cessda saw

Research and Innovation Action

CESSDA Strengthening and Widening

Project Number: 674939   Start Date of Project: 01/08/2015   Duration: 27 months

Deliverable 3.6 Final integrated audit report

Dissemination Level	Public
Due Date of Deliverable	31/07/2017
Actual Submission Date	31/08/2017
Work Package	WP 3
Task	3.2
Type	Report
EC Approval Status	16 November 2017
Version	V1.4
Number of Pages	p.1 – p.69

Abstract: This is a follow-up of D3.2 County report. Focus is on identifying what development steps can be proposed to tackle the obstacles in the way of achieving the widened and strengthened CESSDA ERIC. The maturity of data archive service (DAS) in most European countries was audited. The analysis contains elements of the wider data sharing ecosystem.

History

Version	Date	Reason	Revised by
0.1	30/04/2017	First draft results chapters	Task leader
0.2	16/06/2017	Editorial review	Subtasks lead partners
0.3	23/06/2017	Final integral version	Task leader
0.4	3/07/2017	WP leader and subtasks lead partners approval	IEN, FFZG, FORS, NSD, SND, UKDA
1.0	7/07/2017	Review of final document layout	Task leader
1.1	26/07/2017	Review by the WP leader	NSD
1.2	24/08/2017	Review by the Chair of the Delivery Committee	CSDA
1.3	30/08/2017	Review by the Project Coordinator	CESSDA MO
1.4	31/08/2017	Approval and Submission by the Project Coordinator to the European Commission	CESSDA MO

Author List

Organisation	Name	Contact Information
ADP	Janez Štebe (Task leader)	janez.stebe@fdv.uni-lj.si
FFZG	Marijana Glavica	mglavica@ffzg.hr
FORS	Brian Kleiner	brian.kleiner@fors.unil.ch
FORS	Christina Bornatici	christina.bornatici@fors.unil.ch
IEN	Aleksandar Zdravković	aleksandar.zdravkovic@ien.bg.ac.rs
IEN	Aleksandra Bradić-Martinović	abmartinovic@ien.bg.ac.rs
NSD	Trond Kvamme	trond.kvamme@nsd.uib.no
NSD	Vigdis Kvalheim	vigdis.kvalheim@nsd.no
SND	Ilze Lace	ilze.lace@gu.se
UKDA	Veerle Van Den Eynden	veerle@essex.ac.uk

Time Schedule before Delivery

Next Action	Deadline	Care of
Draft chapters reports ready	30/04/2017	All T3.2 partners
Editorial review	16/06/2017	Subtasks lead partners
1.0 Final integral version	23/06/2017	ADP
1.1 Review by the WP leader	23/08/2017	NSD
1.2 Review by the Chair of the Delivery Committee	29/08/2017	CSDA
1.3 Review by the Project Coordinator	30/08/2017	CESSDA AS
Approval and Submission by the Project Coordinator to the European Commission	31/08/2017	CESSDA AS

Executive Summary

Main focus of a final integrated audit report of the CESSDA SaW project TASK 3.2 - Audit of current status of data archive services in each European Research Area (ERA) country is to identify gaps and obstacles in the way of achieving the widened and strengthened CESSDA ERIC. All European countries CESSDA membership is a goal and the report explore the potential value and benefits from such membership for the member countries. It is a follow-up of a comprehensive study reported in the CESSDA SaW D3.2 County report1, in which the focus was on each of the countries individually. We suggest both reports to be read simultaneously, current giving a broader picture about conditions that affect typical situations that are common for groups of countries.

The maturity of data archive service (DAS) in most European country was audited based on a self-assessment using customised and shortened CESSDA SaW Capability Development Model (CESSDA-CDM). Twenty-three countries with existing DAS participated in this part. CESSDA service provider representative, or organisations expected to be in this role in the future provided information through Web survey form.

The overview of profile and organisational infrastructure shows that most organisations provide a publically available mission clearly declare to carry out main required functions of a typical data archive. While the ambitions and potentials for delivering fully flagged DAS service are common to all types of organisations, the analysis also shows a substantial variety on some of the aspects of maturity self-assessments, both among current CESSDA countries members and among those classified as 'aspiring' members. Countries should provide long-term funds for the establishment and functioning of the DAS for them to be able to fulfil the mission clearly stated in the documents.

The mapping also contains a review of elements of the wider data sharing ecosystem: interplay of structural conditions of social science development, funders open data policies and strategies, and data sharing culture and incentives that increase the data sharing habits of researchers. External stakeholders, in particular funders, can play an important role in improving the national data service sustainability. Research funders are the key stakeholder that can help to provide incentives and remove some of the barriers to data sharing. The biggest differences between CESSDA members and non-members are related to '*Overarching strategy and policy to enable sustainable access and sharing of social science data'*. Advanced policy recommendations, appropriate funding mechanisms and strong DAS can lead to a sustainable data-sharing ecosystem.

Finally, in the last chapter the following countries, where no formal DAS exist, were analysed regarding the potentials of integration of initial RDM support infrastructure: Albania, Austria, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Israel, Kosovo, Latvia, Luxembourg, Macedonia, Moldova, Montenegro, Serbia, Spain and Ukraine. By identifying proto-activities and open access support activities we detected actors and institutions that could play a key role in the elaboration of new national DAS. The list might be of a help to funders and CESSDA MO on national and international level, when planning further development of DAS activities for social sciences in ERA.

---

[1] See http://cessdasaw.eu/content/uploads/2017/07/D3.2_CESSDA_SaW_v1.3.pdf.

Abbreviations and Acronyms

AAI	Authentication and Authorization Infrastructure
CPA	Capability Process Area
CRA	Capability Requirements Area
DAS	Data Archive Service
DDI	Data Documentation Initiative
DCC	Digital Curation Centre
DMP	Data Management Plan
DOI	Digital Object Identifier
DSA	Data Seal of Approval
GERD	Gross Domestic Expenditure on Research and Development
GDP	Gross Domestic Product
IPR	Intellectual Property Right
OAIS	Open Archival Information System
PID	Persistent Identifier
RDM	Research Data Management

Table of Contents

1.   Introduction

Data Archive Services (DAS) do not exist in isolation. One need to have a wider frame of reference established, that includes different important stakeholders' roles. Finally, general conditions that determine status of social science research in the country need to be considered. In such an ecosystem, if it is supportive, DAS can develop and has a multiplying effect back to support further development of social sciences, and in particular with providing access to relevant high quality data enable tackling important societal issues.

The factors considered are:

- Structural factors, in particular the wealth of nations, that often determine support for science in general, including the social sciences, and the extent of the social science community and its research data output. All this provide supply and demand for the DAS: quality data produced and reused in further analysis.

- Broad area of enablers and constraints that influence the data sharing culture in the social science community, and that includes the role and responsibilities play by different stakeholders. In particular, research funders are responsible for establishing the open research data sharing policies. Existing or emerging data services that provide support for data sharing are an important part of the data-sharing ecosystem, together with other scientific information infrastructures, such as research libraries.

Whereas the first report[2] on the maturity of data archive service (DAS) in European countries described each individual country separately, the focus in this report is on the common configurations of factors that form a data sharing ecosystem that appear in groups of countries. Thus, emphasis here is on comparison of countries. The results of this analysis could serve as a point of departure for further investigation into weak points that are common for a certain group of countries. The main purpose of maturity measurement is to call for collective and coordinated action on the European level, where gaps are identified between current and desired states.

At the same time, responsible stakeholders in individual countries or regions need to determine internal goals, while comparing the current gaps in their countries with others, and act correspondingly. The group of countries identified can consider following best practice examples to achieve a mature and supportive open scientific data ecosystem. If principles about open access are agreed upon, both in the scientific community and among policy makers, the results presented here can motivate to find sustainable arrangements for the situation in a country. The intention is to use results to inform stakeholders about steps needed to achieve a certain maturity level.

1.1   Data sharing ecosystem

A metaphor of the data sharing ecosystem was used as an inspiration about how to approach in describing the complex situation regarding DAS in all the variety of European countries, some of them in the initial stage of considering how to start activities, others with long and well-

---

[2] See http://cessdasaw.eu/content/uploads/2017/07/D3.2_CESSDA_SaW_v1.3.pdf.

established data services, highly praised and accepted with the users. The term data ecosystem was coined in a similar context, aimed at describing a variety of data types and actors involved:

> *It is a complex system involving data collectors, stewards, and users as well as sponsors and stakeholders; emergent and historical transparent technologies; and ever-growing data along with their myriad associated artefacts. The system must be understood in totality in order to optimize the whole and not just the individual components.*[3]

If data is water that needs to flow through the system to optimise the quality and efficiency of research, then a data sharing culture is a key systemic component that determines the efficiency and sustainability of a data ecosystem. Much research has been done in the last decade across disciplines and at international level on research data sharing culture, data sharing and management practices, on barriers and enablers. This published literature provides us with much factual information on this topic, which is most likely applicable across all countries.

Both detailed qualitative studies and wider surveys assessing data sharing practices, barriers and enablers amongst researchers at a local, European or international level - some of which focus on specific research disciplines, others look across a range of disciplines – identify numerous perceived or real barriers to data sharing[4]:

- fear of competition, of being scooped and therefore reduced publication opportunities;

- cost in time and money to prepare data and documentation for sharing and absence of funding to do so;

- absence of professional rewards for data sharing;

- lack of standards and data infrastructure;

- ethical and legal constraints (no informed consent for sharing, IPR);

---

[3] Parsons et al. 2011. A conceptual framework for managing very diverse data for complex, interdisciplinary science. Journal of Information Science 37(6) 555–569, DOI: 10.1177/0165551511412705

[4] Borgman, C.L. 2012. The conundrum of sharing research data. Journal of the American Society for Information. Science and Technology 63: 1059-1078. http://dx.doi.org/10.1002/asi.22634; Carlhed, C., & Alfredsson, I., 2009. Swedish National Data Service's Strategy for Sharing and Mediating Data: Practices of Open Access to and Reuse of Research Data–The State of the Art in Sweden 2009. In IASSIST's 35th annual conference Tampere, Finland, May 26-29, 2009. http://www.diva-portal.org/smash/get/diva2:396306/FULLTEXT01.pdf; Kuipers, T. and van der Hoeven, J., 2009. PARSE.Insight: INSIGHT into issues of Permanent Access to the Records of Science in Europe. Survey report. http://docplayer.net/127428-Parse-insight-deliverable-d3-4-survey-report-of-research-output-europe-title-of-deliverable-survey-report.html; Kuula, A. and Borg, S., 2008. Open access to and reuse of research data: the state of the art in Finland. Finnish Social Science Data Archive, 31; Piwowar, H.A., 2011. Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data. PLoS ONE 6. http://plosone.org/article/info:doi/10.1371/journal.pone.0018657; Savage, C.J., and Vickers, A.J., 2009. Empirical study of data sharing by authors publishing in PLoS journals. PloSOne, 4(9): e7078; Sayogo, D.S. and Pardo, T.A., 2013. Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. Government Information Quarterly, 30(1): 19-31. http://dx.doi.org/10.1016/j.giq.2012.06.011; Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., and Frame, M., 2011. Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 6. http://dx.doi.org/10.1371/journal.pone.0021101; Van den Eynden, V. and Bishop, L., 2014. Sowing the seed: Incentives and Motivations for Sharing Research Data, a researcher's perspective. Knowledge Exchange. http://www.knowledge-exchange.info/event/sowing-the-seed; Wicherts, J. M., Borsboom, D., Kats, J., and Molenaar, D. 2006. The poor availability of psychological research data for reanalysis. American Psychologist, 61(7): 726; Youngseek, K. and Stanton, J.M., 2012. Institutional and Individual Influences on Scientists' Data Sharing Practices. Journal of Computational Science Education, 3(1): 47-56.

