cessda saw

Research and Innovation Action

CESSDA Strengthening and Widening

Project Number: 674939

Start Date of Project: 01/08/2015

Duration: 24 months

Deliverable 2.2

Knowledge-Sharing Platform Content and

Management Policy

| Dissemination Level | PU |
|-------------------------|-------------------------------------|
| Due Date of Deliverable | 31/10/16 |
| Actual Submission Date | 31/10/16 |
| Work Package | WP2 Dissemination and Communication |
| Task | 2.1 |
| Туре | Report |
| Approval Status by EC | Not approved yet |
| Version | V1.1 |
| Number of Pages | p. 1 – p. 28 |

Abstract: Building on the results of the Virtual Stakeholder's Forum (D2.1), this report details how the content and functionality of the CESSDA Knowledge-Sharing Platform will be developed, and describes how it will be run, managed and maintained post-project.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability.



History

| Version | Date | Reason | Revised by |
|---------|------------------|--|--|
| 0.0 | 08/08/2016 | Document created, table of contents, first text input | Jonas Recker |
| 0.1 | Until 11/10/2016 | Further text input, comments on and discussion of content | Jonas Recker Birger Jerlehag Iris Alfredsson Irena Vipavc Brvar Sonja Bezjak |
| 1.0 | 20/10/2016 | Finalize text | Jonas Recker |
| 1.1 | 27/10/2016 | Incorporate feedback from Project Manager and CESSDA Technical Officer | Jonas Recker |
| 1.1 | 28/10/2016 | Deliverable revised by Delivery Committee Chair and Project Coordinator | Jindrich Krejci, Ivana Ilijasic Versic |

Author List

| Organisation | Name | Contact Information |
|--------------|--------------------|----------------------------|
| ADP | Sonja Bezjak | sonja.bezjak@fdv.uni-lj.si |
| ADP | Irena Vipavc Brvar | irena.vipavc@fdv.uni-lj.si |
| CESSDA MO | JBM | jbm |
| GESIS | Jonas Recker | jonas.recker@gesis.org |
| SND | Iris Alfredsson | iris.alfredsson@snd.gu.se |
| SND | Birger Jerlehag | birger.jerlehag@snd.gu.se |

Time Schedule before Delivery

| Next Action | Deadline | Care of |
|--|------------|-----------------|
| Determine elements and scope of platform content and management policy | 31/06/2016 | ADP, GESIS, SND |
| Analyze and describe functional requirements, describe outcome of software selection process | 15/09/2016 | GESIS |
| Develop and describe licensing schema | 15/09/2016 | SND |
| Develop and describe collection and operation policy | 15/09/2016 | ADP |
| Develop and describe metadata schema | 15/09/2016 | ADP |
| Develop and describe future content development and management of platform | 15/09/2016 | GESIS |
| Finalize report | 15/10/2016 | ADP, GESIS, SND |

EXECUTIVE SUMMARY

Over the years, CESSDA partners have created a significant body of knowledge on a broad range of topics. Much of this knowledge is captured in digital resources such as papers, presentations, reports, guidelines, and training materials. However, these resources are scattered across the different service providers, sometimes available from the webpage, but in other cases merely stored on internal servers. What is more, no systematic and structured description of these resources exists. Task 2.1 of the CESSDA SaW project is dedicated to remedying this situation by creating a virtual knowledge-sharing platform (KSP) as a central point of access for the body of knowledge created by CESSDA partners.

D2.2 delivers the Platform Content and Management Policy, detailing how content and functionality will be developed during the project and specifying how the Knowledge-Sharing Platform (KSP) will be run, managed and maintained post-project. It builds on D2.1 "Knowledge-Sharing Platform Forum Report", which contains the results of a survey among prospective users of the platform and gives some insight into their expectations relating to the content and functionality of the platform.

Development of platform functionality

The functionality of the platform will be shaped by the software, as well as by the implemented policies and standards for the description and organization the content. To ensure the usability and user friendliness of the platform, the software on which the platform will be built (DSpace) was selected based on a functional requirements analysis taking into account the different stakeholder groups who will be working with the platform (users, depositors, administrators). Furthermore, the OSS Watch criteria¹ for the selection of open sources software were employed.

The discoverability of content and overall user-friendliness of the platform for those looking for resources are also determined by how content is organized and described with (structured and un- or semi-structured) metadata. To support discoverability, the following was developed: a metadata schema drawing on relevant standards for open access repositories (e.g. Dublin Core Library Application Profile, DataCite); a model for the organization of content in the platform making use of the "community" and "collection" structure of DSpace.

Development of platform content

To support both the depositors and the future editors of the platform, a Collection and Platform Operation Policy was developed. It describes the scope of the platform, i.e. which content will be collected, and defines selection criteria to support the review and publication process.

An important aspect to consider when filling the platform with content are Intellectual Property Rights. Thus, when accepting resources for publication, CESSDA has to assure that the depositors have the right to publish the resource and that no rights of third parties are infringed. For this purpose, a deposit contract was created that depositors have to accept when submitting resources. This at the same time allows depositors to determine under which license they want to share their resources.

Maintaining the platform post-project

One of the express objectives of establishing the Knowledge-Sharing Platform is the systematic archiving of CESSDA resources. Thus, the platform must become a long-term effort of CESSDA AS. Accordingly, its continued existence and development must be guaranteed, both on a technical and a content level. To achieve this, the report addresses the following aspects: possibilities for platform hosting, organization of platform administration, submission management, and collection development with the help of an editorial committee.

¹ See <u>http://oss-watch.ac.uk/apps/openness/</u>.

| ADP | Slovenian Social Science Data Archives | |
|-------------|---|--|
| BSD | Berkeley Software Distribution | |
| сс | Creative Commons | |
| CC BY | Creative Commons Attribution | |
| CC BY NC | Creative Commons Attribution - Noncommercial | |
| CC BY NC ND | Creative Commons Attribution - NoDerivatives | |
| CC BY NC SA | Creative Commons Attribution - Noncommercial - ShareAlike | |
| CESSDA AS | Consortium of European Social Science Data Archives | |
| CESSDA MO | CESSDA Main Office | |
| CESSDA SaW | CESSDA Strengthening and Widening | |
| da ra | Registration agency for social and economic data | |
| DOI | Digital Object Identifier | |
| GESIS | GESIS - Leibniz Institute for the Social Sciences | |
| KSP | Knowledge-Sharing Platform | |
| OAI-ORE | Open Archives Initiative Object Reuse and Exchange | |
| OAI-PMH | Open Archives Initiative Protocol for Metadata Harvesting | |
| OpenDOAR | Directory of Open Access Repositories | |
| OPUS | Online Publikationsverbund der Universität Stuttgart | |
| РМ | Person Month | |
| RSP | Registered DSpace Service Providers | |
| SND | Swedish National Data Service | |
| SP | Service Provider | |
| SSOAR | Social Science Open Access Repository | |

