 [CESSDA-PPP 2009a]) and the Statuses [CESSDA-PPP 2009b] specify general requirements concerning the compatibility of systems and procedures (e.g. communication and reporting procedures, required scope and level of services, development and usage of necessary common tools as multi-lingual thesaurus). Moreover, the current proposal of the Service Level Agreement also includes requirements concerning key performance indicators and values, which could be set up as basic quality requirements in relation to both, common CESSDA-ERIC services and national data provision services.

The Data Seal of Approval could be used as a starting point for development of the CESSDA-ERIC seal of approval. Table 1 summarises the requirements of the existing Data Seal of Approval. The CESSDA-ERIC seal of approval could be more specific concerning its aims and target groups; and thus could be also more specific in standards related to the

concrete tasks of the consortium. Its development requires a revision of current data seal guidelines and opens the possibility for greater elaboration and/or new additions.

We recommend establishing a group of experts to examine possible modifications of current guidelines by considering the concrete mission and working environment of the CESSDA-ERIC through inclusion of issues such as scalability of standards, and to take responsibility for the preparation of a proposal for the CESSDA-ERIC seal of approval.

Table 3. Data Archiving Standards for CESSDA

Quality criteria: The research data can be found on the internet; The research data are accessible, while taking into account ruling legislation with regard to personal information and intellectual property of the data; The research data are available in a usable data format; The research data are reliable; The research data can be referred to.
Guidelines: The data producer is responsible for the quality of the digital research data: The data producer deposits the research data in a data repository with sufficient information for others to assess the scientific and scholarly quality of the research data. The data producer provides the research data in formats recommended by the data repository. The data producer provides the research data together with the metadata requested by the data depository.
The data repository is responsible for the quality of storage and availability of the data: data management The data repository has an explicit mission in the area of digital archiving and promulgates it. The data repository uses due diligence to ensure compliance with legal regulations and contracts. The data repository applies documented processes and procedures for managing data storage. The data repository has a plan for long-term preservation of its digital assets. Archiving takes place according to explicit workflows across the data life cycle. The data repository assumes responsibility from the data producers for access to and availability of the digital objects. The data repository enables the users to utilize the research data and refer to them. The data repository ensures the integrity of the digital objects and the metadata. The data repository ensures the authenticity of the digital objects and the metadata. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.
The data consumer is responsible for the quality of use of the digital research data: The data consumer must comply with access regulations set by the data repository. The data consumer conforms to and agrees with any codes of conduct that are generally

> accepted in higher education and research for the exchange and proper use of knowledge and information.
>
> The data consumer respects the applicable licences of the data repository regarding the use of the research data.

Source: Data Seal of Approval, version 1.4 (DSA, 2009)

4.8.2. Development of Best practice standards

In the medium-term, we recommend developing a comprehensive framework of CESSDA Best practice including:

- Definition of minimal requirements for data archival practices in the form of a Seal of Approval. This would serve as a basic certification of ability and quality in social science data archiving and provide one criterion for CESSDA-ERIC membership;
- Provision of an in-depth set of cessda-ERIC materials (dictionary, terminology, standards, guidelines, reference model) defining best practice for social science data archives;
- Creation of a harmonised set of measuring methods and indicators allowing a mapping of activities to reference models, and control of compliance procedures and systems of standards;
- Provision of channels within the proposed training programme for disseminating knowledge and best practices related training;
- Establishing a system of expertise for the maintenance, continual development and updating of standards;
- Facilitate the adoption of prescribed minimal requirements through the provision of the proposed Toolkit to interested parties, e.g. a set of strategic document templates, contractual forms, generic guidelines, use cases, best practice solution scenarios, automated tools, metadata models, etc. [Štebe and Duşa 2009].

Such a strategy should be based on a proactive policy promoting the introduction and implementation of standards in archives of all sizes and at all stages of development. This approach will ensure an extensive, inclusive and professional pan-European network. To achieve these aims we recommend establishing a cessda-ERIC best practice working group that should ideally consist of experts on digital preservation and data activity planning and reflect the experiences of different types of organisation. The process of developing the cessda-ERIC best practice framework is outlined in greater detail in a report produced by Štebe and Duşa [ibid]. They propose 14 recommendations, which are summarised in the table below.

Table 4. Development of the best practices for the cessda-ERIC: recommendations from WP6, T4 (abridged)

> Establish a dedicated CESSDA best practice working group consisting of experts on digital preservation and data activities planning from a wide cross section of organisations.
>
> Engage in designing a first version of the Social Science Data Archives Data Activities Reference Model (SSDA DARM) and sustain its further development.

Make quality requirement proposals, mapped to the proposed SSDA DARM that contains gradation in service level. Set minimum requirements for certain types of organisation, balancing the needs and conditions of a variety of organisations so that general quality of service provision to users in terms of costs and funding opportunities is maximised.

Prepare a harmonised set of measuring methods and indicators of LS service activities related to the above mapping.

Monitor users demand and enable responsiveness to this through the introduction of specific services and products.

Maximise data access and minimise barriers by introducing and promoting open access arrangements, e.g. by introducing data specific Creative Commons licences, and shared responsibility models.

Introduce an additional layer of best practice principles to accommodate decentralised services. Divide requirements between the central hub and nodes.

Facilitate the adoption of minimum standards by producing a CESSDA toolkit: a set of strategic document templates, contractual forms, generic guidelines, best practice solution scenarios, automated tools, suggested metadata models, etc.

Design a generic list of template documents suited to the general requirements of a Social Science Data Archive.

Support organisations in the preparation, documentation, publishing, and implementation of changes in data activities.

Focus on the needs of medium and small organisations, with a less complex organisational structure and governance models, and with smaller throughput. Suggest alternative tools, ready made contractual models, etc. and organisational structures to accommodate a possible lack of resources whilst enabling minimum standards to be achieved. A local federation of nodes could allow for the self-organisation of specialised and common services suited to medium and small organisations.

General requirements should be kept simple and easily understandable, adaptable to variety of settings, while still aiming for quality products and services. Instead of lengthy (and by necessity resource demanding) certification of digital preservation, the proposed expert group could prepare a self-assessment procedure and help support and estimate the cost of improvements, where appropriate.

Assist in the professional development of employees within social science data archive organisations by providing training and online educational material (e.g. wikis, news, references to sources, legislations, standards, tools, etc.). This would assist in balancing general data service and preservation needs with subject specific expertise, thus better reflecting user need.

Source: Štebe and Duşa (2009)

5. Hardware and software requirements for data archiving
5.1. Aims and structure of report

In the anticipated highly collaborative environment of the new organisation new methods of support software production should be proposed. To achieve this it is necessary to move from single subject specialisation to collaboration with digital libraries and non-social science archives. One new organisation uses a different approach to models of supply or software tools that meet the needs of a broader community: The Data Archive Technologies Alliance (Gutmann et al 2009). The self-assessment Interim report (Fábián 2009) suggests that a collaborative framework should be built around the OAIS model (Figure 10). The same approach is used by The Data Archive Technologies Alliance for common wish list assessment, priority areas and pickups for implementation among existing software candidates.

Figure 10. OAIS design for ICPSR
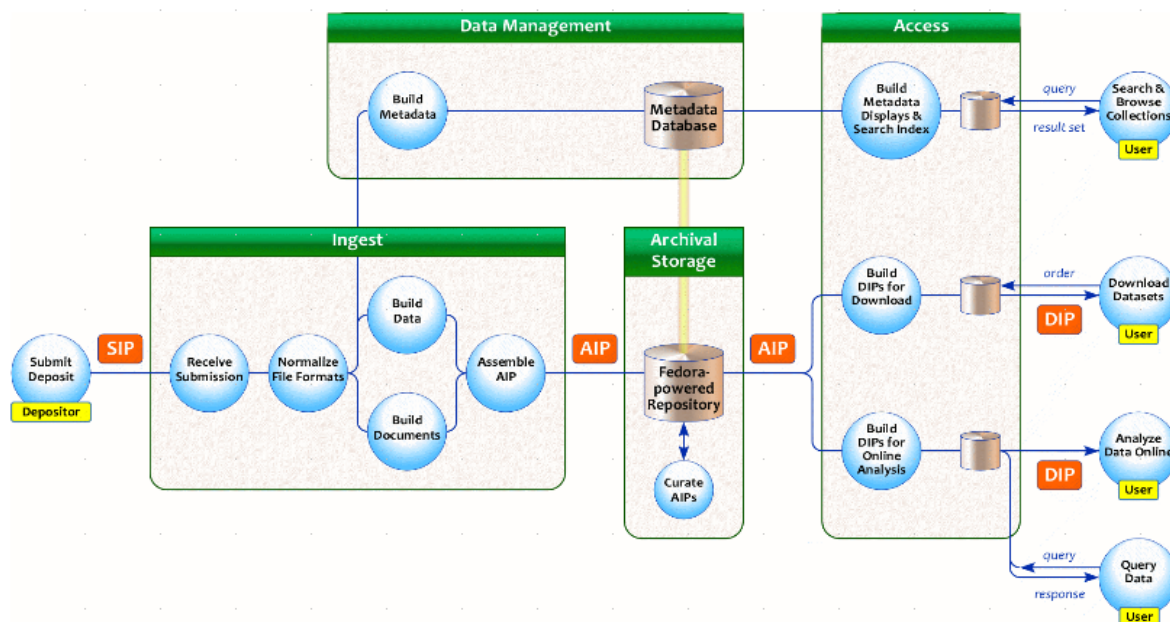
Source: (Gutmann et al 2009)31

The current section builds on the discussion from the WP6 Task 4 report: Development of a strategic framework for a CESSDA Toolkit. Of general interest is a list of tools that are generic to digital preservation life-cycle phases, e.g. the Digital Curation Tools list (http://www.dcc.ac.uk/tools/digital-curation-tools/).