- fear of misuse / misinterpretation of data

- data not documented well enough to be usable.

Enablers of data sharing generally reported in the literature are:

- data sharing expectations of funders and journals

- peer expectations and sharing practices in the research community

- availability of data repositories and standards

- ethical principles and norms in scientific research (that encourage data sharing),

- desire to showcase data quality

- researchers' data management skills

- organizational support

- acknowledgement received for data sharing

- data publication and metrics.

Research also reveals disciplinary and research group differences in data sharing practices. Fewer studies have focussed specifically on motivations and incentives for researchers to share data. A recent report by the Expert Advisory Group on Data Access in the UK on incentives for data sharing, based on interviews with a small number of key stakeholders (research funders, senior academic managers, postdoctoral researchers, a chair of a Research Excellence Framework panel and a senior data manager) and a web survey with researchers and data managers, recommends as essential incentives, that research funders should: strengthen and finance data management and sharing planning requirements; continue funding and development of infrastructure and support services; recognise high-quality datasets as valued research outputs in the Research Excellence Framework; and establish career paths and progression for data managers as members of research teams. In addition, they recommend that research institutions should develop clear policies on data sharing and preservation and provide training and support for researchers to manage data effectively; and for journals to establish clear policies on data sharing and processes, with datasets underlying published papers readily accessible, and with appropriate data citation and acknowledgement[5].

The European RECODE project, investigating values, motivations and barriers towards open access to data in five case studies, found as motivations for researchers to share their data openly: easier access to data for comparison, error testing and to avoid duplication; faster advancement of science; more reliable research results; combined work; encouraging industrial uptake of data for commercialisation, and cumulative knowledge[6]. Overall, the project reported that incentives for providing open access to data were quite weak, and found

---

[5] Expert Advisory Group on Data Access, 2014. Establishing incentives and changing cultures to support data access, http://wellcome.ac.uk/stellent/groups/corporatesite/@mshpeda/documents/webdocument/wtp056495.pdf
[6] RECODE, 2013. Policy Recommendations for Open access to Research Data in Europe - Stakeholder values and ecosystems. http://recodeproject.eu/wp-content/uploads/2013/10/
RECODE_D1-Stakeholder-values-andecosystems_Sept2013.pdf;

a lack of incentives for researchers to participate in data review processes[7]. Boosting data citation metrics and impact factors that reflect data reuse, and the weight that data sharing and publishing may carry in career progression, (similar to the importance of paper citation indices and the impact factors of journal articles), are often flagged up as potential incentives for increased data sharing by researchers[8].

Overall, important factors to consider that influence data sharing culture and RDM practices are:

- data policy at EU, national and institutional level that sets expectations or requirements for research data management and sharing, with funding for data management and infrastructure/support development

- the availability of data infrastructure

- the availability of support services and tools

- the expectations of journal publishers, learned societies and the research community about transparency and publishing of data

- (international) collaborative research as a driver for data sharing

- career progression in academia as a motive for data sharing.

Surveys by the SERSCIDA project in Bosnia-Herzegovina, Croatia and Serbia show that in the absence of data infrastructure and support services, whilst researchers may be very willing to share their research data with the wider scientific community or publicly, in practice data are mostly shared with colleagues and peers within the research group/institution, or not shared at all. All researchers tend to keep their data for the long-term, but mainly stored on local computers and not well documented. In general, researchers express their readiness to archive their data in a specialized data archive for the social sciences[9].

1.2   Methodology

The methodology was based on a multi-method approach that combine **desk research** and various methods of **field research**.

Existing resources were utilised both for **desk research** for gathering information and for **conceptual introduction of chapters and terms definitions**. A basic literature review was prepared for each introductory section of a content area chapter that supported collaborators in finding and pulling together the relevant information to describe situation in a particular country. Preparatory activities for desktop research identify similar studies, make an overview

---

[7] RECODE, 2014. Policy Recommendations for Open access to research Data in Europe - Institutional barriers and good practice solutions. http://recodeproject.eu/wp-content/uploads/2014/09/RECODE-D4.1-Institutional-barriers-FINAL.pdf

[8] Costas, R., Meijer, I., Zahedi, Z. and Wouters, P., 2013. The Value of Research Data - Metrics for datasets from a cultural and technical point of view. A Knowledge Exchange Report. http://knowledge-exchange.info/datametrics; Force11, 2013. Joint Declaration of Data Citation Principles. http://force11.org/datacitation

[9] See SERSCIDA project deliverables: Maping reports for Bosnia and Herzegovina. Croatia and Serbia. http://www.serscida.eu/en/deliverables

and assess the relevance for Task 3.2 information collection, analysing methods used, and note about geographical scope and time reference period (see Appendix 1).

The **field research** was based on:

1. **Self-assessment template**: A structured self-assessment template in the form of a survey was primarily aimed at DAS representatives as informants. SaW project task 3.2 partners responsible for collection of data in a country checked the information and performed the follow-up interviews by telephone or email, aimed at validating data. Where no DAS existed, in addition to utilising existing sources of information via desk research SaW task 3.2 partners responsible for a country conduct selected interviews, and based on that filled in the relevant section of the self-assessment template.

2. **Semi-structured interviews**: Selected interviews were conducted, aimed in particular at collecting information addressing the wider composition of stakeholders for assessing the country's scientific, cultural, policy, legal and funding setting. A representative from one of the key stakeholders groups (respected researcher as representative of the scientific community, OpenAire national representatives and EU commission National Point of Reference for Open Access, and policy makers) was selected as a primary informant for interviewing and as contact person for a country, when seeking additional information.

Interviews and monitoring of information collected were carried out with SaW task 3.2 partners responsible for particular countries. Guidelines and communication protocols for interviewers were followed[10]. The method used in a particular country can be seen from pertaining report in the SaW D3.2 Country report.

A comprehensive data collection Web form instrument was prepared, where information about countries were centralised in a systematic and structured way. Data from all three data collection methods, as well as references to the relevant resources, were entered in the web form[11].

The whole set of ERA Countries were selected for collecting information. For eight countries out of forty-four, no information or only basic information was obtained[12]. All the rest provided information and collaborated in further activities, like revising and approving the final country reports. The outcome of the data collection activity that last from November till December 2016 is a data set that the following analysis are based on.

Since data come from various sources (some of which are more representative and objective than others), comparison of the individual scores between indicators, as well as comparison of scores between countries should be interpreted with caution.

---

[10] See **Appendix 2**, D3.2 Country report on development potentials: Guidelines and communication protocol for interviewers. http://cessdasaw.eu/content/uploads/2017/07/D3.2_CESSDA_SaW_v1.3.pdf

[11] See **Appendix 3**: Questionnaire text extract from the Web form, in the SaW D3.2 Country report on development potentials 1

[12] See SaW D3.2 Country report on development potentials 1 for details about countries covered.

2.   Results of self-assessment of the active data archive service

This section presents the results of the self-assessment conducted by the active data archive service (DAS), following the specification of the CESSDA SaW Capability Development Model. The CESSDA-CDM is defined as 'the basis upon which an assessment of Social Science service provision is made, and can aid in the improvement of the capabilities of existing and future CESSDA service providers'.[13] The CDM model was prepared explicitly with the purpose in mind to serve as a tool that can be used in other tasks of the CESSDA SaW project:

> *One of the objectives of the CESSDA SaW project is to deliver a state of play evaluation of social science data archives and services in ERA countries, identifying gaps and bottlenecks in existing services, and produce national development plans to close the gaps and overcome present barriers. The CESSDA SaW Capability Development Model is generated for this evaluation. The model specifies a set of objectives for establishing and operating CESSDA services for the social sciences and provides a common framework for evaluation of compliance.[14]*

Thus, the purpose of present application of the CESSDA CDM model is to:

- Assess the current status in each country and **identify gaps with regard to minimum CESSDA service provider requirements** to make suggestions for service improvements;
- Provide the base for comparisons across countries, **those more mature serving as an example for others**, also for external benchmarking (i.e. task 5.1 – CESSDA - ICPSR comparison)
- Use results for promotion of the CESSDA membership and SP's sustainability.

For the purpose of SaW task 3.2 self-assessment, we operationalise the model and apply it as part of the self-assessment instrument. By operationalization of CESSDA-CDM, we mean a selection of a limited set of Activities contained in a model, as the whole model would require extensive and time-consuming exercise on the side of the DAS. We thus test the model applicability with a limited set of Activities. The evaluation of the results may in the future lead to a more comprehensive application of the model.

The self-assessment exercise was limited to CESSDA service provider representatives, or representatives of organisations expected to be in this role in the future. The contact person for DAS coordinates which person in the organisation was entitled for providing information on which section. The person responsible for the country report explained and monitored the procedure of collecting the information, and if needed, checked the accuracy of information from external sources (e.g. Desk research), or asked for clarifications and follow-up questions. Self-assessment was the main method in this section. Twenty-three countries participated in

---

[13] See https://cessda.net/CESSDA-Services/Projects/Current-projects/CESSDA-SaW/Work-Packages/WP3/CESSDA-CDM.

[14] https://cessda.net/eng/CESSDA-Services/Projects/Current-projects/CESSDA-SaW/Work-Packages/WP3/CESSDA-CDM/Introduction/Background-of-CESSDA

this part of self-assessment, those that the working DAS could be identified and participation from the organisation obtained.

2.1   Organisational profile of the DAS

One needs to keep the variety of DAS in mind when interpreting the results of a self-assessment, both within current CESSDA member countries service providers and beyond. Country reports[15] contain more details of the complexity and peculiarities of each service provider organisational and technical setting that may explain the current level of activities. Often, in the national report, there are also current and future activities and plans for improvement mentioned, which were given as explanation for scoring current activities somehow lower.