Abbreviations and Acronyms

Table of Contents

Executive Summary

Introduction

Development of Platform Functionality

Requirements

<u>Software</u>

DSpace Data Model and Content Organization

DSpace Publication Workflow and Roles

Metadata Schema

DOI Registration

Implementation process (Technical Tasks)

Development of Platform Content

Requirements

Collection Policy and selection criteria

Content Development

Licensing and Deposit Agreement

Maintaining the Platform Post-Project

Platform Hosting

Platform Administration

Submission Management and Collection Development

Appendix 1: User Stories and Functional Requirements

Appendix 2: Functional requirements and Corresponding DSpace 5.x functions

Appendix 3: Collection and Operation Policy (DRAFT)

<u>Appendix 4: Deposit and licensing agreement for resources uploaded to the CESSDA Knowledge-Sharing</u> <u>Platform (DRAFT)</u>

1. INTRODUCTION

D2.2 delivers the Platform Content and Management Policy, detailing how content and functionality will be developed during the project and specifying how the Knowledge-Sharing Platform (KSP) will be run, managed and maintained post-project. It builds on D2.1 "Knowledge-Sharing Platform Forum Report", which contains the results of a survey among prospective users of the platform and gives some insight into their expectations relating to the content and functionality of the platform.

2. DEVELOPMENT OF PLATFORM FUNCTIONALITY

The functionality - and usability - of the completed platform will mainly be dependent on two factors: firstly, the software and, secondly, the implemented policies and standards for the description and organization of the content. Thus the software defines the general framework for the functionality of the platform. Important aspects to consider are support of standards; out of the box functions for deposit, management, and discovery; openness; and customizability. However, the discoverability of content and overall user-friendliness of the platform for those looking for resources are impacted by the way in which content is organized and described with (structured and un- or semi-structured) metadata. Thus from a user perspective, whether the platform is perceived as well-functioning and easy-to-use does not merely depend on technological features of the software but very strongly on the intellectual organization of content.

2.1 REQUIREMENTS

To ensure that the Knowledge-Sharing Platform is designed to be as user-friendly as possible, it was important to learn more about the needs and demands of future platform users. For D2.1 we determined the stakeholder groups and roles listed in Table 1.

| Group | Stakeholders | Envisioned use |
|-------|--|---|
| I | CESSDA Main Office KSP Editorial Committee CESSDA Service Providers CESSDA Observers and aspiring Service Providers | Extensive/heavy use of the platform in different roles |
| 11 | • Other (social science) data archives | Lighter and more selective use of the platform than Group I |
| 111 | Social science educators Social science researchers Academic support staff (administration, library staff, etc.) CESSDA Members (i.e. ministries) Policy makers on national and European level | Light use of only selected areas/materials on specific topics |

Table 1: Stakeholder groups

These stakeholders will use the platform in different roles. We distinguish:

- Depositors: submit resources to the platform.
- Users: search for and download content from the platform.
- Editors: manage the content-side of the platform.

• Administrators: manage the technology-side of the platform.

The most common role across all stakeholders will be that of user. A considerably smaller group of stakeholders will also deposit to the platform, and the role of managing content and the technological side of the platform will be performed by a yet smaller number of individuals.

To determine the requirements of the "Users", the D2.1 online survey was carried out (see D2.1 Report). To make it easier to translate these results into actual functional requirements for D2.2, user stories were created. User stories are a tool from agile software development, consisting of "short, simple descriptions of a feature told from the perspective of the person who desires the new capability, usually a user or customer of the system" (https://www.mountaingoatsoftware.com/agile/user-stories). The stories were created to better capture and illustrate the desired functionality and to facilitate the communication with the programmers.

To identify needs and demands of the other stakeholder groups, further user stories were created and reviewed by the current Editorial Committee, whose members will use the KSP in the roles of depositor, editor, and administrator during the project. User stories were then matched with software features/functions required to meet the demands expressed in them (see Appendix 1).

2.2 Software

The decision was made to implement the KSP using DSpace, an open source repository software that is widely used for institutional repositories (<u>http://www.dspace.org</u>). Other open source software solutions on the market are, for example, EPrints (<u>http://www.eprints.org/uk/</u>), Fedora (<u>http://fedorarepository.org</u>), Hydra (<u>https://projecthydra.org</u>), or OPUS (<u>https://en.wikipedia.org/wiki/OPUS_(software</u>)). The decision for DSpace was made based on functional/technological and organizational factors (see below).

The DSpace software was developed by the MIT Libraries and Hewlett Packard, with the initial release dating from 2002. Today it is stewarded by DuraSpace, a not-for-profit organization formed in 2009 in a merger between the DSpace Foundation and Fedora Commons (see http://www.duraspace.org/history). The latest version of the software is 5.5, released in March 2016.² According to the Directory of Open Access Repositories (OpenDOAR, <u>http://www.opendoar.org</u>) DSpace is currently the most widely used repository software with over 1,400 current installations (see OpenDOAR information chart).

DSpace meets the functional requirements that were identified in D2.1 and expanded with the help of the user stories developed for stakeholder groups not covered in the online survey. Among the features of DSpace that lead to our decision to use it for the implementation of the KSP are the customizable metadata fields, flexible license management, free workflow definition, user-friendly implementation of multi-lingual user interfaces, full-text and faceted search/browsing (see Appendix 2 for detailed matching between requirements and software features)

In addition to looking at functional requirements, OSS Watch suggests employing the following criteria in the process of selecting an open source software (see Table 2).

