WP4 - Strengthening and widening through knowledge exchange
Development of the necessary administrative, technical, and methodological support needed to establish and develop data archives
Start month: 1
End month: 24

Objectives

The key objective of Work Package 4 is to enlarge and strengthen CESSDA through knowledge exchange between current members and potential or aspiring members. A bottom-up approach is central to this WP. By supplying the necessary administrative, technical, and methodological support in countries with a more developed data archiving infrastructure, the WP will promote the establishment of new data archives and strengthen existing ones. The key objective of WP4 consists of five development support tasks, each focusing on an aspect of the required knowledge exchange. The sixth task of WP4 focuses on understanding how data archives contribute to valorisation of knowledge about society.

Description of work and role of partners

Partners: CESSDA (ADP, NSD, DANS, GESIS, FSD, FORS, UKDA, CNRS, CSDA, SND, LiDA), Charles Beagrie Ltd, EKKE, TARKI Foundation, UTARTU, NUID UCD, ICS-ULisboa

Task 4.1 - Development Support: a rolling programme of hands-on training for data archive practitioners on ingest and dissemination practices, trust (Data Seal of Approval), and standards integration.
This Task will be led by GESIS with 5 PM, with support from ADP (2PM), NSD (2PM), and DANS (3PM). Not all data archives in Europe are on the same level. In the two primary points of contact between an archive and its user community: ingest and dissemination, we find various levels of quality, and it is this quality that has significant impact on the overall quality of data curation and preservation, as well as on the acceptance of data sharing and reuse in the user community. Offering knowledge sharing through hands-on practitioner and community training in these two areas is an efficient means to promote knowledge transfer – both among current CESSDA members and between members and non-member ERA archives working towards building services in line with CESSDA AS quality standards. To achieve that this task will conceptualise and produce online tutorials and webinars that will function as "common ground". The deeper approach will be delivered in the two face-to-face workshops. The training offers are targeted at practitioners from CESSDA AS archives and non-member archives. This includes records managers, information technology staff, and - especially where policies are
concerned - managers and executive staff. Where applicable, the training will make use of an experience-based reference model for data professionals. This reference model contains an overview of skills and competencies and the lifelong learning requirements for this target group. All webinars will be archived and available for future use in the knowledge-sharing mechanism (see WP2).

The task will deliver the following:

- **MS4.1.** Webinar 1: Access policies and usage regulations: licenses
- **MS4.2.** Webinar 2: RDM community training
- **MS4.3.** Online tutorial 1: Elements of the Ingest workflow (incl. pre-Ingest): overview of the entire workflow, highlighting the tasks that archives have to fulfil here; relevance for the overall preservation/curation workflow
- **MS4.4.** Online tutorial 2: Elements of the dissemination workflow: overview of the entire workflow, highlighting the tasks that archives have to fulfil as part of this function; relevance for the overall preservation/curation workflow
- **D4.1.** Workshop 1 report: Trust workshop to support those CESSDA members and aspiring members in need of this training in the process of obtaining the Data Seal of Approval (DSA)
- **D4.2.** Workshop 2 report: Integration workshop on one of the following topics (depending on the results of WP3):  ○ Metadata standards  ○ Strategies for persistent identification (PID)  ○ Strategies for Authentication and Authorization Infrastructures (AAI)  ○ Thesaurus solution  ○ Disseminating data in a cooperative setting: portal solutions, standard protocols, etc.

**Task 4.2 - Development Support: Development of the necessary administrative, technical, and methodological support needed to establish and develop trustworthy data archives.**

This Task will be **led by NSD with 5 person months, with support from ADP (3PM), FSD (3PM) and FORS (3PM).**

Task 4.2 has two main goals:

- to provide resources and guidelines for the development of internal service provider policies;
- to facilitate understanding and cooperation among stakeholders.

Building on results from the FP7 project DASISH, the Reference Model for an Open Archival Information System (OAIS), and the Audit and Certification of Trustworthy Digital Repositories, the aim of T4.2 is to provide high-level guidance to the management of research data services, covering all aspects of the data lifecycle. The outcomes of the task will assist new and developing organisations to prepare and upgrade their own policies and to integrate them into their internal workflows and processes. T4.2 will moreover provide a common CESSDA policy framework for setting priorities in technical harmonisation and recommendations for monitoring research data usage.
Task 4.2 will be divided into two sub-tasks:

**T4.2.1 Context and Purpose**
In this sub-task funder policies (national and H2020), data management planning, and mission statement (objectives, purpose, mandate, etc.) will be mapped and analysed. It will also analyse and describe common supporting services, in particular concerning sensitive data management, common metadata structures, versioning and use of persistent identifiers, and replication among Service Providers.