Of specific interest are tools generated by the DDI community and which support the DDI metadata standard. Indeed, the DDI Foundation Tools Program working principles (http://tools.ddialliance.org/?lvl1=ftp&lvl2=overview) are similar to those of the Data Archive Technologies Alliance (despite that latter's broader outlook). We recommend that the ERIC develops its future plans in collaboration with similar initiatives and organisations.

---

31 We acknowledge the contribution of Cole Whiteman, who designed the graphics for the original presentation. Reproduced with special permission from the authors.

5.2. Related CESSDA PPP workpackage recommendations

In this section we will develop specific recommendations regarding software needs of CESSDA-PPP partner organisations and a strategic framework for the proposed CESSDA Toolkit. The report for WP6, task 4 states the relevant OAIS functional entities as being:

- Ingest;
- Archival storage;
- Data management;
- Administration;
- Access.

The Toolkit report identified the following critical areas requiring additional software creation, configuration and maintenance, or support in existing software implementation:

- "lack of appropriate tools for DDI" (A, B, C)
- "user authentication system" (E)
- "OAI-PMH exposed data catalogue" (E)
- "Lack of tools to produce in-depth data descriptions" (B, C)

Sharing of proprietary tools that were created at particular CESSDA member organisations is not always straightforward due to unresolved technical issues, IP rights and language problems.

Hardware prerequisites

There are different types of software for data archiving, often requiring very different hardware resources. Software development is especially fast-paced, particularly because new versions and updates are released frequently. It is therefore recommended that future software development needs are anticipated during a hardware upgrade.

As noted in the WP6 task 10 report (Duşa, 2009), different software suites are often required to meet higher OAIS standards. Even though CESSDA data archives seem to prefer open-source software, sometimes there is no alternative to commercial software. For example, database software like MySQL is highly regarded by Developers but once there is a large number of datasets (e.g. 1000+) and many hits per second, software such as Oracle software is more fit for purpose.

There are also different hardware requirements for meeting minimal and higher standards. For example, security issues regarding sensitive data (e.g. individual anonymised records) are often resolved by using different servers for the web front end and for data storage. This solution keeps data secure but sees an increase in hardware costs as different servers are required for different archiving activities.

5.3. General Digital preservation tools

Of a general interest is a list of generic tools relating to digital preservation life-cycle phases, e.g. the Digital Curation Tools list (http://www.dcc.ac.uk/tools/digital-curation-tools/). An overview of the tools and software is shown in Figure 2.

One evaluation of digital preservation software is contained in a recently published

UNESCO commissioned report (Bradley et. al. 2007). The report concluded that currently there was no single open source software that would cover all OAIS functional requirements. They suggested collaborative work to rectify this, and acknowledged the increase in OAIS compatible software by repositories such as Fedora and Dspace, who are also listed on a Data Archive Technologies Alliance watch list. RODA has also announced a new software development that supports most of the digital preservation process: http://roda.di.uminho.pt which could be investigated further by the ERIC.

Figure 11 Lifecycle of Digital Curation Tools (Digital Curation Centre 2009)[32]
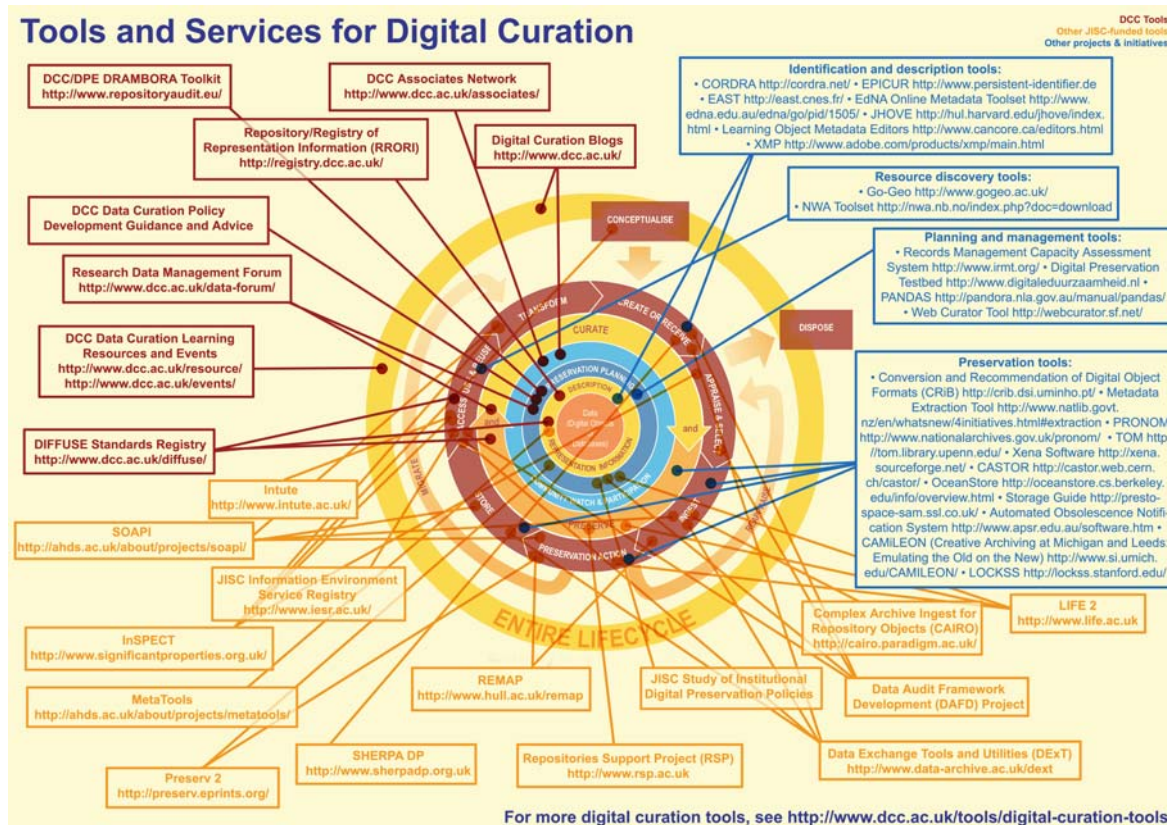
5.4. Specific Data Archive tools

DDI Foundation Tools Program: Overview

Specific to DDI metadata standard are tools that are generated by the DDI community. The DDI Foundation Tools Program working principles are similar to those of The Data Archive Technologies Alliance. See: http://tools.ddialliance.org/?lvl1=ftp&lvl2=overview

Both aim at the widespread adoption of the DDI standard based around the assumption that the sharing of resources and coordination of software development around a common standard benefits all compliant organisations.

Tools Alliance

CESSDA needs to establish close working links with groups and organisations which are

---

32 Special thanks to Sarah Higgins from DCC who gave permission to use this figure.

committed to the development of open source tools for data management. One reason for establishing a Tools Alliance is the effectiveness of working in partnership to develop required special interest, data archive orientated, software and tools which are often overlooked by commercial software development (Gutmann et al. 2009)

Inside a preservation life cycle the following priority areas were identified, closely related to complex organisational workflows (see Figure 25):

Ingest tools
AIP Creation-Validation
SIP Creation-Validation
DIP Creation-Validation
Audit tools

Common DDI standard related planes focus on:

Tools for full variable-level metadata creation not dependent on proprietary software (such as SPSS)
DDI Editor
DDI Converter
DDI 2 to 3 translator

During 2009 the Alliance announced plans to raise funds and find common ground for development work.