Table 2: Open Source Software selection criteria

| Criterion ³ | (How) does DSpace meet the criterion? |
|---|--|
| Reputation : Does the software have a good reputation for performance and reliability? | GESIS's experience from implementing and maintaining SSOAR, datorium, and Leibniz Open is that DSpace is both reliable and has excellent performance. This is specifically due to its use of highly performant frameworks such as SOLR and Cocoon. In |

² Full software documentation for DSpace 5.x releases is available under

https://wiki.duraspace.org/display/DSDOC5x/DSpace+5.x+Documentation

³ All criteria and explanations quoted from <u>http://oss-watch.ac.uk/resources/tips</u>.

| | addition, DSpace is characterized by a very high flexibility, e.g. regarding the user interface. |
|--|---|
| Ongoing effort : Is there clear evidence of ongoing effort to develop the open source software you are considering? Has there been recent work to fix bugs and meet user needs? | The software has continually been developed further since its initial release in 2002. It has big and active user/developer community with well-defined communication channels allowing for the submission of bugs, the contribution of code, or the request of features. ⁴ |
| Standards and interoperability: Choose software which implements open standards. | DSpace supports relevant standards, including OAI-PMH, OAI-ORE, SWORD, WebDAV, OpenSearch, OpenURL, RSS, ATOM (<u>http://www.dspace.org/why-use</u>). |
| Support (Community) : Does the project have an active support community ready to answer your questions concerning deployment? | DSpace has an efficient community support infrastructure, including a number of very active mailing lists (<u>http://www.dspace.org/community</u>). |
| Support (Commercial): Third party commercial support is available from a diversity of companies | DSpace works with "Registered DSpace Service Providers (RSP)". These are companies who "have made an investment in the DSpace technology and a commitment to work cooperatively with DuraSpace organization to best serve the community of users" (http://www.dspace.org/service-providers). |
| Version : When was the last stable version of the software released? | Version 6.0, released in October 2016 |
| Documentation : [] You should be able to trace a clear history of bug fixes, feature changes, etc. | Comprehensive documentation is available on the DuraSprace Wiki: https://wiki.duraspace.org/display/DSDOC6x/DSpace+6.x+Docu mentation . JIRA is used for bug tracking: https://jira.duraspace.org/projects/DS/issues/DS-2325?filter=allo penissues |
| Skill set : Consider the skill set of yourself and your colleagues. Do you have the appropriate skills to deploy and maintain this software? | The technical implementation of the KSP will be carried out by the GESIS department "Knowledge Technologies for the Social Sciences" which has considerable experience with DSpace repositories. Currently, three DSpace repositories are managed at GESIS: SSOAR (<u>http://www.ssoar.info</u>), datorium (<u>https://datorium.gesis.org/xmlui/</u>), and Leibniz Open (<u>http://www.leibnizopen.de/1/home/</u>). |
| Licence | DSpace is distributed under a BSD license (http://www.dspace.org/license). |

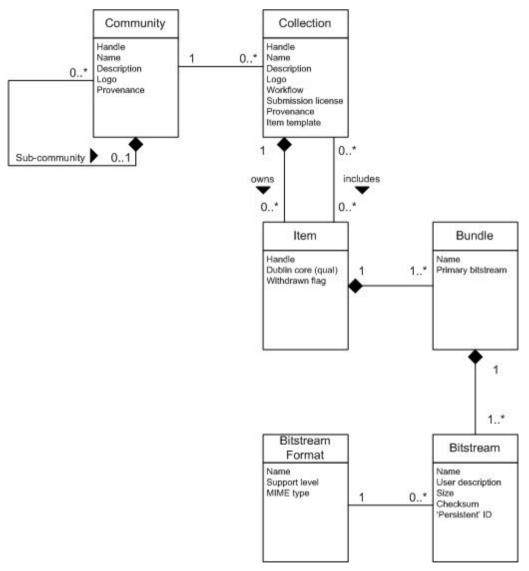
Another factor influencing the decision for DSpace is the fact that GESIS, responsible for the technical implementation of the platform, has long-standing expertise in implementing and maintaining DSpace platforms, both for textual resources and for research data. This also means that routines and solutions for required workflows and functions already exist (e.g. automatic DOI registration with da|ra, suggested citation) and can therefore be re-used for the CESSDA Knowledge-Sharing Platform. This makes the implementation process more efficient.

⁴ See <u>https://wiki.duraspace.org/display/DSPACE/How+to+Contribute+to+DSpace</u>.

2.3 DSpace Data Model and Content Organization

The DSpace data model distinguishes the following types of elements, organized in a hierarchical fashion (see Figure 1). On the highest level, there are *communities*, which contain one or several *collections*. *Collections* are composed of *items*, which can appear in more than one collection. Each item is composed of bitstreams which may be grouped into *bundles* if they are closely related (e.g. all bitstreams that make up a webpage with text and images would be considered a bundle). Each bitstream is associated with a *bitstream format*.

Figure 1 DSpace data model



(Source: DSpace Documentation 5.x,

https://wiki.duraspace.org/display/DSDOC5x/Functional+Overview#FunctionalOverview-DataModel)

While the way that content organized into communities and collections does not have a great impact on the discovery of content (which is very flexibly handled by the DSpace discovery system based on the metadata), it has some repercussions for administering the platform and for use scenarios from the organizational rather than individual user perspective. Thus the following points need to be considered:

1) If the repository ingests different resource types that need to be described with different metadata, collections should reflect resource types. Each resource type collection can have a "tailored" set of metadata

fields associated with it. Among other things this can make it easier for depositors to fill out the submission form (e.g. because they do not have to scroll through long lists of irrelevant metadata fields), and it can help to increase metadata quality (e.g. because there is more flexibility in designating metadata fields as mandatory or optional based on resource type). As we expect different types of resources to be submitted to the KSP - ranging from reports and presentations to audiovisual and web resources - resource types should be reflected on the collection level.

2) A second important question that needs to be considered is the representation of Service Providers in the structure of the KSP. The organization of data in DSpace was originally designed to reflect the typical structure of universities divided into colleges, departments, chairs, etc. This can be used to represent the structure of CESSDA in the content organization hierarchy as well. Thus, each Service Provider and CESSDA Main Office can be represented as a community. This approach has two advantages and one drawback:

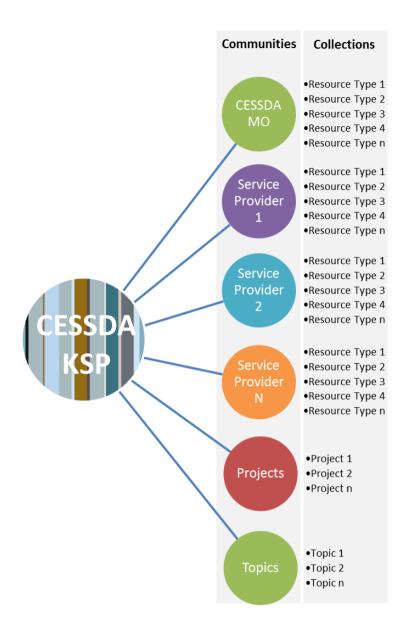
- All publications of one SP can be accessed through a single page, which can for example be equipped with a logo and could be used by the SP to showcase its publications.
- In the unlikely event that for whatever reason a need arises to remove (publications by) one SP from the portal entirely, the community structure allows for this without major effort.⁵
- A drawback of this approach is the additional effort that needs to be made in setting up the communities and collections. However, as the creation of collections and communities is not complicated in DSpace, and seeing as it is mostly a one-time effort during the implementation phase, this is not a significant problem.⁶

In addition to Service Provider community pages it is suggested to create a community for curated collections on different topics, which include lists of "hand-picked" resources on a certain subject, as well as a community for projects, which can contain collections of outputs of bigger collaborative projects (e.g. SaW itself). Taking this into account, the community-collection structure would look as shown in figure 2.