**T4.2.2 Functional Model / Digital Object Model**
The phases of the data lifecycle relevant for data archives will be described and analysed, including:

- **Pre-Ingest, Acquisition, and Ingest:** It will cover the ‘Producer-Archive interaction’ by focusing on tools and standards like the Producer-Archive Interface Methodology Abstract Standard (PAIMAS) topics, such as selection criteria, options for self-archiving, submission agreements, data cleaning, data processing and preferred formats and data types accepted for archiving.
- **Archival Storage:** This will cover preservation strategies, maintenance and retrieval of Archival Information Packages (AIP) and risk management. It will also focus on the organisations’ commitment and approach to ensuring the accuracy, completeness, authenticity, integrity, and long-term protection of data.
- **Data Management:** This will cover issues regarding administrative data used to manage the service. It will also focus on metadata schema, metadata structure and on version control, quality control and change procedures.
- **Access:** This will identify how end users interact with the archive to find, request and receive data and metadata, and covers issues like access and authentication infrastructure, usage promotion and monitoring, citation requirements and references to research project information systems.

**Task 4.3 - Development Support: achieving the Data Seal of Approval**
This Task will be led by DANS with 8PM, with support from GESIS (3 PM), UKDA (3 PM), FSD (3 PM), and ADP (3 PM).

It has become mandatory for all CESSDA Service Providers to achieve the Data Seal of Approval (DSA) in order to get Trusted Digital Repository status. Driving factor here for CESSDA and its related Service Providers will be the adherence to professional excellence. CESSDA needs to develop and promote a set of professional standards. Also clarification of the relation between the DSA Guidelines and the CESSDA Annex 2 Obligations is needed. Central in this task is the formation of a certification work group (“Trust Group”), with the following tasks: monitoring, reviewing, advising, assisting on certification as well as clarifying DSA Guidelines, in particular in relation to the CESSDA Obligations for Service Providers (Annex 2 to the CESSDA Statutes). The main deliverable will be a plan in which the different tasks for the certification (or trust) work group, briefly described in the task objective, will be elaborated.

---

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement number 674939.
This will include:

- Making sure that all the DSA obligations are met by all SPs, in the prescribed timeframe by monitoring the progress with respect to certification;
- Translation of the CESSDA Annex 2 Obligations into the DSA guidelines in the first year (establishing the relationship between these two entities);
- A review of DSA self-assessments;
- Providing assistance on DSA certification and possibly Annex 2 Obligations Compliance (in coordination with other training initiatives within CESSDA), linked with Task 4.1 on Training;
- Giving advice on DSA certification and its relation with the Annex 2 Obligations. As steps towards the production of a plan, initially in the form of guidelines, the already existing ideas and plans have to be evaluated (MS4.10).

This should lead to a proposal on how to set up the Trust Group (MS4.11) and, taking into consideration the tasks to be fulfilled by this group, guidelines for the CESSDA SPs (D4.4). In and before M22, liaison will be made with task 4.1, in particular Workshop 1 (Trust workshop to support aspiring CESSDA members in the process of obtaining the Data Seal of Approval). The experiences and results of that workshop should be included in the final report of this task 4.3. A final report will be delivered in which the different tasks as proposed in the guidelines based on the (ongoing) certification activities of the CESSDA SPs (including aspiring SPs) will be evaluated, if necessary modified and elaborated.

**Task 4.4 Development Support: Establishing the necessary conditions for creating new or reinforcing existing social science data services.**

This Task will be led by DANS with 12PM, with support from TARKI (2PM), CSDA (2PM), FORS (2PM), NSD (2PM), FSD (2PM), and CNRS (2PM).

The objective of this task is to establish the conditions for creating new or reinforcing existing social science data services. In the start-up and consolidation phases of new archives, selected well established CESSDA archives will perform supportive tasks for starting ones, especially offering technical facilities and expertise on software tools, backup services and archival policies, for as long as trustworthy and sustainable services in the countries aspiring CESSDA membership are lacking (note that training is a separate task dealt with in Task 4.1). On the basis of the maturity model developed in Task 3.1, data archiving services throughout Europe will be characterised according to their level of development (ranging from “non-existent” to “well-established CESSDA service provider”) in Task 3.2. This will also offer insight in the gaps in social science data services. A stepwise approach to get data archiving services off the ground and consolidated is useful, starting with a description of the requirements, followed by a feasibility studies, pilot projects, etc. This task prepares the ground for possible later CESSDA membership, for which a data service needs to have a certain maturity and financial sustainability.
The task will perform the following activities, which will each result either in a milestone or deliverable:

a. Establish the demand for development support services (also on the basis of 3.2)
b. Establish the supply of development support services
c. Pilot of delivery of development support services on the basis of a. and b.
d. Develop a sustainability model for development support services
e. Funding models and identifying funding opportunities / possible sources of finance