An overview of organisational profiles begins with the definition of user community that the DAS uses in describing its services. Source of information were quotes, provided by informants that were representing the DAS from a given country. User community definition varies, with research and education community common to all. Additionally, applied research in the public sector, consulting offices, journalists, and best efforts to reach all users was mentioned. Some mentioned limited access to the data for some types of users, or expectation of having basic understanding of research methods and analysis techniques. Some DAS offer specialised services to a wider range of primary users from academic disciplines such as health, medicine, and humanities.

DAS representatives were asked to explicate the mission statement or to refer to an existing publicly available document. From the overview of those answers we conclude that as a rule they included the basic functions pre-ingest, ingest, archival storage administration and management, and access of data in the mission statement. However, in particular cases and with the wider role of DAS this is extended. It can include conducting research, promotion and training about open access; RDM; persistent identification and citing of research data nationally and internationally; sustainability arrangements of access to digital information; facilitating self-archiving; providing devices to access secure data; promotion of secondary analysis; support of policy relevant research; development of data preservation standards and best practice; acting as a competence centre in the personal data protection; and providing legal advice concerning research data. Some DAS are providing, or are part of organisations that provide, support for large-scale survey programmes as well.

Regarding the funding scheme, most DAS services are exclusively publicly funded, however the share of continuous long-term and short-term project funding varies. In several cases, in addition to the government research funding, there is a scheme of institutional funding supported by larger or smaller shares of University and National academy funding.

The DAS in countries vary in size, where some are having from less than one staff member (sometimes even a person taking a role voluntary with no funding at all) to 2 being the largest category, followed by other size ranges evenly distributed among countries, mainly reflecting the range of services, data types that are dealt with, and sizes and varieties of the community

---

[15] Presented separately, see http://cessdasaw.eu/content/uploads/2017/07/D3.2_CESSDA_SaW_v1.3.pdf.

of users. In the present report, we will keep the *Size of DAS* as a background variable that most obviously conditions the level of activities performed by the DAS.

Table 1: Country by Size of a DAS

Size of DAS (estimated FTE rounded)	Countries
1	Belgium, Estonia, Hungary, Italy, Lithuania, Poland, Portugal, Romania, Slovakia
5	Czech Republic, Greece, Ireland, Russia, Slovenia
10	Denmark, Finland, Sweden, Switzerland
20 and more	Germany, Netherlands, Norway, United Kingdom

2.2 Organisational infrastructure

The goal of this part is to self-assess active data archive service general features that enable long-term sustainable functioning and enable trust among users and wider public.

The main objectives are (following the CESSDA-CDM specification) to assess if organisations have adequate funding, valid budget planning, and sufficient numbers of appropriately qualified staff, managed through a clear system of governance, where roles and responsibilities are clearly defined, to effectively carry out the mission of the organisation; and an appropriate organisational structure that fits the objectives, tasks and processes of the organisation.

In the CESSDA SaW context, the aim is to establish, operate and strengthen (CESSDA) service providers for the social sciences.

The main factors that are addressed here in a shortened CESSDA-CDM are:

- Organisational sustainability requirements where specific objectives regarding Mission statement, Identification of types & formats of materials; Designated Community, Service Contracts and Liabilities; and Long-term viability of the repository.
- The availability of documentation, knowledge, trainings and capacity development in research data management (RDM) where specific objectives such as Staff professional development; Appropriate expertise; Compliance to legal and community norms; and the general objective of Effective Documentation.

The maturity level of each activity was assessed on a common 6-point scale (with few exceptions) representing different levels of maturity ranging from *1. Not defined, 2. Initial, 3. Repeated/partial, 4. Defined, 5. Managed to 6. Optimized*. The levels are defined in details for each specific activity, but there are some general/ generic properties that characterise each level[16].

Results of self-assessment are most positive for *Mission Statement* (Figure 1). Mission is obviously a constituting element of any organisation, and is present in all but 2 organisations filling in the self-assessment. High on maturity level are the legal and ethical aspects of organisational activities, which are on at least at the *Repeated* level present in up to 20

---

[16] https://cessda.net/eng/CESSDA-Services/Projects/Current-projects/CESSDA-SaW/Work-Packages/WP3/CESSDA-CDM/Introduction/Model-Components/Three-levels

organisations. *Legal and contractual regulation*, and related legal and ethical activities assessed, the *Data handling requirements* (in dealing with the confidential data), and the *Confidentiality and Disclosure risk* monitoring activities, are among those that achieve the highest self-assessment score. Usually the DAS in national or disciplinary environments represents the leading institution that promotes and advises sound practice in RDM within the academic community, concerning legal and ethical challenges[17]. Still organisations differ regarding the maturity level of each of the mentioned activity, with up to 5 that reach the optimised level in some of the activities.

Figure 1: Organisational Infrastructure
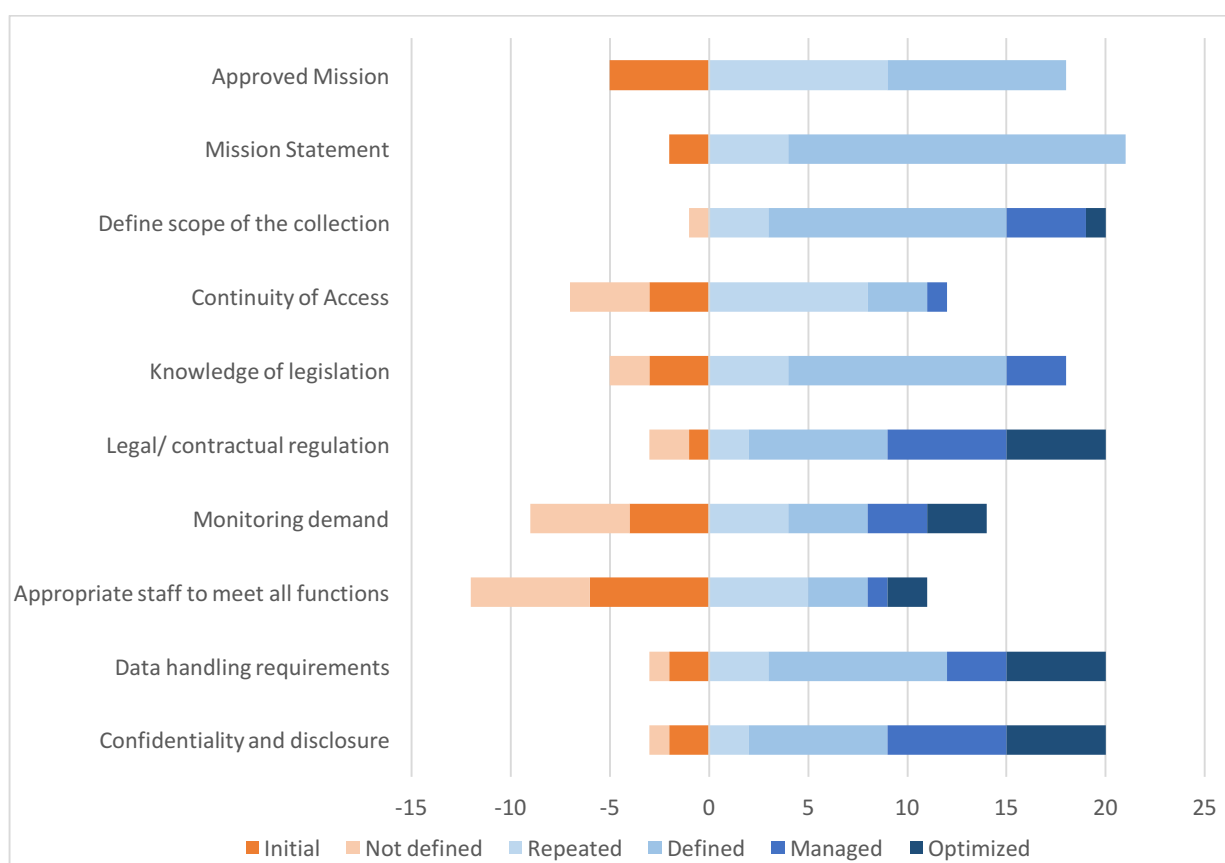
There are 4 organisations that have *Not defined* a contingency plan, 3 have only *Initial*, and 8 *Partial* but limited plans. The Netherlands, Norway and Sweden DAS have self-assessed as *Optimised*: *There are regular reviews and updates to the plans and agreements to reflect changes in the organisation's environment. E.g. due to funding issues or other "crisis".* Those can serve as an example for other that still discuss and negotiate the contingency plan nationally.

---

[17] See the establishment of Data protection official at NSD on http://www.nsd.uib.no/nsd/english/pvo.html.