Figure 2: CESSDA KSP content organization

⁵ It should be noted that by assigning a DOI to the resources in the KSP, we guarantee that at least the landing page of the resource will remain accessible.

⁶ In the event that access to communities and collections needs to be controlled in a more differentiated fashion, this will make it necessary to create an access control model (e.g. RBAC). This can be handled with DSpace without problems, however.

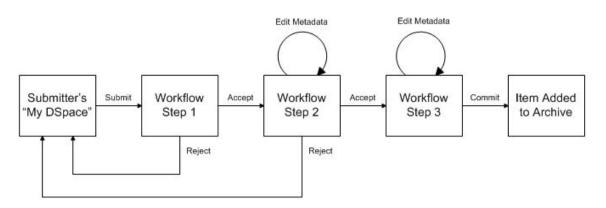


2.4 DSpace Publication Workflow and Roles

The DSpace publication workflow can have up to three steps before items are published in a collection (see Figure 3). Different steps can be assigned to different groups of persons. For the CESSDA KSP we propose a procedure consisting of two steps:

1) After a depositor submits an item, it is reviewed by an editor, who can accept or reject an item and who can also edit metadata if required. 2) When this review is completed, the item is reviewed and then published by an administrator.

Figure 3: Submission workflow in DSpace



Source: https://wiki.duraspace.org/display/DSDOC5x/Functional+Overview#FunctionalOverview-WorkflowSteps

In addition, DSpace allows the assignment of different roles to those using the system. The distribution of roles and responsibilities for the CESSDA KSP is detailed in Table 3.

Table 3: Roles and associated rights

| Group | Rights |
|-------------------------------|--|
| Anonymous | Read all resources and (bibliographic) metadata |
| Registered users ⁷ | Read all resources and (bibliographic) metadata |
| Depositors | Read all resources and (bibliographic) metadata Submit resources to collections |
| Editors | Read all resources and (bibliographic) metadata Submit resources to collections Edit items (metadata) in collections (incl. DOI assignment) Withdraw items from collections Map items into a collection |
| Administrators | Read all resources and (bibliographic) metadata Submit resources to collections Edit items (metadata) in collections (incl. DOI assignment) Withdraw items from collections Map items into a collection Assign registered persons to groups Assign rights to groups Create communities Create collections Publish resources |

2.5 METADATA SCHEMA

The metadata used to describe the resources collected in the Knowledge-sharing Platform (KSP) are crucial for the administration, discovery, and citability of resources. The more detailed information is recorded about resources published through the platform, the easier these tasks become. However, a balance needs to be struck between the desired amount of metadata and the time that depositors are required to spend on the submission of resources. This balance can be achieved by making only those metadata fields a requirement that are absolutely necessary to fulfill the purposes mentioned above, but giving depositors the opportunity to contribute further metadata if desired.

⁷ Registered users can create email alerts for collections, but need not have the same rights as depositors.

To develop the metadata schema for the CESSDA KSP, a number of other schemas was consulted and other repositories were consulted. The schemas and repositories consulted included

- da|ra Metadata Schema version 3.1 (<u>http://www.da-ra.de/en/technical-information/doi-registration/</u>)
- DataCite Metadata Schema version 4.0 (<u>http://schema.datacite.org/</u>)
- Dublin Core Library Application Profile (<u>http://dublincore.org/documents/library-application-profile/</u>)
- OpenAire (<u>https://guidelines.openaire.eu/en/latest/literature/application_profile.html</u>)
- Open Science portal Slovenia (<u>http://www.openscience.si/default.aspx</u>)
- Nordic Health Data (<u>https://github.com/NordicHealthData</u>)
- LeibnizOpen (<u>http://www.leibnizopen.de/1/home/</u>)
- Social Science Open Access Repository (SSOAR) (<u>http://www.ssoar.info</u>)
- EconStor (<u>https://www.econstor.eu/dspace/</u>)

In addition, the requirements of the linked CESSDA-SaW Task 5.2 were taken into account.

On this basis, a metadata schema was developed which is currently in the process of being finalized.

2.6 DOI REGISTRATION

The SaW description of work explicitly requires that resources published through the KSP receive a Digital Object Identifier (DOI), registered through da|ra (<u>http://www.da-ra.de</u>). DOIs enable the unambiguous identification and easy citation of resources. They can be assigned under the condition "that Digital objects referenced via an identifier must be permanently and constantly available, without interruption, under the registered address" (da|ra Policy 5.4.1⁸). Before the CESSDA KSP can begin registering DOIs with da|ra, a Service Level Agreement⁹ has to be signed.

2.7 IMPLEMENTATION PROCESS (TECHNICAL TASKS)

Table 4: Platform Implementation Process

| From | То | Description | Partner(s) |
|------|-----|--|-------------------------------|
| M11 | M15 | Description of functional requirements Definition of data model Definition of metadata schema Definition of licensing scheme Set up da ra cooperation for DOI registration | ADP, SND, GESIS, CESSDA AS |
| M16 | M18 | • Programming/implementation of platform | GESIS |
| M19 | M21 | Beta-testing | All |
| M21 | M21 | Launch of platform on the CESSDA webpage | GESIS, CESSDA AS |

3. DEVELOPMENT OF PLATFORM CONTENT

3.1 REQUIREMENTS

As detailed in D2.1 (see chapter 3.3), the survey carried out among future users and depositors of the Knowledge-sharing Platform suggests an interest in a broad range of resources types and topics. A strong interest exists in resources that can be used for training, "Guidelines and manuals" in particular. Other desired

⁸ <u>http://www.da-ra.de/en/about-us/da-ra-policy/policy/</u>

⁹ http://www.da-ra.de/en/about-us/da-ra-policy/service-level-agreement/

resources included case studies and best practices, technical specifications (e.g., for use of DDI and other metadata frameworks), and encyclopedia type resources. Similarly, the survey showed an interest in a broad range of topics across the curation life-cycle, with a somewhat lower interest only for resources covering the pre-ingest and ingest phases, as well as for resources on Persistent Identifiers. In consequence, the collection policy defines a broad scope of topics and resource types (see below).