5.5.   Discussion and Conclusions

Previously data archive software was often custom-made. However, with the subsequent adoption of common descriptive metadata (the DDI standard) the development of shared tools became more feasible and important. The DDI metadata editor (the statistical package file conversion tool) and data publishing tool was part of the NESSTAR package. Experience of sharing tools still remains limited to partial phases and processes. Often available functionality was too limited and many organisations developed their own parallel systems for, e.g. the publication and access of catalogued digital objects. Custom made applications circumvented the need to use pre-existing, often inflexible, software.

There are more choices today: NESSTAR, which is a commercial operation offering subsidised packages for smaller organisations and Dataverse, which is open source. A range of commercial and open source repository management and preservation solutions are also available, e.g. Fedora, Dspace, LOCKSS.

The future collaborative environment of the ERIC will enable more informed decisions about the use of available software and more co-ordinated new software development. The ERIC should provide financial and personnel support for active participation in initiatives such as the DDI Foundation Tools Program and The Data Archive Technologies Alliance. The governance of the program needs to reflect the development priorities of different types of organisation. Support will be needed when new tools require implementation.

Current experiences with both NESSTAR and Dataverse proved that the introduction of such software is quite demanding, both technically and in relation to installation and ongoing support, despite the use of common standards. DDI3.0 should therefore be introduced gradually, its implementation co-ordinated, whilst bearing in mind future

collaborative work and ensuring compatibility of model conventions.

We propose the following recommendations for the ERIC:

- Recognition of the importance of global co-operation and a full contribution toward a Tools Alliance;
- Internal development or external collaboration with appropriate groups and organisations to ensure the availability of priority tools, e.g., for preservation, harmonisation, multilinguality, common access and user authentication.
- A clear strategy and guidelines for collaborative work are developed;
- Financial resources are ring fenced for collaboration with the Alliance, bearing in mind that work in isolation would be more costly;
- Support teams are set-up for the introduction and adaptation of new tools by all member organisations.

5.6.   Scenarios

In the planned ERIC there will be at least two layers of membership, designed according to different standard levels for professional data archiving activities.

To reach the minimum standard, only certain hardware requirements are needed, while for more advanced standards in data archiving there are more hardware related requirements.

The hardware needed to set up a data archive depends not only on these tasks, but also on the strategic decisions made about types of software for installation. Some software requires more hardware resources, while some others require far less.

With this in mind, there are two possible hardware requirement scenarios:

- A minimalistic scenario, with minimal standards and minimal software;
- An advanced standards scenario, with more requirements both in terms of hardware and software.

5.7.   Research on current practice

For both of the WP6 Tasks 10 and 11, we used a series of interviews with technical personnel from ten of the participating project partner organisations and ICPSR, the largest data archive in the world. Interview participants came from a diverse pool of institutions, both in terms of size and geographical location. Annex 1 lists those data archives which participated in the interviews. Interviews were conducted using Marratech software and notes were taken at the time of interview using the built in whiteboard. Notes could then be shared with interviewees.

5.8.   Software options

In general, there are only three broad categories of software that can be used, for any kind of activity:

- Proprietary commercial software, interested users have to buy the software from a private company;

- Proprietary free software, that does not cost anything but users cannot modify or extend it;
- Open-source software that users can freely decide on modifying, according to their particular needs.

At the beginning of the Internet era and mass-produced computers, the vast majority of software was commercially produced, starting with operating systems, and, with very few exceptions, directed to highly technical users. Proprietary software can shift between being free or commercially available, according to user demand.

Since the beginning of the open-source movement, much software has become available and easily usable for almost anybody with minimal computer literacy. Actually, open source software is often so well developed that it has the potential to successfully replace proprietary software. Currently, institutions surveyed employ both commercial and open-source software, and wherever a web server is involved, our research suggests that at least one open-source piece of software is employed (e.g. Apache, or Tomcat, or MySQL).

Some software performs faster than others, and some consume more resources than others. Although there is a direct relationship between speed and the amount of resources allocated to certain software packages, this is not a hard and fast rule: there is software that performs very quickly, without consuming a lot of hardware resources; inversely there is also software that performs slowly whilst consuming a lot of resources.

There is no benchmark for comparing the speed and resource consumption of commercial and open-source software. However, we contend that commercial software is more resource intensive than any open-source counterpart.

For data archiving purposes, there are two main categories of software employed by CESSDA institutions:

- Commercial software: Windows, IIS, Microsoft SQL Server, Nesstar, etc.
- Open source software: Linux, Apache, Tomcat, MySQL, PHP etc.

5.9. Hardware requirements

The ERIC should have at least two types of standards for institutions to meet: minimal and advanced.

Even though basic data archiving activities will remain more or less the same, there will be a considerable difference in terms of standards compliance. Due to safety and security issues regarding the protection of sensitive data, hardware requirements are likely to become more expensive, both in terms of the cost of individual machines and with regard to the number of machines required for ensuring high-level standard compliance.

Minimal Standards
It is perfectly possible to imagine a small data archive operating with a single server which performs all data archiving services, from the public web page, to variable searches, to particular statistical operations for individual datasets. Commercial and open-source software alternatives are available for all of these tasks.

Any modern hardware machine could perform all these tasks, but to cover all possible

resources required existing software the following configuration would be required:

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

Higher (OAIS) standards

The hardware requirements described above would still be required. However, because it is mandatory for services to be run on different machines, the number of servers required needs to be multiplied by a factor of at least four.

As noted briefly earlier, it is possible that any internet user could gain access to the webpage server. Malicious hackers could exploit weak operating systems and take control of the server. Should this happen, data stored on this server could be at risk of exploitation. Therefore data (and particularly sensitive data) should be stored on a different machine. If security was breached and a web server hacked into, the data storage server would remain safe. Furthermore, hardware failure is prevented by the use of multiple servers and machines – if one server fails, another can perform its operations.

A high-level data archive should therefore have at least four operational servers:

- Web front-end server;
- Data storage server;
- Preservation server to hold long-term archival copies of the data;
- Processing server, for statistical analysis and software development.

These four servers could be installed either on different physical machines, or on different virtual machines on the same physical machine. In the latter case the technical specification and capacity of the physical machine needs to be a lot higher than a normal server, so that the virtual servers are accommodated. Furthermore, in order to prevent hardware failure, a second identical physical machine is needed to perform a live back-up in case the first machine fails to operate.

6. Step-by-step guide to publishing a dataset to the CESSDA portal

6.1. Summary

Even though all CESSDA member institutions have the same basic objective (some have more diverse activities than others), namely data archiving, the *modus operandi* differs substantially in each and every case. We can definitely state there is a definite heterogeneity between CESSDA archives in this respect.

The data archiving world is continually developing and refining professional standards (Krejci, 2009), and CESSDA institutions are more or less compliant with these. Regarding the implementation of these standards, the current perception is that every data archive has a local procedure, depending on particular software tools, servers and expertise.

This heterogeneity is natural, given the different conditions each country has experienced

during their construction phase and subsequently when operating. However, after the CESSDA-PPP project the new infrastructure will have to present some recommendations regarding a "CESSDA way" of doing things - if not an exact procedure then at least a general direction.

For this reason, WP6 organised a series of interviews with technical personnel from more than half of the CESSDA membership and one organisation outside Europe, in order to learn about individual experiences and try to summarise the European web publishing landscape.

Data archives who responded to our interview questions are listed in Appendix 6.
One of the key questions tried to determine not if there are differences between data archives (because this is obvious), but to investigate whether these differences are related to the size of the data archive concerned.

6.2. Common tools and procedures vs. common agreed standards

One natural question regarding the "CESSDA" way of doing things is related to this very "way": is existing practice the best for everyone, cover all possible local needs (or even completely replace local systems with a standard system for all data archives), or should it be a system that can communicate with other systems by use of a common input-output communication standard?

It would probably not be feasible to jump from extreme heterogeneity to perfect homogeneity. However, some things will need to be implemented in order to achieve inter-operability. There can be no "best" operating procedure, for the simple reason that very similar results can be achieved with very different tools.

Additionally, some user-friendly tools are frequently found in commercial software products. However, newly developed institutions are often unable to afford either the software or any required hardware. There are different needs for different institutions and even the global data archiving landscape is rapidly changing: for these reasons, we estimate that different solutions will always exist and there is no reason to change this situation.

Recommendation
The CESSDA-PPP and the future ERIC should encourage the development of different software solutions (both commercial and open-source), under a common framework of standard input-output communication standards.

6.3. Operating systems

There are many differences between how institutions approach operations and publishing procedures. For example, one of the first basic things to consider is the operating system: which one is "best"?

This is an open-ended question, with no easy answer. As in the case of general commercial software versus open-source software, it is (amongst others) an "ease-of-use" versus "power" issue. If this is more than apparent with regard to individual workstations (although Linux is nowadays becoming more than user-friendly), in the case of web servers there is no obvious solution(s). Most of the web servers worldwide use one of the various types of Linux operating system - Linux's popularity being particularly

understandable given its invulnerability against malicious automatic viruses.