Figure 2a, b: Mean organisational Infrastructure DAS self-assessment by group of countries regarding a) Funders' policy*, b) CESSDA membership**
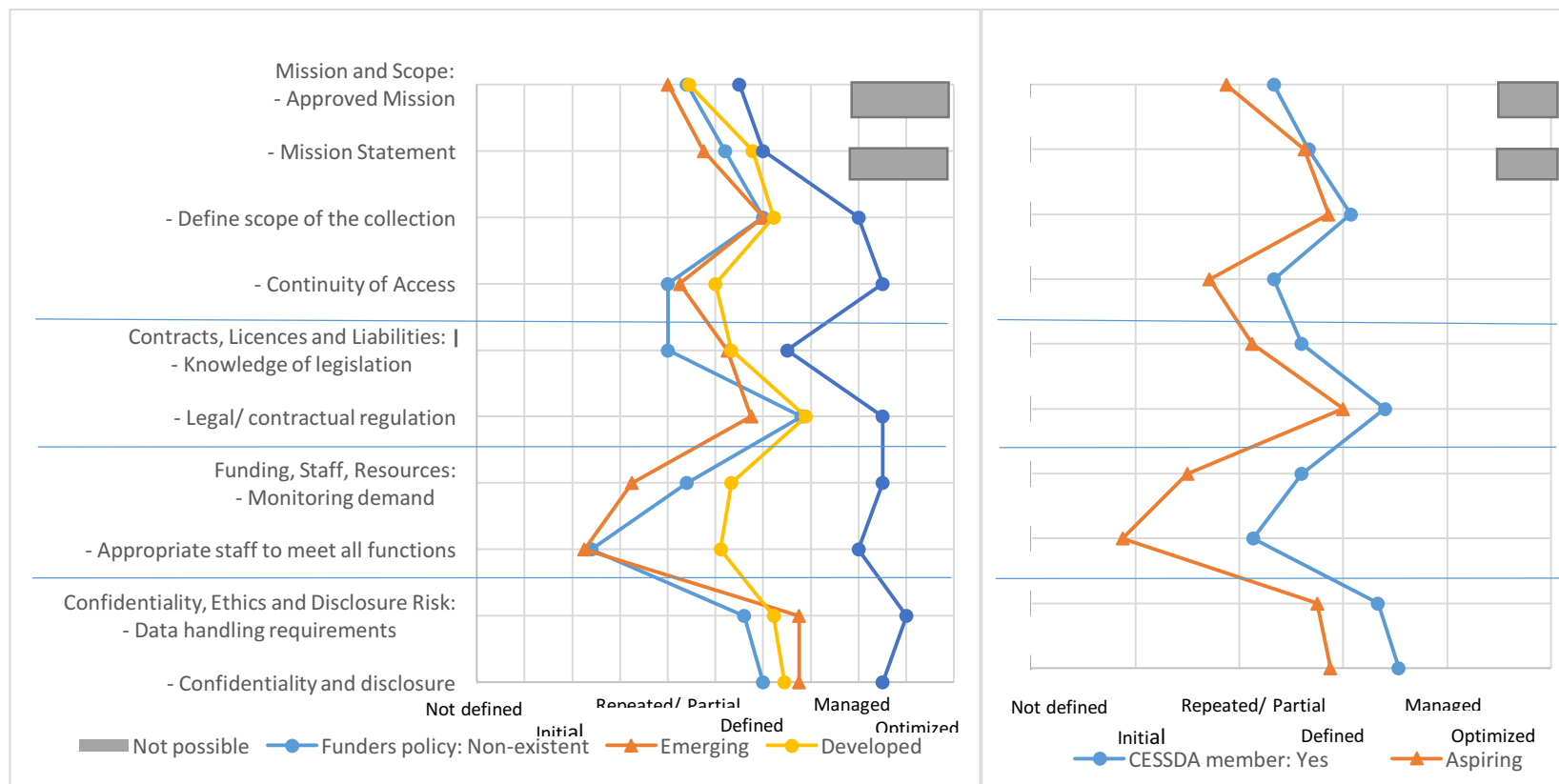
* See Chapter 3.2 Funders RDM policy and support setting for explanation of categorisation.
** CESSDA membership (status at the time of inquiry, January 2017). Members (Belgium; Czech Republic; Denmark; Finland; France; Germany; Greece; Hungary; Lithuania; Netherlands; Norway; Slovenia; Sweden; Switzerland; United Kingdom); Aspiring (Estonia; Ireland; Israel; Italy; Poland; Portugal; Romania; Russia; Slovakia).

Figure 3: Mean organisational Infrastructure DAS self-assessment by size of organisation (rounded)
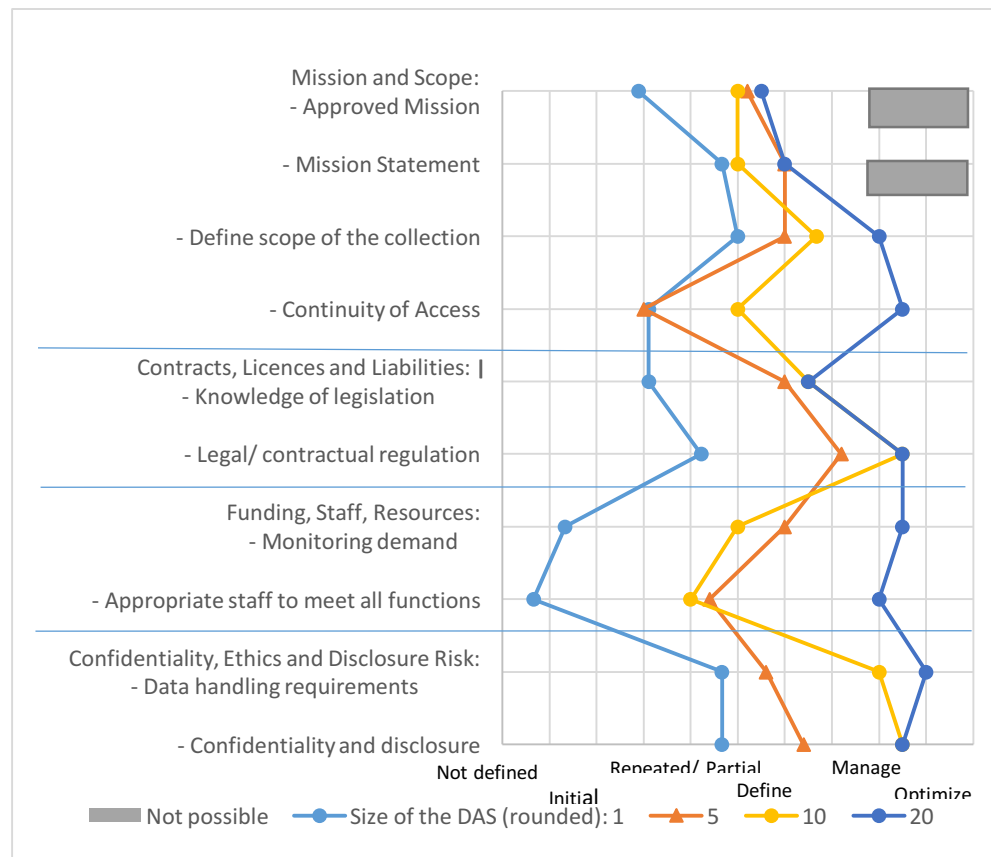
The graphs in Figure 2a, b and *Figure 3* show that the gap is smallest in the first three of Mission and scope criteria, regardless on which of the background characteristic we compare. Any DAS has as a constitution including the mission statement that communicates the range of services to the users' community.

It is sustainability (*Continuity of access*) that shows substantial gap in the first part of organisational monitoring. The gap is largest between the biggest organisations and all others. The middle-sized small organisations (with around 5 staff members) also self-assessed this as their weakest point.

The presence of legally and ethically sound practice also among smaller organisations at least on the *Repeated level* confirms that respect of data subjects is a first principle that DAS are following, thus keeping the high level of trust among the public. Nevertheless, the gap in following the legal and ethical requirements on the highest professional level is still large between the understaffed small organisations and all the rest. Small organisations cannot afford professionalization in a form of narrower specialisation. Certainly at least more than a minimal number of staff is a precondition for accumulating and keeping track of the knowledge in various fields of the data services.

In addition, it is in the criteria of the next activity, the Funding, Staff and Resources, that both the smallest and the biggest organisations by size have clearly ordered themselves, first at the bottom of not defined, second at the Managed to Optimised level. This is also the aspect of self-assessment that the most visible difference can be observed in relation to CESSDA membership and funders' policy development. This reinforces the conclusion that it is appropriate number of staff that can make a difference in organisational viability, as a basis for sustainable and quality services.

17

For sustainability, stable and continuous funding is necessary for the perspective of long-term preservation and keeping the accumulated expertise in human resources. Unfortunately, there are a couple of services that at the time of inquiry did not receive any funding or received only very small amounts, and, as it was explained in the reports, this situation could be extended to a longer period. Chronic lack of funding or unstable funding can cause an initial investment into a DAS service to be lost, and then after some time it requires again the initial funding and extended periods to re-establishing the DAS services anew.

2.3   Digital object management and technical infrastructure

From the CESSDA-CDM, the following definition can be obtained:

> ***Digital Object Management (DOM)*** *consists of the set of processes (selection, acquisition, ingest, management, preservation) required to maintain and provide access to digital information in an authentic form, for as long as required and across changing technical environments. Digital Object Management is closely related to the term "digital data curation". Data curation is the selection, preservation, maintenance, and archiving of digital assets and it establishes, maintains and adds value to data for present and future use. The aim of DOM and digital curation is to mitigate digital obsolescence, keeping the information accessible to users indefinitely.*

The technical infrastructure of an organisation, or an infrastructure, provides the technical underpinnings for it to fulfil its functions and the provision of services to its designated communities. As such, this section of the model can involve a wide range of support technologies and services, but could also be limited to the core technology used to deliver services to the communities.

FAIR principles are at present actualising some of the key functionalities of the DAS in the data access ecosystem. DAS consist of peoples and technology. While organisational activities and digital object management activities depends mainly on people, the technical infrastructure can make systems talk with one another and in providing interface from digital objects to users. Thus, Findable and Accessible are both related to the preservation and descriptive metadata content and formats, PID services, citation formats etc. Interoperable again depends on the data and metadata standards that are followed, and Reusable is related to the legal and ethical clearance of conditions of access, in addition to the technological aspect of keeping data transparent for further reuse.

The data archives and services in the social sciences have a long tradition. The assessment of some of the activities that support basic functions are on the high level in all of the DAS that were identified and available for self-assessment (see Figure 4). Metadata and documentation requirements, clear communication about requirements for deposit and available standardised licences and agreements are in place, and data and metadata standards are respected for most common types of digital objects.

Less common are concepts related to execution of digital preservation activities in the narrow sense: preservation strategies, succession and technical plans and risk assessments. Also partly missing are some of the features of linked open scientific identities environment, such as AAI, persistent identifiers, and citations. Again, we can probably observe the difference in capabilities that requires investment into more advanced technical development and larger specialisation of staff to carry on separate functions and roles in the organisation.

Figure 4: Digital Object Management and Technical Infrastructure
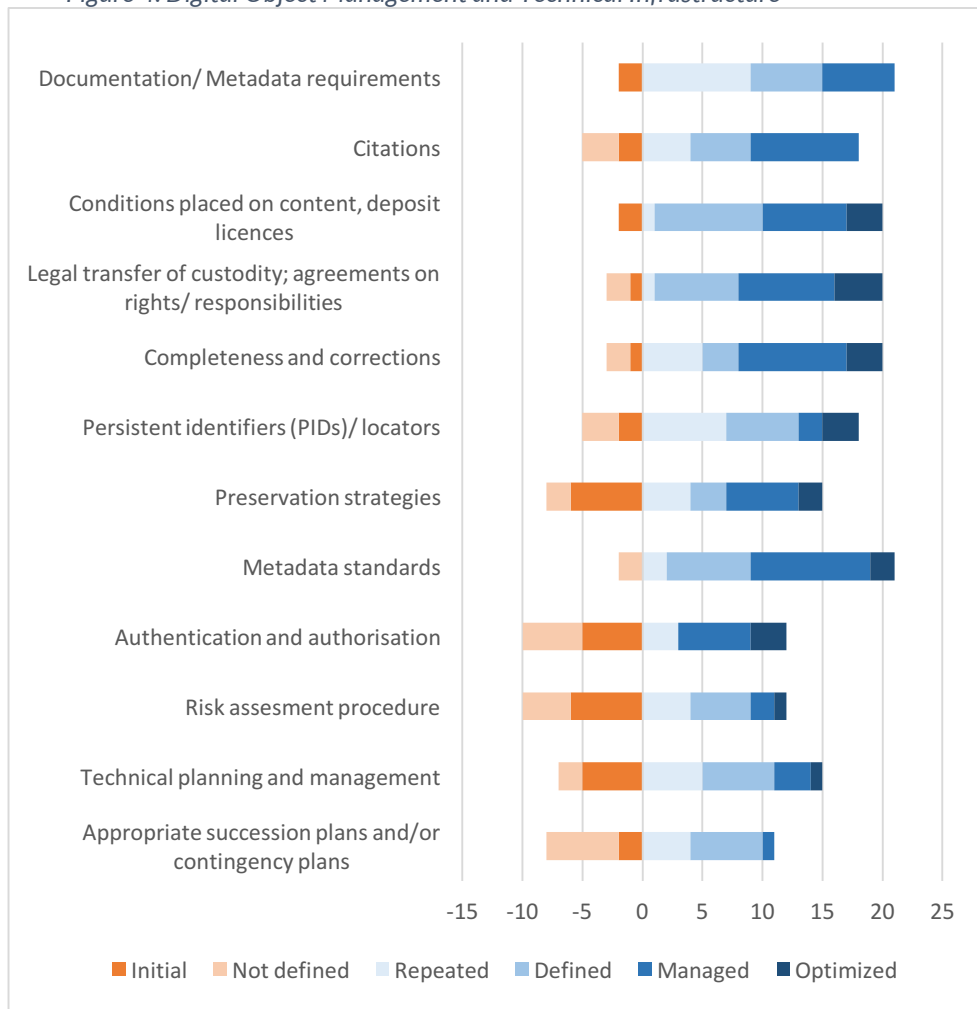