3.2 Collection Policy and selection criteria

The collection policy is a document that describes which types of resources on which topics are collected by the CESSDA Knowledge-sharing Platform (KSP). This fulfills a twofold objective: Firstly, it helps to give those using the platform to access or deposit resources a better idea of the kind of content they can expect in the platform. Secondly, it supports the editors, who decide whether a submitted resource is accepted, in determining whether the resource fits the scope of the collection. To further support this decision process, selection criteria are derived from the definition of the scope of the collection.

The collection policy and selection criteria for the KSP are embedded in a broader policy document detailing the way the platform will be operated. See Appendix 3 for the draft version of this Collection and Operation Policy.

3.3 CONTENT DEVELOPMENT

The main focus of Task 2.1 is the conceptualization and implementation of the platform. However, as part of the implementation it will be necessary to test all aspects of the platform - organizational and technical workflows in particular. This is only possible with "real" content, so that the process of content development will begin as we begin to implement the platform.

In preparation for testing the platform a call for the preparation of content will be sent out to CESSDA Service Providers and other potential depositors part of SaW (e.g. CESSDA observers or archives that were part of the CESSDA network before it became CESSDA AS). The call will include information on the collection policy and selection criteria of the KSP as well as on legal aspects to keep in mind (specifically, Intellectual Property Rights). When the testing phase begins, all partners in Task 2.1 - in particular those with only 0.5 PM in the task will deposit the prepared content.

In this phase, an additional focus will be put on two other sources for content to be deposited:

- Firstly, output from other SaW work packages will be identified and prepared for inclusion in the platform. In particular, this will concern the training resources created in Task 4.1 as well as the outputs of Task 5.2, which has close links with Task 2.1.
- Secondly, we will use resources created as part of the CESSDA Expert Seminars to gain a better understanding of the issues that need to be dealt with when acquiring resources for the KSP that were created in the past by authors who may or not still be employed at a CESSDA SP. CESSDA Expert Seminars have taken place annually since 1987 (see http://cessda.net/CESSDA-Services/Resources/Events/Expert-Seminars). Resources associated with the Expert Seminars include presentations but also websites with schedules. Among the issues that we expect to encounter are difficulties with clarifying the situation regarding Intellectual Property Rights and obtaining the rights from the respective rights holder to publish the materials via the platform.

3.4 LICENSING AND DEPOSIT AGREEMENT

To avoid future disagreement about the usage and ownership of the resources presented in the Knowledge-sharing Platform (KSP), there is a need for common understanding about the legal rights of the respective partners involved. There is also a need for the users of the KSP to know how they can use the resources. This is resolved by using commonly used licensing systems. In this context, two relationships need to

be considered: 1) that between depositor and the KSP, and 2) that between the KSP and those who access and download resources from it.

Thus, the KSP has to obtain the depositor's permission to store and disseminate a given resource. At the same time, depositors have to confirm that they have all rights required to grant the KSP the respective usage rights and that no rights of third parties are being infringed. In turn, users who access and download resources via the KSP have to commit to using resources only in the way permitted by the rights holder.

While relationship 1) will be governed by a "Deposit and licensing agreement" (see Appendix 4), relationship 2) will be governed by the license assigned to each resource by the depositor.

The stakeholder survey carried out for D2.1 showed very clearly that there is a strong demand for resources to be as openly accessible as possible. Therefore, in the KSP guidelines for depositors (yet to be developed), it will be crucial to highlight the benefits and the importance of publishing resources under open licenses if at all possible. This is also reflected in the Collection and Operation Policy (Appendix 3).

At the same time the survey hinted that many respondents did not have detailed knowledge of copyright regulations and open licences. Therefore it seems necessary to provide general guidance on this topic for both users and depositors to make sure that they understand the implications of the respective licenses for use.

4. MAINTAINING THE PLATFORM POST-PROJECT

One of the express objectives of establishing the Knowledge-sharing Platform is the systematic archiving of CESSDA resources. Thus, the platform must become a long-term effort of CESSDA AS. Accordingly, its continued existence and development must be guaranteed, both on a technical and a content level.

4.1 PLATFORM HOSTING

Hosting the DSpace platform will require running a server (LINUX) and about 1 PM/per year for administrative/maintenance tasks. Currently, two options for hosting are being discussed with the CESSDA Technical Working Group: 1) The KSP will be hosted by CESSDA as part of its future cloud infrastructure platform (2017). 2) GESIS will host the server - during a pilot phase, or indefinitely.

4.2 PLATFORM ADMINISTRATION

Administrators have responsibilities in three areas: 1) user/rights management, 2) collection management, and 3) publication of resources (see table 5).

Table 5: Administrative tasks

| User/rights management | Register new users Assign users to user groups associated with certain rights (e.g. read, add, write; see table 3 above) Remove users |
|------------------------|---|
| Collection management | Create new communitiesCreate new collections |
| Resource publication | Publish resources that were positively reviewed by editors |

The administrative work will be shared between CESSDA Training and CESSDA MO. It is expected that initially the workload will be a little higher, as a greater number of new users will likely register, resources published or created in the past will be continued to be added to the platform, and SaW project results will be published through the platform as well. However, after this initial phase post-project, the expected workload per year does not exceed 1 PM (to be shared between CESSDA Training and CESSDA MO).

4.3 SUBMISSION MANAGEMENT AND COLLECTION DEVELOPMENT

In order to represent the interests of CESSDA SPs but also to support CESSDA MO and CESSDA Training in maintaining the platform an Editorial Committee will be formed, consisting of KSP administrators, editors, and liaisons.

As described in table 3 above, editors fulfill tasks in submission management. They

- reject or accept submissions based on the collection policy and selection criteria;
- review and edit metadata if necessary;
- assign DOIs.

The expected overall workload (to be shared between all editors) is 1-1.25 PM per year.

To support editors, especially in regard to collection building, each SP should assign a liasion. Liasions monitor the CESSDA-related public output created by staff of the SP, and either encourage creators to submit the resources or get the permission to deposit resources for them. The workload somewhat depends on the size of the SP and its publication output, but should overall be very small, ranging from 0.1 to a maximum of 0.25 PM per SP per year.