Ad a. The task will begin by identifying the demands for development support services in those countries of the ERA where, according to the outcomes of Task 3.2, data services for the social sciences are either lacking or where the maturity is low. A workshop will be organised to bring the demand information together (MS 4.12).

Ad b. Parallel to (a), the potential support that is on offer by CESSDA member countries with a more mature service infrastructure will be described. Such services will be in the following categories: technical facilities; expertise on software tools; backup services; advice on archival policies (MS 4.13).

Ad c. On the basis of (a) and (b) supply and demand will be matched and a selection of support services will be implemented. For instance, while a social science data archive is under development in country X aspiring later CESSDA membership, a backup service can be provided by a CESSDA member in country Y with a more stable infrastructure, so that the aspiring country can offer guarantees conforming to the Data Seal of Approval (Result: D4.5).

Ad d. Activity (c) is guaranteed during the CESSDA-SaW project, but it needs to be clear what will happen afterwards. For this, a sustainability model needs to be in place. This activity will make sure that the provided service will persist if necessary; otherwise, if the cross-national service is no longer required as the maturity of the service in country X has reached a certain maturity, it must be clear how the cross-national service is ended and data is transferred (Result: D4.6).

Ad e. This activity develops experience and provides advice regarding funding models and sources of finance. It will explore compliance with European and other funding opportunities. Lobbying work, capacity building, monitoring funding opportunities, support for proposal preparation based on successful examples as showcases, and information exchange on applicable rules and principles for proposals are part of this.

Task 4.5 - Strengthening and widening through establishing a CESSDA partner network of established non-member data services to support processes of integration into CESSDA membership and establish a wide collaborative environment among European data service providers.
This task will be led by TARKI with 15PM, supported by CSDA (5PM), ICS-ULISBOA (4 PM), EKKE (4 PM), and ISSDA (4 PM), and SND (4 PM). The goals of the task are twofold: (1) to
This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement number 674939.

discover and monitor the capacities of European data services outside the current CESSDA membership, and (2) to establish a collaborative network between CESSDA and non-member data services. CESSDA is looking for possible ways to organise collaboration with non-CESSDA data services.

The first activity within this task is to delve into the technical, organisational and administrative development level of the data services, as well as into the financial sources on which they operate (meaning state/European funds, private investments for services provided, etc). The second activity in this task is to create a collaborative structure between the partners in a widened CESSDA network.

The activities will result in both: (1) establishing the mechanism for dissemination of CESSDA know-how and assistance support processes for integration of potential new CESSDA members; and (2) widening collaboration in promoting and implementing CESSDA objectives at the European level. The work will be performed in close connection with other support tasks of Work package 4 and with activities within Work package 3 devoted to the development of data infrastructures and the audit of state-of-the-art data archives in ERA countries.

To reach the objectives of the task, the first milestone is a collaborative workshop on establishing partnerships between CESSDA and service providers convergence point with Work Package 3, to which service providers, ministries and official representatives of the CESSDA member states and potential partner countries will be invited. The workshop aims to establish an informal collaborative network of current and potential CESSDA partners and will collect information regarding the needs of future cooperation. Using the information from the non-CESSDA member data services, the first deliverable of the task will summarise the current state and needs of these data services. Based on this, in the second stage of task 4.5, models of becoming full member of CESSDA AS/ERIC will be formulated. At the end of this task, there will be a second workshop on the processes of integration of data services into CESSDA (together with WP 3) and a final report on formal models of collaboration and development towards structuring a widened CESSDA network. This task builds on insights and results of other collaborative projects aimed at integration, structuring and development of European data infrastructures, especially the projects CESSDAPP, SERSCIDA and DASISH, as well as on the former networking activities of EDAN (East European Data Archive Network) and REGIO (Regional Data Archive Network in the Central and East European region).

**Task 4.6 – Understanding the economic impact of social science data archives.**

This task will be led by Charles Beagrie Ltd with 10PM, with support from ADP (2 PM), FSD (2 PM), LiDA (1.5PM), UTARTU (1.5PM) and UKDA (1.5PM).