Figure 5 (below) shows little differences in DAS digital object processes and technical infrastructure approaches between countries with the developed and those with emerging and non-existent funders' data sharing policies. Largest is a difference in the technical safeguards guaranteed for long-term preservation, articulated through the 'Appropriate succession plans and/or contingency plans'. It seems that the wider ecosystem readiness to support the long-term preservation of high-quality research data may encourage organisational response to introduce appropriate measures. The succession plan itself usually requires that some of the stakeholders outside the DAS organisation, funder or another organisation of national importance such as an academy, national library or national archives, recognise the importance of the sustainable arrangements of digital preservation.

Figure 5: Digital Object Management and Technical Infrastructure by Funder RDM policy development
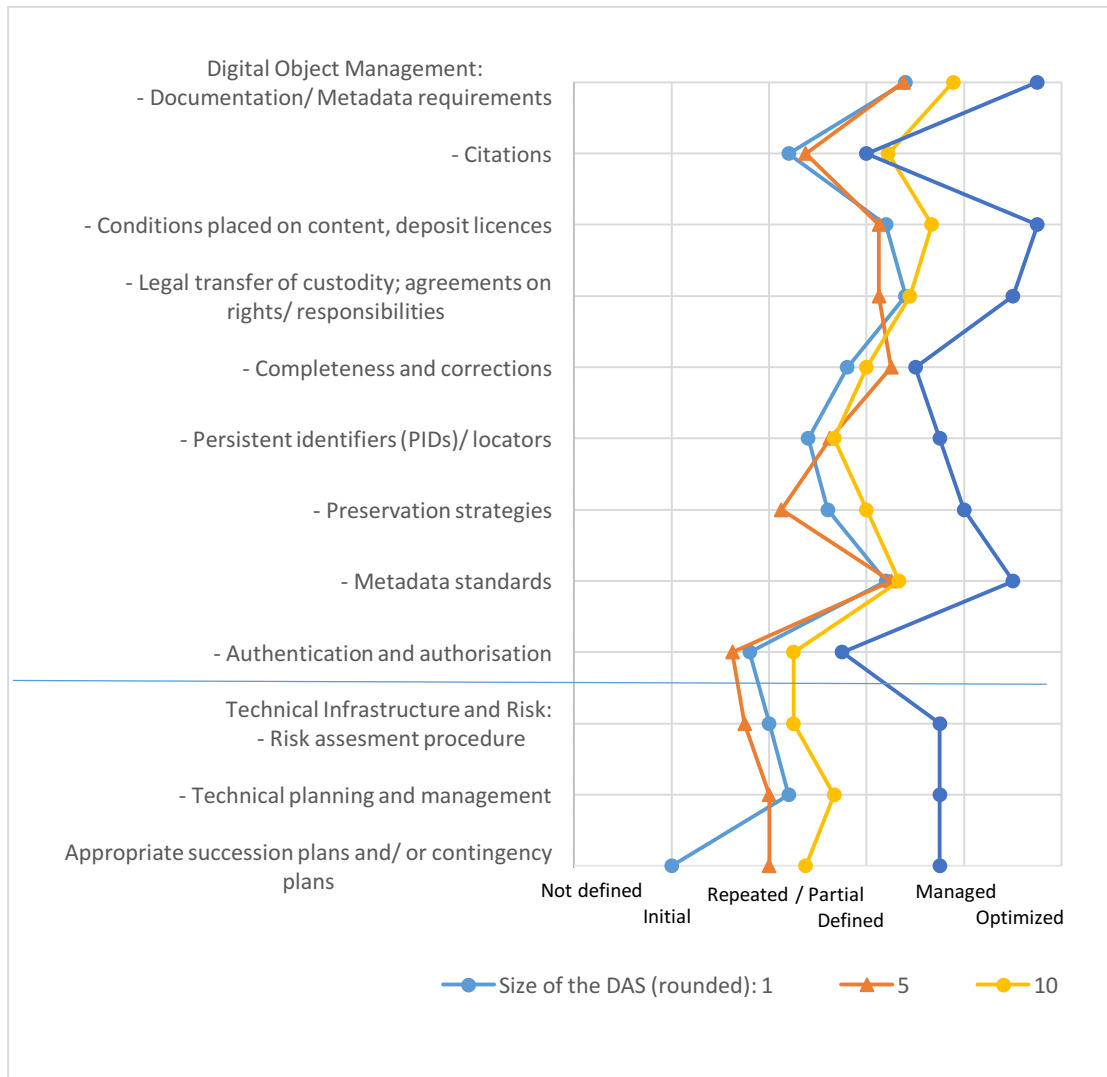
As this showed to be a weak point for most included organisations perhaps the leading examples can be followed and customised in other countries. From comments provided in the web form, given supporting the self-assessment, we learn that more developed arrangements indeed include agreements with the organisations with the national reputation: National libraries, Academies of science, and funding organisations. In some cases, special task forces are foreseen, to manage during the reserved period of time technical and organisational aspect of a succession, where a succession plan includes the key representatives of (former) organisations and external stakeholders into the membership of the task force.

Citations are the second aspect of an organisational setting that shows the largest variation, according to the existence of funders' policy. Providing citation services, being one of the most powerful enablers and incentives for the research data producers to share data through the established DAS, clearly is in congruence with the funders' policy, which in the developed form would entail the stimulating evaluation of the researchers and groups, based mainly on the data citation records.

Figure 6: Digital Object Management and Technical Infrastructure by country CESSDA Membership
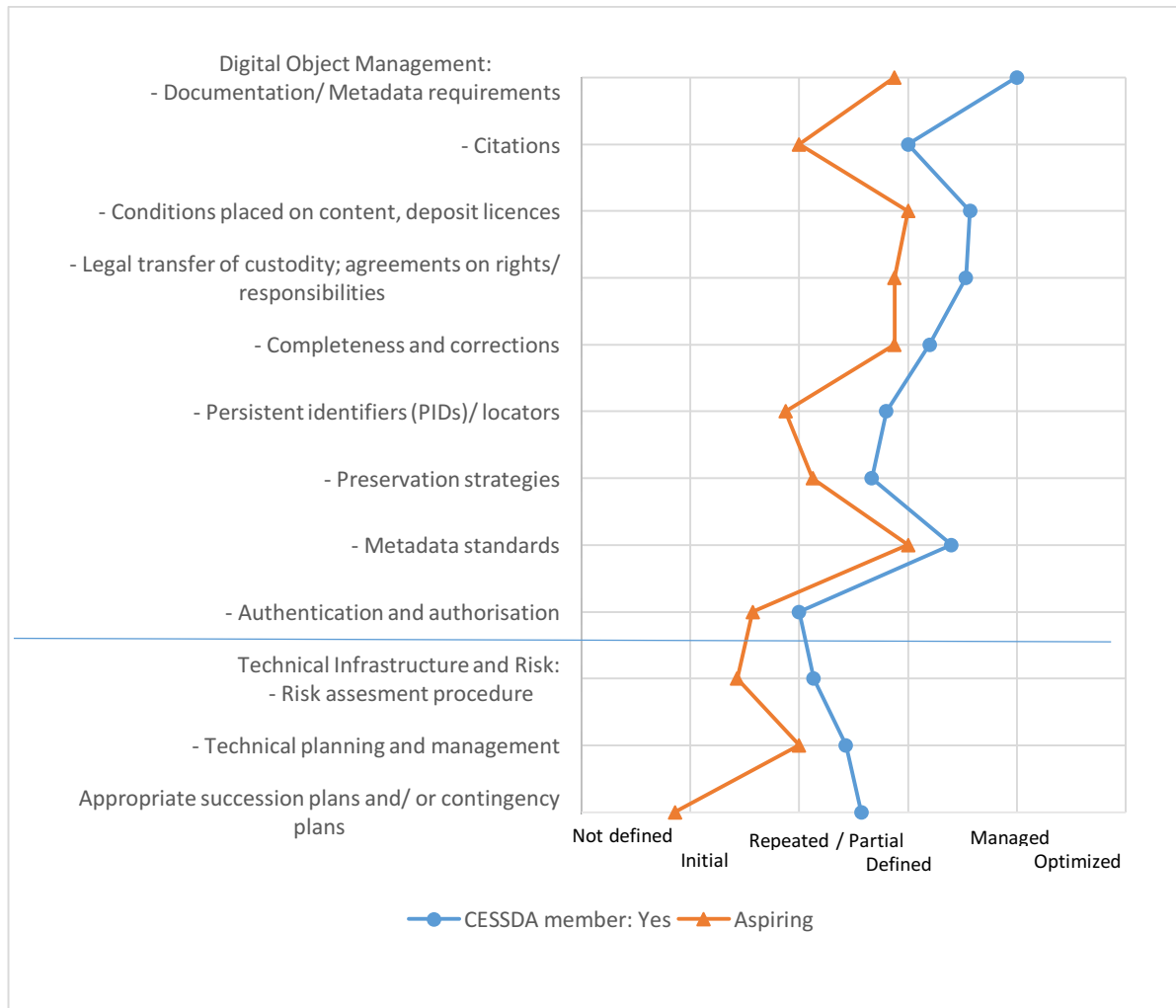
Again, citations management, as one of the aspects of the modern digital research data ecosystem (see FORCE11, DataCite), appears also in discerning the CESSDA members and aspiring DAS. CESSDA ERIC may use this observation in setting the priorities among areas, taking into account where the gaps are largest among more developed and less developed services, where the demands from external stakeholders and the designated community are most urgent, and where the common standardised solutions can be proposed and designed in such a way, that can be applied among all DAS. The current debate around the adoption of the FAIR principles[18] for the mature infrastructure services, as CESSDA clearly emphasises, is that to be Findable and Interoperable the data resources and information services need to be based on standard citations, including the comprehensive metadata and persistent identifiers services. The later also lags

---

[18] See https://www.force11.org/group/fairgroup/fairprinciples.

behind among the DAS from aspiring countries. The CESSDA Persistent identifier policy, which has been proposed, may offer such a solution for those DAS that are starting to implement the PID systems. A similar area that is subject of further integration, coordinated within CESSDA, is the authentication and authorisation infrastructure that helps to 'provide access to research data in an effective and secure way'[19].