Together, all members of the Editorial Committee should actively support and promote collection building/development. This can be done, on the one hand, by encouraging colleagues to submit resources after CESSDA Expert Seminars or other (training) events. On the other hand, members of the Editorial Board can be responsible for certain topics ("curated collections") and enhance these by adding metadata records for relevant resources from outside CESSDA to the Knowledge-Sharing Platform.

| Group | Number of persons | Selection process | Duration of commitment |
|----------|-------------------|---|---|
| Liasions | 1 per SP | Self-selection/assigned by SP. | Suggestion: 2 years. Individual agreements can be made. |
| Editors | 4 | SPs should be asked whether they are willing to contribute human resources for this task, and which amount. This will be different for smaller and bigger SPs. | Suggestion: 2 years. Individual agreements can be made. Rotating among SPs if desired; while rotation is desirable, it should be kept in mind that it will make training of new editors necessary whenever a change occurs. So longer-term commitments are encouraged |
| Admins. | 2 | Assigned by CESSDA MO | Indefinitely/not fixed |

Table 6 Composition of the Editorial Committee

Appendix 1: User Stories and Functional Requirements

1. Deposit

| As a depositor | Corresponding features/functions | Technical solution? | Priority according to D2.1 |
|---|--|---------------------|----------------------------------|
| I want easily accessible information on which types of resources are accepted by the KSP so that I don't submit something out of scope. | Collection policy Selection criteria | No | n/a |
| I want to upload resources to the KSP in a simple procedure so that I don't have to spend too much time on this. | Easy registration procedure Simple upload procedure Only necessary amount of mandatory metadata fields | Partly | Moderate |
| I want the registration process to be simple because I don't want to spend much time on "formalities". | Registration with an email address associated with a CESSDA SP | Partly | n/a |
| I want to freely choose the license under which a resource is shared because I want to have control over how a resources is distributed. | Default licenses to choose from Custom license text | Partly | n/a |
| I want to understand what a license means even if I am no copyright expert. | Licensing FAQ/Guideline | No | n/a |
| I want to be able to get support with uploading resources in case I have questions so I don't have to spend more time than necessary getting familiar with the system | User manual Short explanatory texts Helpdesk | Partly | n/a |
| I want my shared resources to receive a PID so that they are easily citable. | (Semi-)automated procedure for DOI registration and assignment | Yes | High |
| I want to be notified when a resource I uploaded is published so I don't have to go to the platform and check again and again myself. | Email notifications when status of deposited resource changes | Yes | n/a |
| I am interested in knowing how often a resource I uploaded was accessed. | Usage statistics | Yes | Moderate |

2. Access

| As a user | Corresponding features/functions | Technical solution? | Priority according to D2.1 |
|--|--|---------------------|----------------------------------|
| I want to have a clear idea of the kind of resources and information that I can find on the platform. | Collection policy | No | n/a |
| I want to discover openly accessible resources on the KSP and elsewhere on the web so I don't have to search in two or more places. | Simple searchAdvanced searchBrowsing | Yes | High |
| I want to search the KSP without having to think about metadata fields or controlled vocabularies because I want the search process to be as easy and comprehensive as possible. | Simple searchFull text search | Yes | Moderate |
| I want to easily download interesting resources to use them in my work. | Access to metadata and resources without registration/log-in or other barriers Limit search to resources where full text is available | Yes | High |
| I want to browse topics to discover interesting resources. | Browse by topic Facets to drill down search results | Yes | High |
| I want to specify search criteria so that I can find exactly what I am looking for. | Advanced search Simple search + drill down via facets | Yes | High |
| I want to browse curated collections on topics to discover interesting resources. | Curated collections on topics | Yes | High |
| I want to download bibliographic metadata to import it into my reference management software for easy referencing and citing resources | Export function for metadata in BibTex format | Yes | High |
| I want guidance on how to cite a resource so I can easily reference works in my writing. | "Suggested citation" automatically generated from the metadata | Yes | High |
| I want to be alerted to new resources on a certain topic by email, so I can keep up to date. | Alert functions for registered users | Yes | Moderate |
| I want easy access to the KSP from the CESSDA Homepage so that I don't have to visit too many different pages when looking for information. | Link to KSP prominently from the CESSDA homepage Search KSP directly from CESSDA homepage | Yes | High |

| I want the KSP and its resources to be as openly accessible as possible so that I don't have to waste my time with complicated access procedures and so that I can use the resources from the platform easily in my work. | Access to metadata and resources without registration/log-in or other barriers Guidance concerning the meaning of a specific license | Yes | High |
|--|---|--------|------|
| I want to be able to search for resources in a specific language to make sure that the search results are relevant to me. | Limit search by language (metadata) | Partly | n/a |

3. Management of content

| As an editor | Corresponding features/functions | Technical solution? | Priority according to D2.1 |
|--|--|---------------------|----------------------------------|
| I want to be notified when new resources are submitted to the KSP so I can review them. | Alert function as part of review workflow | Yes | n/a |
| I want a set of criteria that helps me decide whether a resource falls within the scope of the KSP or not, so I can make the decision to reject or publish a resource easily and dependably. | Selection criteria (collection policy) Guidelines for editors | No | n/a |
| I want to be able to review and edit the metadata of a submitted resource before publication to ensure good quality of the resource description. | Platform publication workflow: edit submission | Yes | n/a |
| I want to be able to assign a DOI to a submitted resource accepted for publication in a simple process. | Platform publication workflow: register DOI | Yes | n/a |

4. Management of platform (incl. user management)

| As an editor | Corresponding features/functions | Technical solution? | Priority according to D2.1 |
|---|--|---------------------|----------------------------------|
| I want to be able to get in touch with depositors to let them know that their resource needs further editing. | Publication workflow: Accept/reject submission | Yes | n/a |

| As an administrator Corresponding features/functions | Technical solution? | Priority according to D2.1 |
|--|---------------------|----------------------------------|
|--|---------------------|----------------------------------|

| I want to have easy access to all information necessary to decide if a newly registered depositor falls within the scope of the KSP's intended community of depositors. | List of organizations and email domain names | No | n/a |
|---|---|-----|-----|
| I want to accept or reject depositors who have registered with the platform to give me control over who can publish things in the KSP. | Registration workflow | Yes | n/a |
| I want to be able to assign editors to take care of the publication workflow for certain communities and collections | User management routine | Yes | n/a |
| I want to be able to customize the design of the KSP to adapt it to the CESSDA corporate design. | Customizable style sheets | Yes | n/a |
| I want to generate statistics about the size of the collection and its use to better understand the interests and needs of the user community. | Statistics for Items (by topic, by institution) on the platform Number of users Use by topic | Yes | n/a |

5. Other

| User story | Corresponding features/functions | Technical solution? | Priority according to D2.1 |
|--|---|---------------------|----------------------------------|
| As a user, I want to have the possibility of suggesting resources to the editorial committee for inclusion in the platform without too much effort. | Form on the webpage that allows suggestion of resources | Yes | n/a |