Table 5. Operating systems employed by survey participants

Organisation	Windows only	Mixed	Linux only
FSD		x	
SND	x		
EKKE		x	
ADPSS	x		
TARKI			x
FORS		x	
GESIS	x		
ADP		x	
ISSDA	x		
ICPSR		x	
UKDA	x		

In Table 5, it can be seen that most of the data archives interviewed use Windows for at least as one of their main operating systems, if not for all data archiving and publishing procedures. In the case of GESIS, an IBM Unix machine is used, not as the main operating server, but for marginal activities such as the preservation of downloadable PDF files.

Generally, it is accepted that Windows servers are easier to administrate than Linux servers (which require quite a bit of IT expertise). In other respects Linux is also highly regarded, but Windows is still used for no other reason than operational continuity and exploitation of expertise accumulated over the years.

If they were to start a new data archive, many respondents indicated a preference for Linux operating systems. In practice, half of the data archives employ a mixed environment, with both Windows and Linux machines. In some cases, Windows is only used for the purpose of running the Nesstar Server, which is unanimously accepted as the most user-friendly software for publishing activities.

6.4. Software

Just as in the case of operating systems, software used for data archiving activities is as diverse as the number of institutions interviewed. There is also a mix of commercial and open-source software employed.

Two of the most common well-known myths about open-source software are related to: a) reliability and; b) post-installation support.

It is a common belief that proprietary software is more reliable simply because there are hired people working on the software, further developing it and making sure that significant problems (bugs) are eliminated. However, neither of these arguments are convincing. In relation to open-source development, there are indeed people who work without being paid. There is also the question of the number of developers: on the commercial side there are a limited number of programmers, while literally thousands of people contribute to open-source development. Identifying and solving bugs is also much easier in an open-source environment: many of the testers not only report bugs but also

propose solutions and amendments to the (open) code.

Support is also a big issue. It is often assumed that open-source software is unsupported, and therefore unsuitable for well established, mature, organised and standardised institutions. Whilst this might be true for some open-source software, it is not a generally true. In fact, there are many examples of commercial software with poor support and open source software with excellent support. For example, few SPSS users actually use the SPSS official help (and therefore they do not benefit from actual "official" support). On the other hand many R users receive answers to their questions almost instantly (sometimes even on Sunday evenings!).

Returning to data archiving software, both less resourced and well established data archives employ a mixture of both commercial and open-source software. For example, it is hard to imagine these days how information on the web could be published without well known servers like Apache, or why MySQL would not be considered even for relatively complicated database structures, or how would we cope if Java were to disappear.

Some open-source software or environments have almost become the "standard" tools that are used in various proportions by all institutions. Linux itself is one example of a well developed open-source initiative. The most heavily used open-source software includes (in random order): HTML, Apache, Tomcat, PHP, Python, Java, Genshi (used by Turbo Gears), CVS, MySQL, SQLite. PostgreSQL, Perl, Zope, Plone or Django (CMS – Content Management Systems), DataVerse, Lucene (providing Solr, a search engine), XML, and R.

On the commercial / proprietary side, there are various Windows based applications like Oracle or DB2 (large scale database systems), Microsoft Access for smaller scale databases, Windows' IIS, .Net, Oxygen (an XML editor), the "serious" suite of commercial statistical engines like SPSS, SAS, or Stata plus the well known Nesstar Server and Publisher.

In this vast expanse of software possibilities, probably the most comprehensive advice is given in the Task 12 report from WP7 (Dusa, 2009).

Recommendation
Where cost is an issue, new data archives should turn to open-source software. Where ease-of-use is a factor, new data archives should investigate commercial software alternatives.

Even if cost is not an issue, many data institutions (including large and well developed archives) still use open-source software for the simple reason it is well developed enough to satisfactorily solve many problems.

6.5. Producing DDI files

As expected, this is one area of expertise that reveals large differences between the data archives. The exception to this is Nesstar Publisher, which, having developed from CESSDA collaborations, with EC funding, seems to be preferred by many institutions.

The actual process of producing DDI-XML files varies between archives, even though many of them use Nesstar Publisher in one phase or another. Almost every institution participating at this survey maintains a relational data base, external to the Nesstar suite.

Among data base (DB) software preferences, Oracle, PostgreSQL, Microsoft SQL Server, MySQL or even Microsoft Access for smaller scale institutions, are used. Where thousands of data sets are involved, Oracle or Microsoft SQL Servers are preferred; the other data archives use another DB software solution.

Almost every data archive starts with an XML template file containing a DDI2 compliant structure, which is usually produced automatically using the information stored in the relational DB. This process is based on highly customised scripts in different languages, in order to fit particular workflows developed over the years (e.g. Visual Basic, use of a OAI-PMH server, specially constructed software like Hermes or even specially developed in-house code).

Once generated, the template is further manually updated with information using either specialised XML editors (Oxygen, XML Spy) or Nesstar Publisher, after which a full DDI file is produced, containing metadata for both study level and variable level information. Nesstar Publisher offers the possibility to integrate both data and metadata in the same file (a proprietary NSDstat format). With only one exception, all data archives prefer storing data and metadata as separate files.

Whilst the preferred format for metadata is an XML file with a DDI2 structure, data are preserved in various formats: from the convenient SPSS file (either the last version or the portable format), to the platform and software independent ASCII format, with columns usually delimited by a tab sign. For long term preservation, ASCII seems to be the preferred format, even though it does not currently support all UTF character codes (there are solutions for this too, most notably a special ASCII file with any special characters replaced with their corresponding escape codes).

> **Recommendation**
> Data and metadata should be preserved in separate files.

6.6. Publishing procedures

Just as in the case of DDI files, publishing final data and metadata files on a web server is a highly specialised process, varying between one data archive and another. However, producing a DDI-XML file is an intrinsic part of the publishing process, therefore many of the software and procedures presented in the previous section remain the same. There are two main publishing activities for almost all data archives, with the exception of those using the Nesstar Server as their main interface:

- Publishing a small size metadata file in the web catalogue, containing very basic (and mandatory) fields for all studies (i.e. title, study number, etc.);
- Publishing a complete metadata file for browsing purposes.

Even though the Nesstar Server provides a sophisticated catalogue for all browsing purposes (from minimal information to full variable level information), this is a comparatively recent development compared to the timeframe of traditional catalogues, where data archives usually publish information about datasets.

This tradition is hard to replace, especially because the Nesstar Server is a self-sufficient, fully fledged and tightly integrated system. Many of the respondents to this survey mentioned difficulties in integrating their traditional website (including catalogue) with the

new Nesstar Server.

For this reason, most of the data archives still maintain a separate catalogue in a traditional format (different in every case), and in some cases another catalogue in the Nesstar Server (basically the same information plus a wealth of other information down to variable level). The format of the CESSDA Portal catalogue is yet to be decided - this makes the task of proposing a step-by-step guide almost impossible. For the moment, however, given that the CESSDA catalogue is currently served by the Nesstar Server, proposing a guide is very simple:

- Create a DDI-XML file using the Nesstar Publisher, eventually fine-tuning it with various XML editors;
- Publish the information in the Nesstar Server using the highly developed Nesstar Publisher tool.

However, if the CESSDA catalogue will be based on a different, new structure (other than the Nesstar Server), then an appropriate new procedure will be required. Although a very sophisticated tool, Nesstar Publisher has one crucial weakness: it publishes only to the Nesstar Server. It can be used perfectly to construct a DDI-XML file, but it cannot be used as a standardised publishing tool, which is valid for all kinds of data providing servers. It is true that such a server does not even exist (this is actually the reason why each data archive employs a different system), but the goal of the future cessda-ERIC should be the creation of a standard server with very precise data and metadata publishing procedures, making use of standardised communication protocols.

> **Recommendation**
> The CESSDA Portal should use an open standard to publish data and metadata to its catalogue, use plain XML files for metadata and ASCII format for the data files.

In brief, such a procedure could be summarised as follows:

Create the data and metadata files in a standard, agreed format; the data archives are free to choose their own tools and procedures, as long as the two files are produced in a standard format. Use standard communication protocols and a commonly agreed procedure to publish those files on to the CESSDA portal.

Details about standard publishing processes will only be available following decisions regarding the format of the CESSDA Portal catalogue.