The gap is large also in the 'Documentation/metadata requirements'[20]. This activity requires that 'The organisation clearly specifies the information (documentation, metadata, provenance) that needs to be associated with the data that is to be deposited' which on the average is close to *Defined* level: 'A written formal specification of required information is explicitly defined (e.g. in a collection policy); requirements are compliant with metadata standards that are used and can be understood by the Designated Community (e.g. DDI); metadata requirements are accessible and communicated to users/depositors.' As a rule, DAS in the field of social sciences are very much aligned to the DDI standard and are capable of negotiating the content of metadata and documentation with the depositors. It is the level of advanced digital object management that requires documentation and metadata requirements to be aligned with policies and on continuous improvement based 'on technology watch, monitoring of, and communication with Designated Community and other relevant stakeholders'.

The remaining areas, where we discovered the relative weakness in aspiring countries compared to existing CESSDA members are Preservation strategies, Risk assessment procedures and in particular appropriate succession and /or contingency plans[21]. As for 'Continuity of access' from the organisational type of activities, among technical activities formalised and documented digital preservation activities and practices present a challenge for the organisations that may not afford specialised experts in the field.

This conclusion is further confirmed below in Figure 7, where comparison among the organisations of different size shows that technological digital preservation infrastructure can only remain partially complete in the circumstances of the lack of personal and financial resources. Priority in such situations is usually oriented toward more visible immediate ingest and access services offered to satisfy basic users' needs.

Technical infrastructure and Risk is clearly the highest – on the average close to *Managed* level - in the largest organisations. These can afford the technical development, having more technically

---

[19] See **RA2.3.2.3**, https://cessda.net/eng/CESSDA-Services/Projects/Current-projects/CESSDA-SaW/Work-Packages/WP3/CESSDA-CDM/Part-2-CRA2-Digital-Object-Management/CPA2.3-Access

[20] See **RA2.1.1.2: Documentation/ Metadata requirements** in the https://cessda.net/CESSDA-Services/Projects/Current-projects/CESSDA-SaW/Work-Packages/WP3/CESSDA-CDM/Part-2-CRA2-Digital-Object-Management/CPA2.1-Data-Acquisition-and-Ingest

[21] See *RA2.2.3.3: Preservation strategies,* *https://cessda.net/eng/CESSDA-Services/Projects/Current-projects/CESSDA-SaW/Work-Packages/WP3/CESSDA-CDM/Part-2-CRA2-Digital-Object-Management/CPA2.2-Data-Preservation-storage-curation-and-planning;* **CPA3.1 - Risk Assessment**, https://cessda.net/eng/CESSDA-Services/Projects/Current-projects/CESSDA-SaW/Work-Packages/WP3/CESSDA-CDM/Part-3-CRA3-Technical-Infrastructure/CPA3.1-Risk-Assessment; *SO3.5.1: Business continuity plan, https://cessda.net/eng/CESSDA-Services/Projects/Current-projects/CESSDA-SaW/Work-Packages/WP3/CESSDA-CDM/Part-3-CRA3-Technical-Infrastructure/CPA3.5-Technical-Resilience-Disaster-Planning*

specialised staff at their disposal. Larger organisations also tend to have more sustainable digital preservation strategy arrangements, including the specified succession and/or contingency plans.

Largest organisations in the area of *Digital object management* exceed a bit more from all others on the aspects of *Conditions placed on content*, and *(meta)data standards management*. Both aspects require extended elaboration of the data life-cycle stage interdependencies. The developed access conditions arrangement requires elaboration of the deposit agreement related to specific data types, and the technical and organisational aspects of granting access, following those conditions. The same can be said about the formats registries, which can be a challenge to follow and cover related to the extended spectrum of different data types, especially for small organisations.

Figure 7: Digital Object Management and Technical Infrastructure by DAS Size
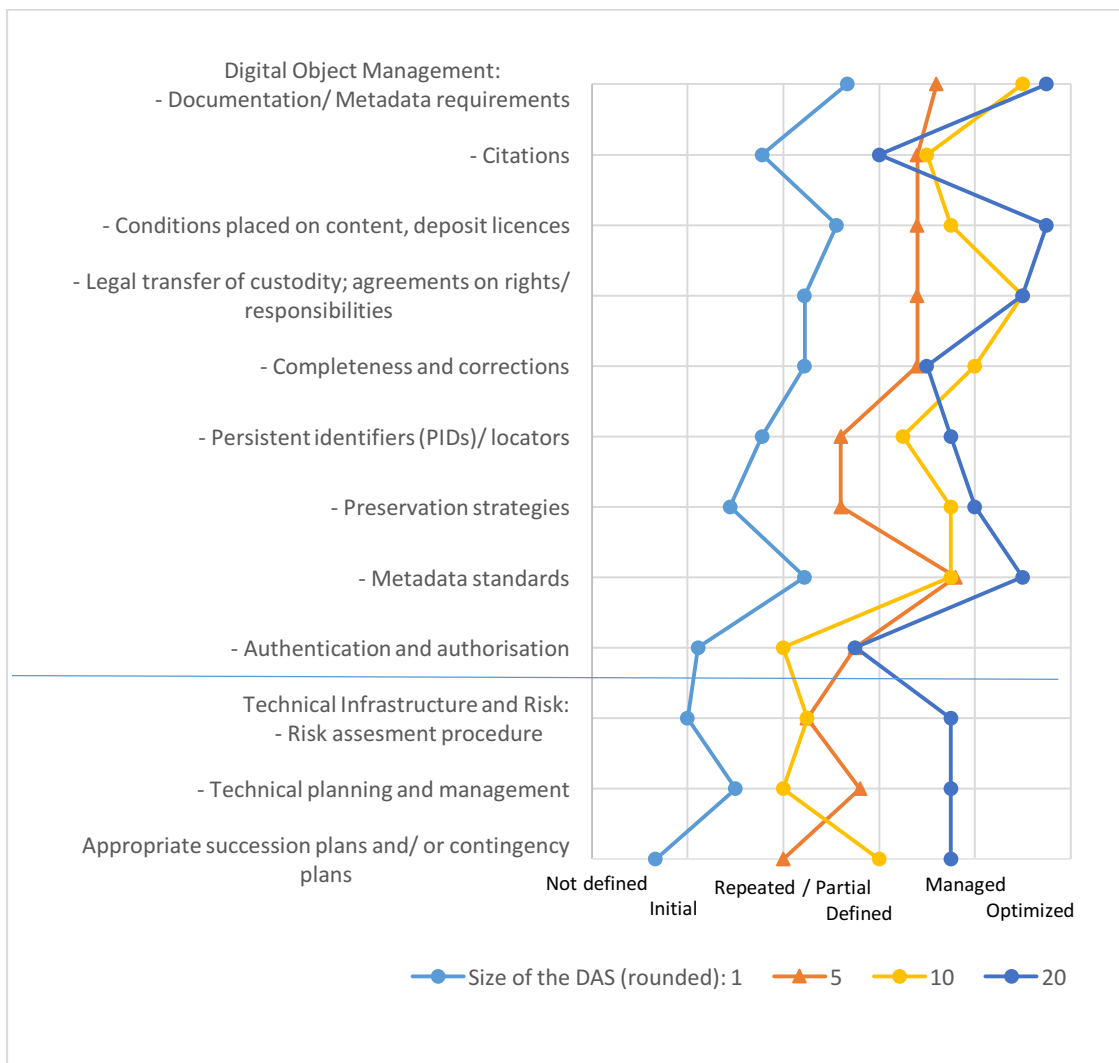
2.4 Conclusions about DAS Capability Development

Both DAS among current and candidate CESSDA ERIC members have many things in common. They use similar definitions of designate (users) community. Mission statements are as a rule public and explicit in defining OAIS functional entities.

Appropriate Funding, Staff and Resources are a condition that severely hampers more than half of the European national data service providers. In many places, the expertise established cannot be utilised yet to full potential due to the lack of basic organisational conditions. This also goes against the sustainability arrangements that pertain to the long-term value of data resources kept and made accessible through the data services.

Basic functions of the DAS are on a high level in all services that were identified and available for self-assessment, like metadata standards used, requirements communicated to users and depositors, and support offered. There is still slow adoption of some of the more advanced information technologies that require investment in more demanding development and larger specialisation of staff. Different organisational, digital management, and technical aspects that require the narrow specialised expertise and continuous monitoring of digital preservation environment, or substantial investments into technical support, all are candidates for being accepted as common CESSDA tasks that can run in collaboration. It is clear that larger organisations, which all reside in the countries that are members of CESSDA, are also in a most favourable position in acquiring professional knowledge and other resources. Both further development of various solutions and spreading established ones them among other members are the tasks that can be achieved in collaboration. Some areas, such as adoption of PID and AAI services, are still rather low in maturity in many of the groups of DAS. Thus, gains in making the adoption of the agreed CESSDA PID policy system and a coordinated mature AAI implementation may be greater when more countries are involved.

3. Broader data sharing ecosystem development results

Results of this section are aimed towards exploring how different enablers and incentives for sharing data characterize the broader context of DAS activities and development status. It included aspects of general development of social science sector, policy level support and scientific community acceptance that are reflected in data sharing culture.

3.1 Structural conditions

The aim of this block of indicators was to assess overall development of the social sciences and research data production as a condition that determine status of social science research in the country. Lack of social science development can be considered as barrier for development of other component of data sharing ecosystem, including the DAS organisation position in such a system. Having in mind the complexity of measuring all characteristics that determine overall development, we focused mostly on the issues of financial stability, research capacities and results achieved in the social sciences in a country. In measuring the development of research data production we took into consideration both volume and average quality of the data produced, and potential of data reuse.

Overall development of the social sciences is likely the most important determinant of the quantity and quality of the research data produced and potentials for their reuse. In recent decades big efforts have been made to measure the development of social sciences, mostly by the national institutes of statistics, UNESCO, OECD, and the European Commission, being the main examples of organizations starting to collect systematically data on the development of science and technology[22]. A well-developed social science sector enhances the potential of the researchers to produce data of higher quality, with regard to availability, usability, reliability, relevance and presentation quality[23].