In Task 4.6 we will develop a benefit/cost advocacy programme and supporting tools such as workshops, factsheets, and (four) case studies for data services. Once developed, individual organisations will be able to apply the programme to provide nationally relevant economic and financial information, taking account of the ‘hidden’ benefits and impact of data-sharing, to support requests for sustainable funding for new national data services (a requirement for
This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement number 674939.

CESSDA membership). We expect the cost/benefit advocacy programme to have a significant impact by assembling an evidence base to support the negotiation with ministries and funding organisations. The evidence base and toolset should also support advocacy with other core stakeholders such as data creators and data users. The support and engagement of these different constituencies will be critical to starting, growing, and sustaining European social science archives. The development of the cost/benefit advocacy programme will require the gathering of information from members and potential new members.

The methods for this activity will be similar to those applied during the CESSDA PPP and which were further developed and enhanced for the SERSCIDA project. They include surveys, workshops, round table meetings and focus groups as appropriate to the information need. Charles Beagrie Ltd will lead on the development of core documents and materials for the task with support from CESSDA AS for the gathering of information from members and potential new members and user testing with them of the advocacy programme and draft supporting materials. The four case study partners will assist with detailed user testing and case studies for their countries. The cost/benefit advocacy programme will be able to draw on a range of pre-existing work by partners in the consortium. However it will need to develop the methodology and a toolset of documents in order for this to be applicable to a range of European countries and in new and emerging as well as established social science data archives.

The generic work on costs and benefits undertaken for the Keeping Research Data Safe (KRDS) project (Beagrie et al 2008, 2010, 2011) and the 4C European project on digital preservation cost models (Grindley et al 2014) provide a sound approach and basis for considering the costs and qualitative benefits of social science archives. The Netherlands Data Archiving and Networked Services (DANS) has also developed a digital preservation cost model (specific to DANS but potentially more widely applicable if generalised) based on the Activity-Based-Costing (ABC) method and combined it with the Balanced Scorecard Method (Palaiologk et al 2012). This may provide complementary data and approaches. The study of the value and economic impact of the Economic and Social Science Data Service (ESDS) in the UK (Beagrie and Houghton 2012) adds a widely based quantitative evaluation of the economic impacts of a social data archive (as well as the application of the KRDS Benefits Framework to the qualitative benefits).

Extending this existing work to a wide range of European social science archives in different countries will require adaptation of existing tools and findings to fit a range of circumstances. Existing Activity Based Costing models and economic impact evaluations are time and resource heavy tools to implement but have been applied in the largest and most well resourced countries. We can describe and make available these available cost and economic impact evaluation models and support materials for other countries who can implement them. What is needed in addition, is a set of lighter, easier to adopt projections and approximations based on existing work that may be more suitable for differences in scale, resource, and available time in other European countries and can support adoption of the CESSDA developmental model. A more summary approach to costs distilling findings from across 13 organisations has been
pioneered by the KRDS project looking at long-term archiving costs and proportional costs for ingest, archiving and access costs. With updating and further additions, we believe this could provide an excellent basis for a set of lighter, easier to adopt projections and approximations for costs for data archiving. A set of lighter, easier to adopt projections and approximations for value and economic impact will require more work to produce as there has been less processing of existing studies for this purpose. Tools such as the Balanced Scorecard Method adopted by DANS and the KRDS Benefits Framework adopted by the UK Data Archive could provide a good basis for this on the qualitative understanding of value. However currently the study of the value and economic impact of the Economic and Social Science Data Service (ESDS) in the UK is the only source that has quantified economic impact for social science data archives. Its findings on Returns on Investment and cost/benefit ratios are a starting point but for transferability the projections will need to consider and adjust for the impact of service maturity and critical mass of data collection, between different European countries and their data archives. Although based on data archives and services for other disciplines, subsequent economic impact studies for the Archaeology Data Service, the British Atmospheric Data Centre, and EMBL – European Bioinformatics Institute build on, and can contribute to, the generic lessons from the ESDS study for projections of economic impact from data sharing and the services that data archives provide.

We will follow an iterative approach to developing the advocacy programme, commencing with a short survey and workshop to gather and validate requirements with potential users, followed by development of draft components of the toolset, user testing in a set of round table meetings and focus groups, finalisation of the toolset, and promotion and dissemination by CESSDA. The toolset of documents needed to accompany the generic cost/benefit advocacy programme will include a short “how to” guide, factsheets, and a set of four case studies of the value and impact of mature, developing/emerging, or new/planned social science archives that could be used by peer European countries.