7.   APPENDICES

Appendix 1. WP6 Procedure for drafting CESSDA training and staff exchange programmes
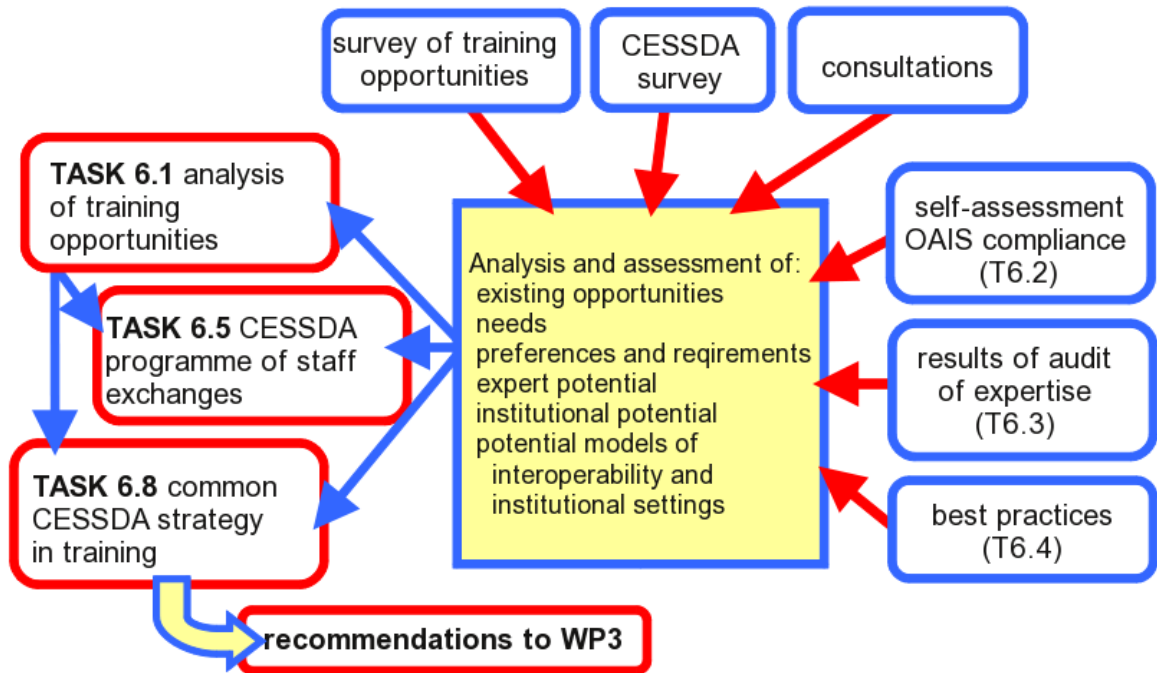
Source: WP6 task 1.

Appendix 2. CESSDA Survey questions on general and training issues

Basic information about your organisation
2. How would you describe your organisation?
Data Archive (CESSDA member)
Data Archive (non-CESSDA member)
Research Institute
Research data repository
Other

4. Is your organisation:
Legally independent
A department within a university
A department/section of your national archive
Other
Other (please specify)

5. Does your organisation have any of the following:
1. Structural diagram ....YES/NO
2. Strategic plan ....YES/NO
3. Business plan ....YES/NO
4. Training plan ....YES/NO

8. Does your organisation take part in any of the following activities?
Data exchange agreements ....YES/NO
Training on research infrastructures ....YES/NO
Implementing R&D projects ....YES/NO
Disseminating information at national level...................................YES/NO
Disseminating information at European or International level ....YES/NO
Public consultations or deliberations..YES/NO
Preparation of new data resources (e.g. teaching datasets) ....YES/NO
Other (please specify)

9. In the year 2006 - 2007, how many staff were employed in your organisation as:
Full-time data professionals
Full-time other staff
Part-time data professionals
Part-time other staff

10. Has the number of your data professionals changed during the last 2 years? If so, please quantify and say whether there has been an increase or decrease in staff.

13. Does your archive employ any of the following specialists for data archiving and management?
Archivists...YES/NO
Librarians...YES/NO
Information scientists ....YES/NO
Information technologists..YES/NO
Statisticians..YES/NO
Survey research methodologists ....YES/NO
Sociologists..YES/NO
Other social scientists ....YES/NO

Human scientists...YES/NO
Others, please specify below ....YES/NO

14. In the year 2006 - 2007, did your archive systematically collaborate with specialists from other departments or teams within your organisation or outsource work to fulfil data archiving tasks?
Please do not include short term, occasional collaborations and if you answer yes, indicate the specific area(s) of expertise in the comment box.
Yes, the archive collaborated with experts from different departments or parts of our organisation
Yes, the archive collaborated with external experts
Yes, the archive collaborated with both experts from different departments and external experts
No
Comment

15. Briefly describe your training plan (e.g. areas of training, frequency, priorities, methods, assessment, funding)

16. How many of your data professionals have received systematic, job-related training (internal, external or both) in the last year?

17. What types of job-related training and professional development activities have your employees participated in during 2006-2007?
In-house training seminar ....YES/NO
External company-paid seminar ....YES/NO
Individual participation in external training seminar or programme paid for from predominantly external sources...YES/NO
Mentor training with internal expert(s) ....YES/NO
Mentor training with external expert(s)..YES/NO
Visiting programme or time spent at other organisations................................YES/NO
Other (please specify below)..YES/NO

18. Please list the number of your staff who have taken part in each of the following types of training in the last year 2006 -2007.
In-house training seminar
External company-paid seminar
Individual participation in external training seminar or programme paid for from predominantly
External sources
Mentor training with internal expert(s)
Mentor training with external expert(s)
Visiting programme or time spent at other organisations
Other (please specify)

19. For 2006-2007, has your archive organised any job-related training seminar or programmes that were also made available to external participants?
Lead by internal or domestic experts for a domestic audience
NO/YES, ONE/ YES, MORE THAN ONE
Of international scope
NO/YES, ONE/ YES, MORE THAN ONE

20. List all external job-related training seminars or programmes in which your employees were trained in 2006-2007.

21. List all internal job-related training seminars or programmes in which your employees were trained in 2006-2007.

22. During the last 3 years, have any of your employees undertaken the role of lecturer in any training programmes or seminars organised by other institutions?
no
yes, occasionally
yes, often

23. Does your archive provide online tutorials or other web-based learning materials for data professionals?
no
yes

24. Please give URL(s) for your web-based and online training materials (if any).

25. Do you expect that reaching and maintaining appropriate standards for digital archiving (e.g. OAIS compliance) and/or the implementation of new technologies in connection with the CESSDA PPP will require new skills to be developed beyond your existing training capacity?
No need for additional training beyond current practices
Available expert knowledge and training opportunities are good but require some improvement
Available expert knowledge and training opportunities require significant improvement
Improvements to available expert knowledge and training should have the highest priority

26. List up to 5 priority areas for external training opportunities needed by your organisation to reach and maintain appropriate standards of digital archiving and/or to implement new technologies in connection with the CESSDA PPP

27. If offered by CESSDA, which of the following methods would you use for training and professional development of your employees?
CESSDA training seminars n
CESSDA summer school or special training camps
Visiting programs in other CESSDA archives
Staff exchange programme
On-line tutorials or learning guides
Virtual Centre of Competence (VCC)
More CESSDA expert workshops

28. Do you have any other method of training or professional development you would like to propose for inclusion in the CESSDA training programme? If yes, please specify.

29. Please comment on conditions under which your organisation would be willing to host data professionals from other data archives to provide them with know-how concerning practices in your archive (for example, funding conditions, maximum numbers of visitors per year, time limit on length of visits).

Appendix 3. Table 6. URLs of English language catalogues

Archive	English on-line catalogue
CNRS-RQ	NO
DISC	http://www.ssd.gu.se/?p=search
ADPSS	http://www.sociologiadip.unimib.it/sociodata/eng/index.php?w=cerca
GESIS	http://www.gesis.org/en/services/data/retrieval-data-access/data-catalogue/
EKKE	//www.gsdb.gr/scripts/en/search_ds_en.pl?code=15
ISSDA	http://www.ucd.ie/issda/data.htm
RODA	Not available temporarily
ARCES/CIS	http://217.140.16.67/cis/opencms/EN/1_encuestas/catalogo.html
ADP	http://www.adp.fdv.uni-lj.si/opisi2/index.xml
DDA	NO
SDA	http://archiv.soc.cas.cz/czech/archivtop-f.html
FSD	http://www.fsd.uta.fi/english/data/catalogue/index.html
TARKI	http://www.tarki.hu/en/services/da/da_services.html
UKDA	http://www.data-archive.ac.uk/search/searchStart.asp
CEPS/INSTEAD	NO
DANS	http://www.dans.knaw.nl/en/data/
NSD	http://www.nsd.uib.no/nsd/english/datatjenester.html

Table 7. URLs of NESSTAR servers

Archive	NESSTAR
CNRS-RQ	NO
DISC	http://nesstar.snd.gu.se/webview/
ADPSS	http://149.132.156.160/webview
GESIS	http://zacat.gesis.org/webview/
EKKE	http://194.177.213.23/webview
ISSDA	NO
RODA	NO
ARCES/CIS	NO
ADP	http://nesstar2.adp.fdv.uni-lj.si/webview/
DDA	http://nesstar.dda.dk
SDA	http://147.231.52.119/webview/
FSD	http://fsd2.uta.fi:8080/webview/index.jsp
TARKI	http://www.tarki.hu:8088/webview/
UKDA	http://www.data-archive.ac.uk/orderingData/exploreOnline.asp
CEPS/INSTEAD	NO
DANS	http://nesstar.steinmetz-archief.nl/webview/index.jsp
NSD	http://nsddata.nsd.uib.no/webview/

Appendix 4. CESSDA Survey Questions about expertise issues

9. In the year 2006 - 2007, how many staff were employed in your organisation as:
Full-time data professionals
Full-time other staff
Part-time data professionals
Part-time other staff

10. Has the number of your data professionals changed during the last 2 years? If so, please quantify and say whether there has been an increase or decrease in staff.