3.1.1 Development of the social science sector

CESSDA membership is conditioned on existence of established service provider in a country. We see already that one of key determinants of comprehensive DAS is related to size and financial sustainability. Results of this section confirm the gap in SSH development determine the current CESSDA membership. Less developed countries in Europe cannot sustain the conditions of CESSDA membership without further external support.

With the first heading concept, we aimed for overall assessment of financial stability, research capacities and results achieved in the social sciences in a country (funding, human resources and infrastructure conditions; impact and prestige in society). The following set of objective official statistics indicators was selected to assess the development of social science disciplines, consisting of a few figures about funding capacities and human resources:

---

[22] See http://unesdoc.unesco.org/images/0024/002458/245825e.pdf.
[23] Cai, L. and Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, 14, p.2. DOI: http://doi.org/10.5334/dsj-2015-002.

1.1.1 Gross Domestic Expenditure on Research and Development **(GERD) in SSH as %** Gross Domestic Product (**GDP**)   This indicator shows general intensity of investment in SSH

1.1.2. GERD in SSH as % of GERD   This indicator shows the relative intensity of investment in SSH (relative to other scientific disciplines)

1.1.3. Number of researchers in SSH per capita (in 100,000 pop)   This indicator shows human resource potential in SSH

1.1.4. GERD in SSH per researcher in SSH (in EUR)   This indicator shows investment in human resource in SSH

Development of the social sciences is also made visible by the supply of support services that are at the disposal for users. **1.1.6. Existence of support services** was measured by the informant's interview or self-assessment in the web form. **Final score consists of a count of the following** options checked: software licenses, datasets, and full-text databases.)

Table 2: Development indicators by CESSDA Membership status[24]

CESSDA member	GERD in SSH as % GDP	GERD in SSH as % of GERD	Number of researchers in SSH per capita (in 100,000 pop)	GERD in SSH per researcher in SSH (in EUR)	Existence of support services (options count)
Yes*	.18	10.7	131	41722	2.1
Aspiring**	.12	12.5	101	14777	2.2
No or initial***	.09	23.1	66	10494	1.6
Total	.13	14.2	101	22739	2.1

* Belgium; Czech Republic; Denmark; Finland; France; Germany; Greece; Hungary; Lithuania; Netherlands; Norway; Slovenia; Sweden; Switzerland; United Kingdom;

** Croatia; Estonia; Ireland; Israel; Italy; Latvia; Poland; Portugal; Romania; Russia; Serbia; Slovakia;

*** Albania; Bosnia and Herzegovina; Bulgaria; Cyprus; Kosovo; Macedonia (FYRM); Montenegro;

CESSDA members (status at the time of inquiry, January 2017) are distinct by the absolute amount of investment into SSH (GERD in SSH per researcher in SSH (in EUR)) more so than by any other indicator. The relative 'investment' also has a similar meaning, both determined by the size of GDP and corresponding share for science. Following this is the 'Number of researchers in SSH per capita (in 100,000 pop)', reflecting the size of the science community, which in turn could determine the demand for the infrastructure services. The only indicator that has an inverse relation with the

---

[24] See Table 1.4.1 in SaW D3.2 Country report on development potentials 1., which describes the countries status regarding CESSDA membership.

D3.6 – v. 1.4

CESSDA membership status, is the 'GERD in SSH as % of GERD', which probably follows the traditional pattern of overall science underdevelopment, where SSH can survive with relatively larger numbers of researchers with minimal costs. This obviously does not encourage the corresponding social science infrastructure development.

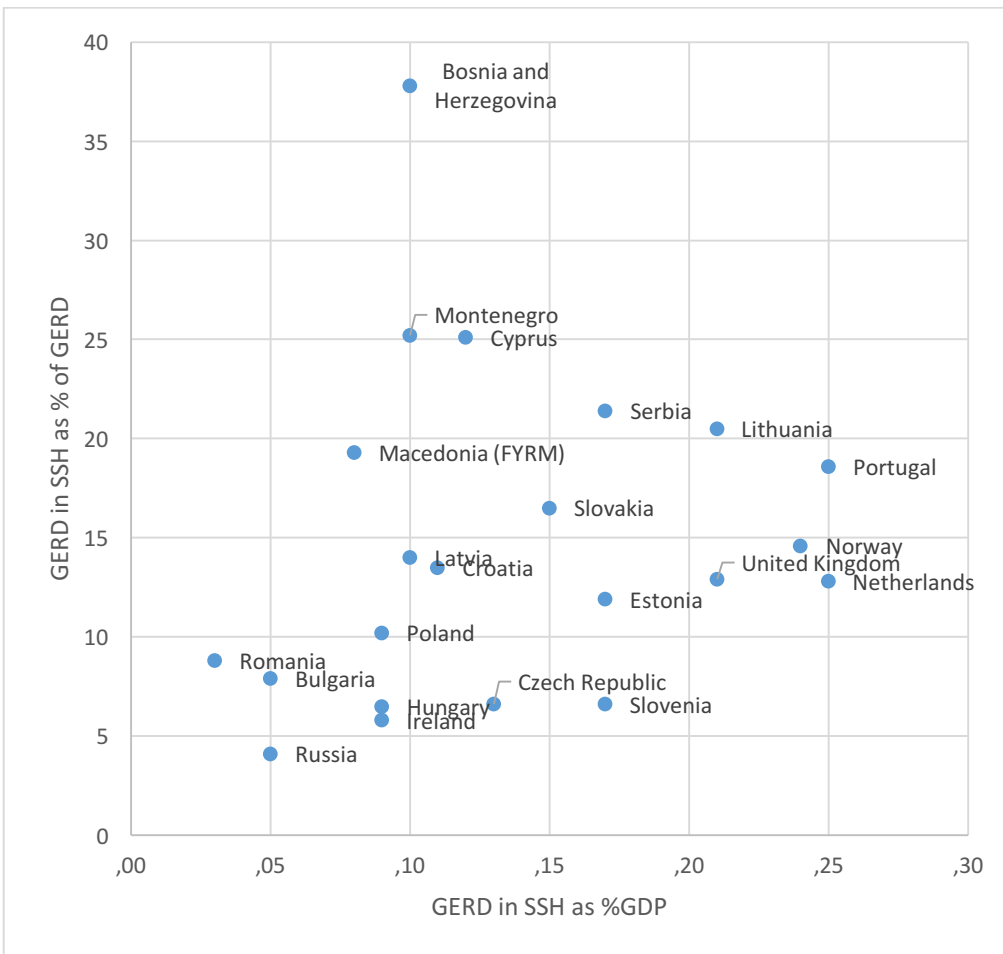*Figure 8: GERD in SSH as % of GERD by GERD in SSH as %GDP*

Figure 8 shows the position of countries regarding the relative and general intensity of investment in SSH. We can discern a group of steady state SSH developed countries (Lithuania, Portugal, Norway, UK and Netherlands) and the underdeveloped (Romania, Bulgaria and Russia in the extreme). The third group (BiH, Montenegro and Cyprus) is distinct with strong relative SSH and weak general position. Certainly it is the size of GDP that mainly determines the absolute development of SSH, where such distinct countries as Norway and Portugal would fall apart. Overall, lower GDP also influences lack of R&D investment in expensive infrastructure for research in sciences and technology. This may be beneficial for future CESSDA activities involvement in new countries, as investment for SSH infrastructures can be considered as disproportionally lower compared to other sciences, and thus perhaps easier to prioritise.

Each of the indicators' values was first divided into quantiles of thirds and assigned values from 0 to 2. Indicator 1.1.6 on support to SSH in the form of software licenses and access to databases gives three options to check (software licenses, datasets, and full-text databases). If all options

27

were checked, the score of the indicator is 2, if two were checked the score is 1, and otherwise it is 0.

The overall assessment of financial stability, research capacities and results achieved in the social sciences in a country (funding, human resources and infrastructure conditions; impact and prestige in society) sums the quantiles values, and distributes the final values again approximately into thirds. If values were missing for any of the indicators, the overall score was determined based on the values of remaining indicators. Thus, three distinct categories of **Development of social science sector** were derived:

> **0 - Low level:** *Funding of SSH and productivity of the researchers are in the lowest quantile; impact on designated community is small or non-existing*;

> **1 - Medium level:** *Funding of SSH and productivity of the researchers are in the mid quantile; impact on designated community is limited*;

> **2 - High level:** *Funding of SSH and productivity of the researchers are in the highest quantile; impact on designated community is strong.*

Clearly, the propensity of being a CESSDA member is somehow related to the overall conditions of the development level of the social science sector in the country, which at least to some extent reflects also the general economic development and wealth of the country.

Figure 9: Development of social science sector by CESSDA Membership status
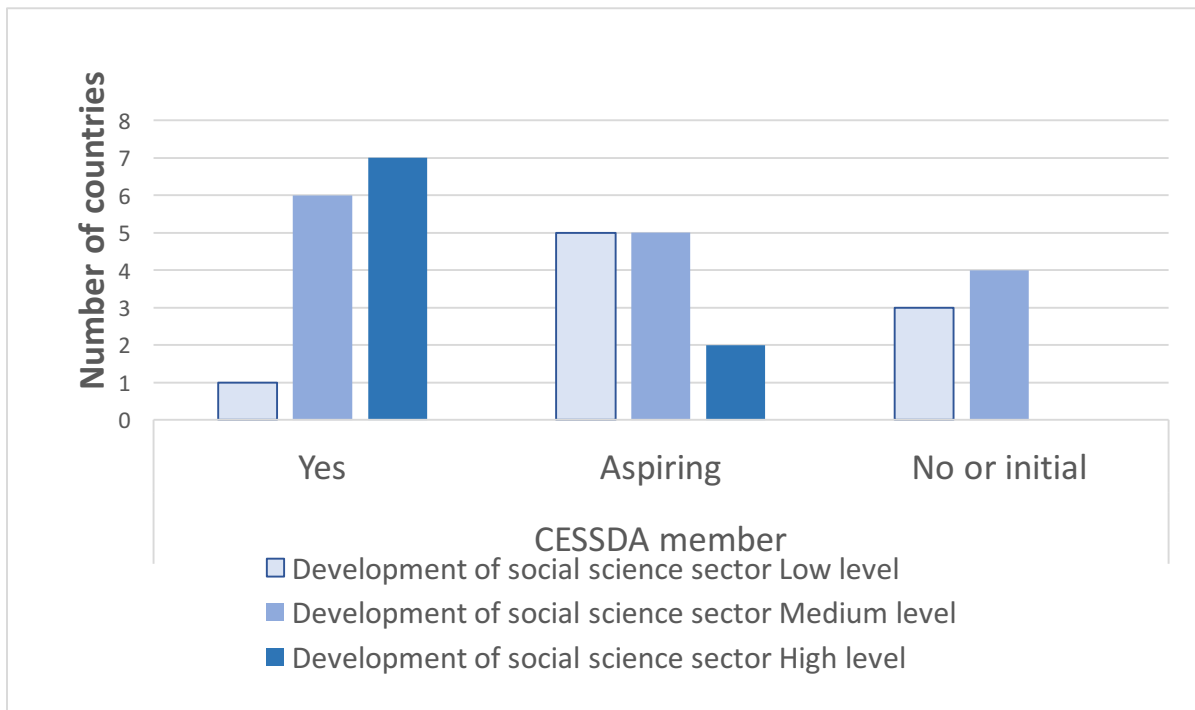
3.1.2 Development of research data production in SSH

The second dimension assessed to explore the potential for the DAS to have a prominent role in supporting quality science production in a country, is the amount and quality of the research data output. This is conceptually considered as one of the enablers of the data sharing culture to flourish. More particularly, we include the following indicators of development of research data production:

- International collaboration and national studies as the drivers of data production (see 1.2.1 and 1.2.2) and
- Volume and quality of data produced (see 1.2.3).