Appendix 2: Functional requirements and Corresponding DSpace 5.x functions

1. Deposit

| Required features/functions | DSpace 5.x support | Priority according to D2.1 |
|--|--|----------------------------------|
| Easy registration procedure | Different registration and authentication processes supported - self registration by web form and subsequent log in with user name and password, or authentication via LDAP or Shibboleth. | Moderate |
| Simple upload procedure | The following features can be used to make the deposit procedure as easy/efficient as possible: Fully customizable metadata Choice management and authority control to help users entering controlled metadata values Customizable licenses, including support for Creative Commons; | Moderate |
| Designation of mandatory/optional metadata fields | Supported. Fully customizable. | Moderate |
| Flexible licensing of content | DSpace offers support for Creative Commons; license texts are fully customizable by both admins (to define options to choose from) or by users who want to enter a specific license text. | n/a |
| Explanatory/help texts in context | Short info texts can be shown when hovering the cursor over a given word/icon on the page. | n/a |
| (Semi-)automated procedure for DOI registration and assignment | By default, DSpace uses the CNRI Handle System to assign globally unique identifiers to communities, collections and items. The GESIS WTS department has developed a routine that allows to assign a DOI to items in the repository using the da ra service. This functionality will be re-used for the CESSDA KSP. | High |
| Email notifications when status of deposited resource changes | Supported. | n/a |
| Usage statistics for individual resources | Supported. | Moderate |

2. Access

| Required features/functions | DSpace 5.x support | Priority according to D2.1 |
|-----------------------------|---------------------------------------|----------------------------------|
| Simple search | DSpace uses the search engine Lucene. | High |

| Advanced search | The search function is fully customizable: keyword search in all or limited to specific metadata fields is supported. | |
|--|---|----------|
| Full text search | Supported for text based contents | Moderate |
| Facets to drill down search results | Faceted browsing is supported for any metadata field. | High |
| Browsing | Users can browse the following indices: title, author, issue date, subject terms. It is possible to limit browsing to specific parts of an index, e.g. to only browse items in a particular collection or community. | High |
| Access to metadata and resources without registration/log-in or other barriers | Anonymous discovery and retrieval are supported. | High |
| Collections on topics | DSpace supports the creation of so-called "Communities" and "Collections". These can be used to organize the items stored in the repository, e.g. by topic. A community contains collections, defined as "groupings of related content". It is possible for a collection to appear in different communities. | High |
| Export function for metadata in BibTex format | Not supported out of the box. Implementation effort: 0.25 PM | High |
| "Suggested citation" automatically generated from the metadata | Not supported out of the box. Implementation effort: 0.25 PM | High |
| Alert functions for registered users | Registered users can subscribe to collections in DSpace and are notified by e-mail about new items in these collections. | Moderate |

3. Management of content

| Required features/functions | DSpace 5.x support | Priority according to D2.1 |
|--|---|----------------------------------|
| Alert function as part of review workflow when new items are submitted | Supported | n/a |
| Platform publication workflow: edit metadata of submission | Supported | n/a |
| Platform publication workflow: register DOI | The GESIS WTS department has developed a routine that allows to assign a DOI to items in the repository using the da ra service. This functionality will be re-used for the CESSDA KSP. | n/a |

4. Management of platform (incl. user management)

| Required features/functions | DSpace 5.x support | Priority according to D2.1 |
|---|--|----------------------------------|
| Publication workflow: Accept/reject submission | Supports a customizable submission workflow which allows for the rejection or acceptance of submissions. | n/a |
| Registration workflow: accept/reject depositors | Supported. "Deposit" rights have to be actively assigned to newly registered users. | n/a |
| Assigning roles and associated rights to users | DSpace supports an authorization system based on so-called "Resource Policies". These policies are used to determine which actions can be performed for a given object and which group of users has permission to do so. Users are given a set of permissions by assigning them to a certain group. | n/a |
| Customizable style sheets | DSpace supports two user interfaces, one is based on JavaServer Pages (JSP), the other one on the Apache Cocoon framework (XMLUI). These can be customized to match the CESSDA corporate design. | n/a |
| Statistics for Items (by topic, by institution) on the platform Number of users Use by topic | Usage metrics are provided based on SOLR. Statistics are available on item, collection, and community levels.They include page visits and downloads. In addition, system statistics allow to generate reports on the content and use of the repository drawing on DSpace's log files. This includes user logins and popular searches among other things. | n/a |

5. Other

| Required features/functions | DSpace 5.x support | Priority according to D2.1 |
|---|---|----------------------------------|
| Integration with CESSDA Webpage | Depending on the preferred solution, this will require an implementation effort of 0.5 PM | n/a |

APPENDIX 3: COLLECTION AND OPERATION POLICY (DRAFT)

1. Introduction

The CESSDA Knowledge-sharing platform (KSP) is a central point of access for the body of knowledge created by CESSDA partners. This policy is intended to serve as a guide to the development, management and care of the collection to achieve the overall mission of the KSP. It defines mission, purpose, scope, selection criteria and stakeholders' roles.

2. Mission

Over the years, CESSDA partners have created a significant body of knowledge on a broad range of topics. The KSP is intended to support the systematic and structured description of these resources (1). It thereby aims to promote a broad exchange of knowledge among CESSDA Service Providers (SPs) and the communities they serve (2). To foster knowledge exchange in the best possible way, resources should be freely accessible to everyone to use and republish as they wish, with as little restrictions from <u>copyright</u> or other mechanisms of control as possible. The KSP should be as open and barrier free as possible (3).

The KSP aims to become a special collection: in a specifically defined field of knowledge that strives to be exhaustive, as far as is reasonably possible:

- Exhaustive collection of digital materials
- Extensive collection of records about published materials (metadata).

3. Purpose

The KSP serves the systematic collection, registration, description of, and long-term access to different resources produced by CESSDA SP or for CESSDA SP with the purpose of fostering knowledge-exchange and skills transfer in the European Social Science Data Archiving landscape.

The KSP minimizes the risk that resources are lost or become inaccessible.

4. Scope

The KSP mainly keeps a current focus, but it will also seek to gather resources retrospectively. This includes documentation of CESSDA Expert Seminars and other CESSDA events.

The KSP focuses on digital resources, but also maintains metadata records about publications or other related materials.

The KSP acquires materials in the national languages of the CESSDA SPs, however, a minimum metadata should be provided in English for all materials included in the collection.