11. Please give the number of your employees with the highest educational levels listed below.
Some high school or less
Graduated from high school
Attended some college courses
Two year degree
Three year degree
Four year degree
Postgraduate study without degree
Master's degree
Doctoral degree

12. Please give the number of your employees with the following skills.
Elementary SPSS
Moderate SPSS
Strong SPSS
Elementary SAS
Moderate SAS
Strong SAS
Elementary R
Moderate R
Strong R
Elementary MS Access
Moderate MS Access
Strong MS Access
Other database or statistical analysis software

13. Does your archive employ any of the following specialists for data archiving and management?
Archivists...YES/NO
Librarians..YES/NO
Information scientists ....YES/NO
Information technologists..YES/NO
Statisticians...YES/NO
Survey research methodologists..YES/NO
Sociologists...YES/NO
Other social scientists...YES/NO
Human scientists...YES/NO
Others, please specify below ....YES/NO

16. How many of your data professionals have received systematic, job-related training

(internal, external or both) in the last year?

Appendix 5 The original expertise task questions

Survey questions to directors:

How many are people working in the archive?
....

Out of which, how many in:
FTE (30 hours or more per week):.......................................
PTE (less than 30 hours per week):....................................

Please describe your data archive in terms of departments and/or sections
....
....
....
....
....
....

Survey questions to employees:

Which department or section do you work in the archive (if any)?
....

What is your current position at the archive?
....

Do you work full-time or part-time?
1. full-time
2. part-time

What is the highest educational level that you have attained?
1. Some high school or less
2. Graduated from high school
3. Attended some college
4. Two-year degree
5. Four-year degree
6. Post-graduate study without degree
7. Master's degree
8. Doctoral degree

Which field have you earned your degree(s)?
....
....
....

Have you received any special training or additional qualification after graduation?
Yes
No --> Go to question Q11

In which field?

How would you rate your knowledge of:

SPSS for Windows	elementary	moderate	strong	none
SAS	elementary	moderate	strong	none
R	elementary	moderate	strong	none
MS ACCESS	elementary	moderate	strong	none
Other(s)	elementary	moderate	strong	none

Appendix 6. Names of CESSDA member organisations (short name and country)

CESSDA member organisations who responded to survey request(s):

ADPSS-Sociodata (ADPSS; Italy)
Central Archive for Empirical Social Research (GESIS-ZA; Germany)
Centre of Sociological Research (CIS, CESSDA archive is ARCES; Spain)
Danish Data Archives (DDA; Denmark)
Data Archiving and Networked Services (DANS; Netherlands)
Finnish Social Science Data Archive (FSD; Finland)
FORS Data Archive (Switzerland)
Greek Social Data Bank (GSDB; Greece)
International Network for Studies in Technology, Environment, Alternatives,
Development (CEPS/INSTEAD; Luxembourg)
Irish Social Science Data Archive (ISSDA; Ireland)
Norwegian Social Science Data Services (NSD; Norway)
Réseau Quetelet (CNRS-RQ; France)
Romanian Data Archive (RODA; Romania)
Social Science Data Archives (ADP; Slovenia)
Sociogical Data Archive (SDA; Czech Republic)
Swedish National Data Service (DISC; Sweden)
TÁRKI Joint Research Center (TARKI; Hungary)
UK Data Archive (UKDA; UK)

CESSDA member organisations who did not respond to survey request(s):

Estonian Social Science Data Archive (ESSDA; Estonia)
Vienna Institute for Sociological Documentation and Methodology (WISDOM; Austria)

Appendix 7: List of data archives interviewed (in random order)

- FSD – Finland
- SND – Sweden
- EKKE – Greece
- ADPSS – Italy
- FORS – Switzerland
- TARKI – Hungary
- GESIS – Germany
- ADP – Slovenia
- ISSDA – Ireland
- ICPSR – USA
- UKDA – United Kingdom

Appendix 8. Answers from the WP6 survey re: Full-member checklist*

No	Requirements	Barriers	RODA - Romania	GESIS - Germany	SDA - Czech Republic	TARKI - Hungary
1	Make a full contribution to the cessda-ERIC budget.	Inadequate national funding	4	2	?	4
2.1	Be fully compliant with the DDI metadata standard.	Inappropriate training, lack of knowledge of standards	4	1	2	2
2.2		lack of appropriate tools for DDI	4	2	1	2
3	Adopt and apply the cessda-ERIC common Single Sign-on user authentication system.	Lack of IT specialists who can configure and maintain single sign-on user authentication systems	3	2	2	3
4	Enable the harvesting of all catalogue records for inclusion in the cessda-ERIC data portal	Lack of IT specialists who can configure and maintain an OAI-PMH server	3	2	2	2
5.1	Make data holdings downloadable through common data gateways	Legal and technical barriers: lack of IP rights management, no contracts for distribution.	3	2	2	3
5.2		Same barriers as described in points 3 and 4 above.	3			2
6	Maintain their local language(s) within the multi-lingual thesaurus.	Insufficient number of staff, inadequate training for maintaining localised ELSST	4	3	3	3
7.1	Share data archiving tools (under the IP conditions set out in the Intellectual Property Rights annex to these	No tools to share	3	1	2	2

No	Requirements	Barriers	RODA - Romania	GESIS - Germany	SDA - Czech Republic	TARKI - Hungary
	statues).					
7.2		Tools available only in local language	3	1	4	4
7.3		Unclear IP rights	3	1	2	1
8.1	Adhere to the OAIS reference model and/or the agreed cessda-ERIC Seal of Approval for archival practices.	No clear standards are set by the cessda-ERIC (Recommendations are expected from WP6, Task 9)	N/A	N/A	N/A	N/A
8.2		Lack of awareness/knowledge among management team(s)	3	3	3	3
8.3		Inadequate control of intra-institutional processes	3	3	3	2
9	Contribute to the cessda-ERIC's cross national data harmonisation activities.	Insufficient number of staff, inadequate professional background and experience.	4	2	3	3
10.1	Contribute material and/or expertise to the cross-national question bank.	Lack of tools to produce in-depth data descriptions,	4	1	3	2
10.2		Lack of human resources: insufficient number of properly trained staff.	4	3	3	4
11	Help and support designate members of the ERIC to achieve Full or Associate membership.	Insufficient number of properly trained staff / inadequate funding for providing support	4	3	3	4
12.1	Where possible, facilitate access to national government (and	Inappropriate national legal systems.	4	4	3	2

No	Requirements	Barriers	RODA - Romania	GESIS - Germany	SDA - Czech Republic	TARKI - Hungary
	research-funded) microdata.					
12.2		Lack of awareness of, or unfavourable attitude toward, open access policy by legislators and governmental officials.	4	3	3	4
		Total	**67**	**39**	**44**	**52**

* (Scores: 4 – greatest problem; 3 – severe problem; 2 – minor problem; 1 –Not a problem)

Appendix 9. List of abbreviations:

CESSDA - Council of European Social Science Data Archives
CET - Cost Estimation Toolkit
DARM - Data activities reference model
DSA - Data seal of approval
ERI - European Research Infrastructure
DAF - The Data Audit Framework
DRAMBORA - Digital Repository Audit Method Based on Risk Assessment
LIFE - Life Cycle Information for E-Literature
LS - Levels of Service
PLEDGE - PoLicy Enforcement in Data Grid Environments
PREMIS - PREservation Metadata: Implementation Strategies
OAIS - Reference Model for an Open Archival Information System
SSDA - Social Science Data Archives
TRAC - Trustworthy Repositories Audit & Certification: Criteria and Checklist

AIP: Archival Information Package – a basic OAIS concept
CESSDA PPP: CESSDA Preparatory Phase Project, for which this report has been prepared.
CESSDA: Council for European Social Science Data Archives – http://www.cessda.org
CMS: Content Management System. An application for software editing, publishing, and the searching of various types of digital documentation. Web CMSs are those implemented as web applications.
DDI: "The Data Documentation Initiative is an international effort to establish a standard for technical documentation describing social science data. A membership-based Alliance is developing the DDI specification, which is written in XML." (http://www.ddialliance.org/)
DIP: Dissemination Information Package – a basic OAIS concept
DR: Digital repository
DSA: Data Seal of Approval - A minimum set of criteria for digital research data prepared by DANS (Data Archiving and Networked Services). DANS is a social science archive in the Netherlands and member of the current CESSDA network. (http://www.datasealofapproval.org/)
ELSST: European Language Social Science Thesaurus
ERIC: A "European Research Infrastructure Consortium" is the new legal framework for a legal entity, recognised in all Member States
METS: Metadata Encoding and Transmission Standard "The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation." (http://www.loc.gov/standards/mets/)
NESSTAR: A software tool for data publishing and online data analysis: http://www.nesstar.com/
NESTOR: The German Network of expertise in Digital long-term preservation. NESTOR "is a cooperative project of libraries, archives and museums as well as of leading experts forming a network of expertise in long-term preservation and long-term availability of digital resources. A goal of this project (funded by the BMBF, German Ministry of Education and Research) is the constitution of a permanent form of

organization for all issues of long-term preservation as well as the development of national and international agreements and the assignment of tasks." http://www.langzeitarchivierung.de/index.php?newlang=eng

OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting. This is used to harvest (or collect) record metadata descriptions in an online internet archive. By using OAI protocol distributed services can be built on metadata sourced from many archives.

OAIS: OAIS is the ISO reference model for Open Archival Information Systems. The OAIS reference model is defined by a recommendation from the Consultative Committee for Space Data Systems (CCSDS). (http://public.ccsds.org/) The recommendation was later approved as an ISO standard. The full ISO standards are available at: (http://public.ccsds.org/publications/archive/650x0b1.pdf).

SIP: Submission Information Package – a basic OAIS concept

8.   References

Alma Swan & Sheridan Brown Skills, Role & Career Structure of Data Scientists & Curators: Assessment of Current Practice & Future Needs. JISC Report.
http://www.jisc.ac.uk/publications/publications/dataskillscareersfinalreport.aspx

APA (Alliance for Permanent Access) 2008 Keeping the records of science accessible: can we afford it? Report on the 2008 Annual Conference of the Alliance f
Permanent Access, Budapest, 4 November 2008 http://www.alliancepermanentaccess.eu/documenten/Alliance2008conference_report.pdf

Beagrie, Neil; Julia Chruszcz; and Brian Lavoie (2008) Keeping research data safe. JISC Report.
http://www.jisc.ac.uk/publications/publications/keepingresearchdatasafe.aspx

Beedham, Hilary, Julie Missen, Matt Palmer, Raivo Ruusalepp. 2005. Assessment of UKDA and TNA compliance with OAIS and METS Standards. UKDA, TN
(http://www.data-archive.ac.uk/news/publications/oaismets.pdf)

CCSDS (Consultative Committee for Space Data Systems). 2002. (January). "Reference Model for an Open Archival Information System (OAIS)" CCSDS 650.0
1 BLUE BOOK. Washington DC. CCSDS Secretariat. (http://public.ccsds.org/publications/archive/650x0b1.pdf)

CCSDS 2004 (May). "Producer-Archive Interface Methodology Abstract Standard"

CCSDS 651.0-B-1 BLUE BOOK. Washington DC. CCSDS Secretariat. (http://public.ccsds.org/publications/archive/651x0b1.pdf)

CESSDA PPP. 2008. Background document on Legal, Governance and Funding issues relating to the CESSDA-PPP for Funders Meeting, London . 27 October
2008. Draft document with restricted access. CESSDA-PPP Web - Management Board section.
http://www.cessda.org/ppp/board/doc/27.10.08.Background.pdf

CESSDA-PPP-MB (2009) The Statutes of the Council of European Social Science Data Archives – European Research Infrastructure. Draft_Statues_for_CESSD
ERI_v0. CESSDA Management Board Meeting, Athens, Greece: 15 January 2009. http://www.cessda.org/ppp/board/doc/Draft_Statues_for_CESSDA-
ERI_v0.pdf

CESSDA-PPP. 2008. Annex I - "Description of Work". Annex to the Grant agreement for CESSDA-PPP, project No. FP7-212214.

CESSDA-PPP. 2008. CESSDA PPP Survey Questionnaire. CESSDA-PPP Web. http://www.nsd.uib.no/cessda/ppp/surveys/

CESSDA-PPP. 2009a. Service Level Agreement. For the provision of data access, archiving, dissemination, preservation and user support services. Preliminary d
distributed before the CESSDA General Assembly and CESSDA-PPP Management Board meeting in April 2009.

CESSDA-PPP. 2009a. The Statutes of the CESSDA European Research Infrastructure Consortium (CESSDA-ERIC). Proposed Statuses for the new organisation
CESSDA-PPP. Available on-line from Management Board section of the CESSDA-PPP Web: http://www.cessda.org/ppp/board/doc/legal-framework-

statues.pdf

CESSDA-PPP. 2009b. Reforming the Council of European Social Science Data Archives. Strategic document prepared by CESSDA-PPP project for information potential funders. Restricted access. Available on-line from Management Board section of the CESSDA-PPP Web: http://www.cessda.org/ppp/board/doc/legal-framework-docs-notes.pdf

CESSDA-PPP. 2009b. The Statutes of the CESSDA European Research Infrastructure (cessda-ERI). Version - 2. Preliminary proposed Statuses for the new organisation. CESSDA-PPP. Available on-line from the Management Board section of CESSDA-PPP Web: http://www.cessda.org/ppp/board/doc/cessdaERI_legalframework_statues_v2.pdf

Chen, Robert S., W. Christopher Lenhardt (2002) SEDAC Inputs to SEEDS Levels of Service Workshop

Corney, D.R., De Vere, M., Folkes, T., Giaretta, D., Kleese van Dam, K., Lawrence, B. N., Pepler, S.J., Strong, B. (2004) Applying the OAIS standard to CCLRC British Atmospheric Data Centre and the Atlas Petabyte Storage Service. Central Laboratory for the Research Councils http://www.allhands.org.uk/2004/proceedings/papers/156.pdf

CRL (CRL, The Center for Research Libraries), OCLC (Online Computer Library Center, Inc.). 2007. "Trustworthy Repositories Audit & Certification: Criteria Checklist.": http://www.crl.edu/PDF/trac.pdf

Dale, Robin, with contributions from G. Sayeed Choudhury, Tim DiLauro and Marie Waltz. (2006) ICPSR Audit Report. For the period ending 24 October 2006 Center for Research Libraries. Auditing and Certification of Digital Archives Project, retrieved 13. 2. 2009, from http://www.crl.edu/PDF/ICPSR_final

Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE). 2007 (February), "DCC and DPE Digital Repository Audit Method Based on Risk Assessment, v1.0., retrieved 26.10.2008, from http://www.repositoryaudit.eu/download

DSA. 2009. Data Seal of Approval: Quality guidelines for digital research data. Version 1.4. DANS. Available on-line: http://www.datasealofapproval.org/sites/default/files/Data_Seal_of_Approval_1-4.pdf [accessed 23/06/2009].

Dusa, Adrian, Brigitte Hausstein, Janez Stebe, Jindrich Krejci, Peter Hegedus, Zoltan Fabian. 2008. Report on the WP6 Workshop on Strengthening the CESSDA RI. Report from the Workshop held in Bucharest, Romania, on 16 July 2008. CESSDA-PPP, Workpackage 6.

Dusa, Adrian, Jindrich Krejci. 2008. Summary of the CESSDA PPP WP6 Workshop on "Strengthening the CESSDA RI". CESSDA-PPP Project.

ERPANet (ELECTRONIC RESOURCE PRESERVATION AND ACCESS NETWORK) (2003): Digital Preservation Policy Tool http://www.erpanet.org/guidance/docs/ERPANETPolicyTool.pdf

Eurostat Quality 2009. http://epp.eurostat.ec.europa.eu/portal/page?_pageid=2273,1,2273_47140765&_dad=portal&_schema=PORTAL

Fábián, Zoltán and Peter Hegedűs. 2008. Audit of expertise within the CESSDA network of excellence. Interim report from Task 6.3. CESSDA-PPP, Workpacka

6.