The Collection consists of:

- resources describing or covering different areas of work in social science data services, such as: Data Access, Dissemination, and Open Data, Archiving (Curation and Preservation), Metadata and Standards, Research Data Management, Pre-ingest, Data Acquisition, Data Processing and Documentation, Ingest, Training, Management of Data Archives, Data Protection and Ethics, Persistent Identifiers, and Other.
- Different resource types, such as: Training resources: Guidelines or manuals, Training resources: Webinars, Training resources: e-tutorials, Software tools, Policy or advocacy documents, Presentations, Reports and also User satisfaction surveys, Scholarly publications (e.g. articles, collections, monographs), Blog posts or other social media, other Training resources.
- Materials in all languages used by CESSDA SPs, with a minimum of descriptive metadata in English.

5. Responsibility

The KSP takes responsibility for the long-term access to and protection of resources in the collection. Trainings and management are developed accordingly by the CESSDA Main Office in cooperation with the Editorial Committee.

6. Selection criteria

- 1. <u>Scope:</u> Resource should be related to the scope as defined in this document.
- 2. <u>Content</u>: Resource should be relevant to other CESSDA members or observers, it should bring new (added) value. Duplication of documents with the same or very similar content should be avoided, exceptions: resource is offered in a different language.
- 3. <u>Language</u>: Resource can be in any language used by CESSDA members, but a minimum of descriptive metadata should be provided in English.
- 4. <u>Copyright</u>: Resource should preferably use an open licence (e.g. Creative Commons or GPL). If no open license is provided, the KSP may decide only to record the metadata of the resource.
- 5. <u>Duplication:</u> If a resource has been already published elsewhere and received a PID, the KSP collects only metadata.

Selection criteria are to be used by (potential) depositors and editors.

7. Stakeholders

Stakeholders perform different roles for the KSP. These are distinguished as follows:

- Depositors: submit resources to the platform.
- Users: search for and download content from the platform.
- Editors: manage the content-side of the platform.
- Administrators: manage the technology-side of the platform

Stakeholders come from:

- CESSDA Main Office
- KSP Editorial Committee
- CESSDA Service Providers
- CESSDA Observers and aspiring Service Providers
- Other (social science) data archives
- Social science educators
- Social science researchers
- Academic support staff (administration, library staff, etc.)
- CESSDA Members (i.e. ministries)
- Policy makers on national and European level
- Other

8. Platform use and management

Stakeholders as defined in the following are invited to share and access resources, for example to support the process of setting up or running a data service as a service provider to CESSDA.

8.1 Depositors

Depositors come from:

- CESSDA Main Office
- KSP Editorial Committee
- CESSDA Service Providers
- CESSDA Observers and aspiring Service Providers

Depositors submit resources to the platform. They:

1) Check if the resource fits selection criteria

2) Provide metadata and description in English, even if the resources is in a language other than English.

To support the deposit process, it is recommended that CESSDA SPs assign a liaison responsible for depositing resources to the KSP.

8.2 Editors

Editors come from:

- CESSDA Main Office
- KSP Editorial Committee

Editors manage the content-side of the platform.

Editors follow selection criteria to reduce personal bias by setting individual selection decisions and to ensure continuity and consistency in selection and revision.

Editors are elected by CESSDA Main Office / Board of Directors / General Assembly for the mandate of two years. They can be re-elected.

Number of editors and their mandate is determined by CESSDA Main Office.

8.3 Administrators

Administrators manage the technology-side of the platform. They are appointed by CESSDA Main Office.

8.4 Users

Anyone can use the resources collected in the KSP. However, it is envisioned that users mainly come from:

- CESSDA Main Office
- KSP Editorial Committee
- CESSDA Service Providers
- CESSDA Observers and aspiring Service Providers
- Other (social science) data archives
- Social science educators
- Social science researchers
- Academic support staff (administration, library staff, etc.)
- CESSDA Members (i.e. ministries)
- Policy makers on national and European level

Users search for and download content from the platform.

The KSP distinguishes Anonymous and Registered Users.

KSP resources used in (research) publications must be cited accurately and in sufficient detail. Sources are cited within the text, within tables and graphs, and in publication references.

8.5 KSP Editorial Committee

The role of the Committee is to vouchsafe the long-term development and management of the platform, and act as an instance of quality assurance.

The Editorial Committee consists of KSP administrators, editors, and liaisons.

If no other agreements are made, members of the Editorial committee are appointed for a period of 2 years.

9. Revision of the document/policy

Policy is to be reviewed and updated according to the development and needs.

Appendix 4: Deposit and licensing agreement for resources uploaded to the

CESSDA KNOWLEDGE-SHARING PLATFORM (DRAFT)

This deposit agreement is modeled on the contract used by the University of Gothenburg in its library's open access system. It has been checked by the University lawyer for validity.

1. Agreement for publishing on the CESSDA Knowledge-Sharing Platform

The Owner of the Resource undertakes to grant the right to the Consortium of European Social Science Data Archives (CESSDA) to publish the Resource on the CESSDA Knowledge-Sharing Platform (KSP). The right to publish includes the right to make the Resource available. This agreement shall not impose any limitations of the Owner's right to make use of the Resource.

2. Duration and cancellation of the agreement

The agreement shall remain in force until the Owner notifies CESSDA that the right shall be cancelled. Upon a cancellation CESSDA no longer has the right to publish the Resource on the CESSDA KSP and shall immediately remove the Resource. CESSDA can, on its behalf, cancel the agreement with immediate effect and hence remove the Resource from the KSP. No motivation needs to be stated for the removal.

3. Origins of the Resource

The Owner shall ensure that she/he/they has the right to use included material where appropriate (illustrations etc), and has the right to dispose of the Resource for publication in accordance with the present agreement.

Should it be known that the Owner does not have the permission to use part of the Resource or the Resource in its whole, or lack the necessary permissions as stated above, the Owner shall ensure to indemnify CESSDA.

4. Publication license

Please choose under which licence the Resource can be used by third parties.

- CC BY <u>https://creativecommons.org/licenses/by/4.0</u>/
- CC BY NC https://creativecommons.org/licenses/by-nd/4.0/
- CC BY NC SA <u>https://creativecommons.org/licenses/by-nc-sa/4.0/</u>
- CC BY NC ND <u>https://creativecommons.org/licenses/by-nc-nd/4.0/</u>
- Other _____ (please provide a name/description of the preferred license)

If you need help deciding on a license, use this license selector: https://ufal.github.io/public-license-selector/#

5. Reimbursement

No reimbursements are to follow from this agreement. Any associated costs are the responsibility of each part respectively.

6. Validity of the agreement

The present agreement shall come into force immediately after acceptance by the Owner.