Fábián, Zoltán(2009) Self-assessment procedures for the CESSDA RI infrastructures. Interim report from Task 6.2. Version 0.9 – draft. CESSDA PPP, WP6, 200 01.19.

Faundeen, John (2005) Selection, Appraisal and Retention of Digital Scientific Data. Workshop on Strategies for Permanent Access to Scientific Information in Southern Africa: Focus on Health and Environmental Information for Sustainable Development.

Fontaine, Kathy (2006) Earth Science Data System Working Groups, NASA, WGISS-22 Annapolis, MD, September 13, 2006

Fontaine, Kathy, Greg Hunolt, Arthur Booth, Mel Banks (2008) Observations on Cost Modelling and Performance Measurement of Long-Term Archives

Green, Ann (2008) Guidelines and Tools for repository planning and assessment presented at the DataShareproject meeting University of Edinburgh, Feb 5-6, 20

Harmsen, Henk (2009) Data Seal of Approval. Dissemination, assessment & procedures in 30 slides. Digital Preservation Workshop - January 30th 2009 - The Hague. Retrieved 13. 2. 2009, from http://www.datasealofapproval.org/files/20090130_Harmsen.ppt

Hausstein, Brigitte, and Paul de Guchteneire. 2002. Social Science Data Archives in Eastern Europe. Results, Potentials and Prospects of the Archival Developme Berlin, Cologne, Paris: E. Ferger Verlag. Available also on-line: http://www.gesis.org/en/research/networks/data-infrastructure/edan/#Publication

Higgins, Sarah and Ross Harvey (2008) The DCC Curation Lifecycle Model. 4th International Digital Curation Conference "Radical Sharing: Transforming Science?"

Humphrey, Charles. (2006) "e-Science and the Life Cycle of Research" retrieved 13. 2. 2009, from http://datalib.library.ualberta.ca/~humphrey/lifecycle-science060308.doc

Humphrey, Charles. (2008) Radical Data Services. In Celebration of the first 25 years of the University of Edinburgh Data Library. http://datalib.ed.ac.uk/25anniversary/Presentations/Humphrey.ppt#273,16,Slide 16

Hunolt, Greg (2003) Data Service Provider Reference Model - Requirements / Levels of Service. SEEDS Levels of Service / Cost Estimation Study Working Pap Five, NASA. http://esdswg.gsfc.nasa.gov/st_cost.html#Documents

Hunolt, Greg (2006) Users' Guide. Cost Estimation Toolkit (CET). Version 2.1 Funded by NASA Contract NNG05CA99C

ICPSR (2008): Strategic Plan. Retrieved 13. 2. 2009, from http://www.icpsr.umich.edu/ICPSR/strategic/strategic-plan.pdf

ICPSR (Inter-university Consortium for Political and Social Research) (2005) Guide to Social Science Data Preparation and Archiving. Best Practice Throughout Data Life Cycle, 3rd Edition. http://www.icpsr.umich.edu/ICPSR/access/dataprep.pdf

IHSN (International Household Survey Network) (2006) Tools and guidelines. http://www.surveynetwork.org/home/?lvl1=tools&lvl2=overview

Jacobs, Jim. (1991) "Providing Data Services for Machine-Readable Information in an Academic Library: Some Levels of Service." The Public-Access Compute
Systems Review 2, no. 1: 144-160.

JISC Monitoring Unit - http://www.mu.jisc.ac.uk

Juran, J. M. and Frank M. Gryna. 1993. Quality Planning and Analysis: From Product Development Through Use. Third edition. New York: McGraw-Hill.

Kevin Bradley, Junran Lei, Chris Blackall (2007): Towards an Open Source Repository and Preservation System. Recommendations on the Implementation of an
Open Source Digital Archival and Preservation System and on Related Software Development. National Library of Australia and UNESCO Memory of
World Sub-Committee on Technology. Paris, June 2007
http://portal.unesco.org/ci/en/files/24700/11824297751towards_open_source_repository.doc/towards_open_source_repository.doc

Krejci, Jindrich and Tomas Cizek. 2008. Current training practices and training opportunities for the CESSDA. Interim report from Task 6.1. CESSDA-PPP,
Workpackage 6.

Krejci, Jindrich, Brigitte Hausstein and Adrian Dusa. 2009. Establishing CESSDA Training Programme: Challenges and possible strategies. CESSDA-PPP Interi
Report from Task 6.5 and Task 6.7. CESSDA-PPP, Workpackage 6.

LeFurgy, William G. (2002) Levels of Service for Digital Repositories. D-Lib Magazine. May Volume 8 Number 5. retrieved 13. 2. 2009
http://www.dlib.org/dlib/may02/lefurgy/05lefurgy.html

LIFE (2008) LIFE (Life Cycle Information for E-Literature). University College London (UCL) and the British Library. LIFE Homepage. retrieved 13. 2. 2009,
http://www.life.ac.uk/

Moore, Reagan (2008): Towards a Theory of Digital Preservation, Vol 3, No 1. http://www.ijdc.net/index.php/ijdc/article/view/63/42

Myron Gutmann, Ben Evans, Deborah Mitchell, Kevin Schürer (2009): The Data Archive Technologies Alliance: Looking towards a Common Future. Presentati
at the IASSIST/IFDO 2009 Conference in Tampere, Finland http://www.fsd.uta.fi/iassist2009/presentations/F1_Gutmann.pps

NESTOR (Working Group on Trusted Repositories Certification). 2006. Catalogue of Criteria for Trusted Digital Repositories, Version 1 (draft for public comm
Frankfurt am Main: nestor c/o Deutsche Nationalbibliothek, urn:nbn:de:0008-2006060710

Ramapriyan, H. K. (Rama) (2008) Metrics Life Cycle & Archive Policy, NASA/GSFC, Metrics Planning and Reporting (MPAR) WG, 7th Earth Science Data
Systems Working Group, Philadelphia, PA, October 21-23, 2008

RELU (Rural and Land Use) (2006) Guidance on Data Management. http://www.data-archive.ac.uk/relu/RELUaug2006.pdf

RIN (Research Informaiton Network) (2008) Stewardship of digital research data: a framework of principles and guidelines Responsibilities of research institutio
and funders, data managers, learned societies and publishers. http://www.rin.ac.uk/data-principles

Ruusalepp, Raivo (2009): Building Blocks of a Certification Process: DRAMBORA and TRAC. Digital Preservation Workshop - January 30th 2009 - The Hague. Retrieved 13. 2. 2009, from http://www.datasealofapproval.org/files/20090130_Ruusalepp.ppt

Schürer, Kevin (2006) Project on Enhancing Poverty Measurement, Monitoring and Analyses in Russia. Consultant's Report on providing assistance to Rosstat in Open Access to the Household Budget Survey. UK Data Archive

Sesink, Laurents, René van Horik en Henk Harmsen (Eds.) ([2008]): Data Seal of Approval (Datakeurmerk): Quality guidelines for digital research data in the Netherlands. DANS: http://www.datasealofapproval.org/files/datasealofapproval1-2_colour.pdf visited on 2.2.2009.

Smith, MacKenzie, Reagan W Moore (2007): Digital Archive Policies and Trusted Digital Repositories. International Journal of Digital Curation, Vol 2, No 1, http://www.ijdc.net/index.php/ijdc/article/view/27/30

Štebe, Janez and Adrian Duşa. 2009. Task 6.4 Report: Recommendations concerning best practices. CESSDA-PPP, Workpackage 6. Available on-line: http://www.cessda.org/ppp/wp06/WP6_T6.4.pdf

Swan, Alma and Sheridan Brown. 2008. The Skills, Role and career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs. Report to the Joint Information Systems Committee (JISC). Key Perspectives Ltd. http://www.jisc.ac.uk/publications/publications/dataskillscareersfinalreport.aspx

UCL (University College London) and The British Library (BL) (2008) The LIFE2 Final Project Report. A JISC-funded joint venture project under 03/06, Repositories and Preservation Capital Programme, and supported by the LIBER Access and Preservation Divisions

UKDA (UK Data Archive) (2008a) Managing and Sharing Data: a best practice guide for researchers. http://www.data-archive.ac.uk/news/publications/managingsharing.pdf

UKDA (UK Data Archive) (2008b) Processing and preservation. Some things never change... 40 years of the UKDA. http://www.data-archive.ac.uk/ukda40/about/processing.asp

UKDA. 2008. UK Data Archive Preservation Policy. Created by Matthew Woollard. http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0308.pdf

Vardigan, Mary and Cole Whiteman. 2007. "ICPSR meets OAIS: applying the OAIS reference model to the social science archive context." Archival Science (Volume 7, Number 1 / March): pp. 73-87.

Woollard, Matthew (2008) UK Data Archive Preservation Policy. UK Data Archive, University of Essex, Ver. 3.00 http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0308.pdf