Table 3: Development of research data production in SSH by CESSDA Membership status

CESSDA member	Existence of international collaborative research or cross-national studies as a driver for data production* (Mean Count)	Existence of studies of national importance as a driver for data production** (Mean Yes=1) (Else=0)	General volume and frequency of data production in SSH in a country*** (Mean 'Frequently, institutions have well established tradition in data production'=1) (Else=0)
Yes	6.9	.93	.60
Aspiring	6.1	1.00	.33
No or initial	4.0	.43	.00
Total	6.1	.85	.38

* Desk research, official sources, literature review
** 'Are there any existing studies of national importance produced by SSH researchers in your country?'
*** 'In your experience how would you characterize the average production of research data by the SSH institutions in your country?'

Both existing and aspiring CESSDA members are on the average in the mid count (6 studies) or more of involvement in international comparative research data producing projects or programmes. Data production that follows from studies of national importance is also present in both categories. The only difference is in the last indices, the General volume and frequency of data production assessment, where the aspiring have below 50% (33%), and existing above 50% established as 'Frequent' a well-established data production.

High-quality and larger volume research data produced is expected to be one of the enablers for further development of the data-sharing ecosystem, including the internationalisation and integration of national data services under the European infrastructure unit. The countries in the initial stage of joining CESSDA are lacking also the propensity of international data production, with more than half of them lacking studies of national importance.

Summary categorisation of **Development of research data production in SSH** (*Prevalence of high quality research data with high potential for reuse*) into **Underdeveloped:** *Rare or no data producing*

research projects, dispersed and low quality existing data, absence of studies of national importance; **Developing:** *Some examples of research excellence, particular streams of research outstands, either qualitative or quantitative, some examples of international collaborative research*; and **Developed:** *Well established streams of research traditions, national and international, great variety of important types of research data*, is based on three above-mentioned indicators. The development status assignment follows the same logic as described previously: Higher values (above 6) on the Existence of international collaborative research or cross-national studies, existence of important national studies, and high volume of quality data production leads to Developed category[25].

Figure 10: Development of research data production in SSH by CESSDA Membership status
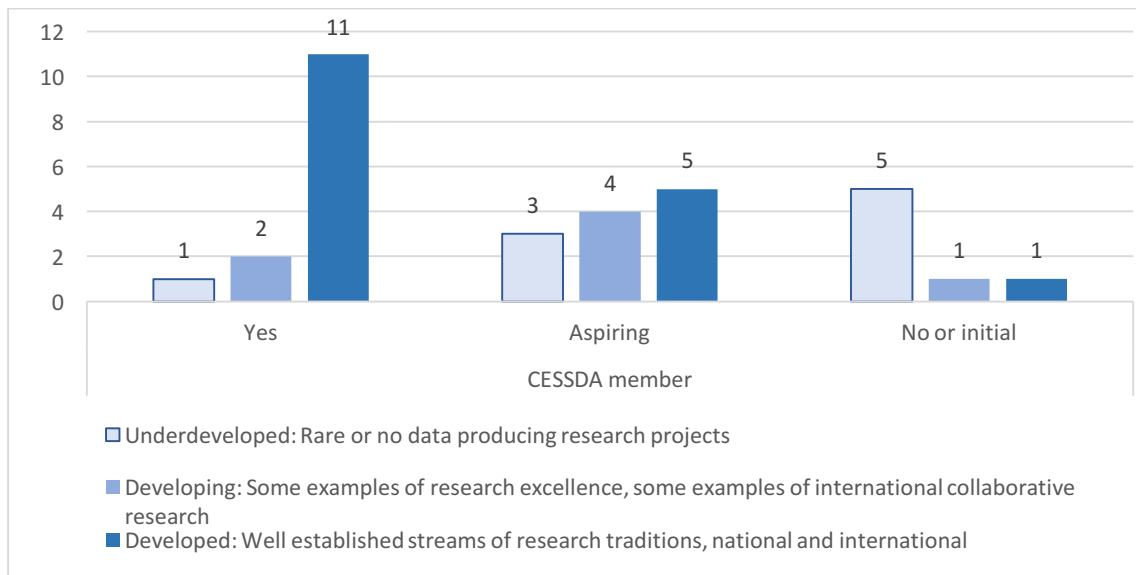
Again, the CESSDA members appear to have consistently higher levels of development also in the quality data production, and countries within initial membership status correspondingly lower.

---

[25] Details of heading concept categorisation are described in Appendix 4, SaW D3.2.
http://cessdasaw.eu/content/uploads/2017/07/D3.2_CESSDA_SaW_v1.3.pdf

3.2   Funders RDM policy and support setting

Research funders are the key stakeholder that can help to provide incentives and remove some of the barriers to data sharing. The aim of this section is to explore research funders' data sharing policies and underlying strategies that enable sustainable access to publicly financed social science research data. This can range from declared awareness about principles and soft recommendations to explicit requirements, the fulfilment of which is proactively supported and rewarded. The established policy environment of open access to research data provides sets of incentives, requirements and rewards that support practices of data sharing. In such an environment, data archive services can have a recognized and important support role.

As Science Europe Roadmap noted:

> *Quality-assured research data are key building blocks of the research process, and are the basis of economic and societal innovation. Research data often generate impact that goes well beyond their initial purpose. They are highly valuable in terms of supporting new research. They are also indispensable in verifying research findings, and in this way, the sharing of data contributes significantly to good scientific practice. The collection of research data can be a major part of a research project, and the wider sharing and reuse of this data can help to maximise the value of the original investment.*[26]

However, 'RDM funding is generally not (yet) seen as a part of the standard research process, nor is it part of the normal research budget, and the specifics of RDM and the budget scope for funding data facilities are usually not clearly defined'.[27]

G8 Science Minister Statement, EU Commission Open science policies recommendations, Science Europe Roadmap, and the RECODE project, all has in common stressing importance of clear research data policy in each country. Main points of such policy that are addressed in this chapter are:

- giving incentives for scientists to archive and share their data, by promoting data management plans and support for proper research data management,
- open access to research data as default principle,
- recommendations about appropriate place of deposit,
- policy about selection of data based on quality and reuse potential for long-term curation, giving financial and other incentives to promote data sharing, including importance of legal and ethical guidelines to attain clarity on the legal conditions framing the envisaged re-use of research data,
- develop data-intensive research skills,
- advocate data and scientific software contributions.

---

[26] SE, 2013. Science Europe Roadmap. http://www.scienceeurope.org/wp-content/uploads/2014/05/ScienceEurope_Roadmap.pdf
[27] Knowledge Exchange Research Data Expert Group and Science Europe Working Group on Research Data, 2016. Funding research data management and related infrastructures. http://www.scienceeurope.org/wp-content/uploads/2016/05/SE-KE_Briefing_Paper_Funding_RDM.pdf

Cross-country evidence about the research data policies is available in some of the reports, which were consulted. IFDO has conducted a survey about countries' open access research data policies. Results were presented in a special report, concluding that: 'The future success of efforts in this area relies on the ability of policy makers and funders to move from high-policy statements to policy enforcements and monitoring and from short-term funding to long-term funding and institutional models that build trust and confidence.'[28]

EU Commission published a report on current adoption of its recommendations among member state countries.[29] Three groups of countries are suggested according to the existence of policies on open access to data, to their level of implementation and to the existence or not of supporting infrastructure and/or initiatives specifically established to foster open access to scientific data:

- 'No plan for a more developed policy in the near future',
- 'Some plans for a more developed policy in place or to be developed in the near future', and finally
- 'Open access policies already in place at national or regional level and/or several institutional strategies in place or important subject-based initiatives'.

Research funders' open access policies SHERPA/JULIET portal is an updatable source of information on open access policies of funders around world[30]. It includes sections on open research data. There is a recent SPARC Europe report on Open Data and Open Science Policy in Europe, which was prepared together with the Digital Curation Centre (DCC).[31]

Recent expert review[32] found that data policy initiatives that follow 2007 OECD principles and guidelines on access to public research data are usually of three kinds: sticks (mandatory rules), carrots (incentives), and enablers (soft and hard infrastructure).

3.2.1 Analysis and results

Most advanced funders are considering or already mandate the requirement that a data producer prepare a Data Management Plan (DMP) as an integral part of a funding application and of on-going project activity, aiming at Open data as the default. The informant report that '*DMP is a requirement, clear guidance is issued, support and tools are provided, the content of DMP and exemptions from full open access are defined*' regarding the first indicator criteria, represents the highest (*Managed*) development assessment[33].

What is encouraging (see Figure 11) is that not only 'CESSDA members' but also 'aspiring members' and 'non- members' are at least considering the DMP requirement. Guidance and support for a

---

[28] Kvalheim, V. and Kvamme. T., 2014. Policies for Sharing Research Data in Social Sciences and Humanities. http://ifdo.org/wordpress/wp-content/uploads/2015/07/ifdo_survey_report.pdf

[29] European Commission, 2015. Access to and Preservation of Scientific Information in Europe. http://ec.europa.eu/research/openscience/pdf/openaccess/npr_report.pdf#view=fit&pagemode=none

[30] SHERPA/JULIET, http://www.sherpa.ac.uk/juliet/index.php

[31] SPARC Europe report on Open Data and Open Science Policy in Europe. http://sparceurope.org/open-data-open-science-policy-europe/

[32] OECD, 2015. Making Open Science a Reality. OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jrs2f963zs1-en

[33] See details about measurement in Appendix 3, CESSDA SaW D3.2 Country report on development potentials 1, **Indicator 2.1.1**. http://cessdasaw.eu/content/uploads/2017/07/D3.2_CESSDA_SaW_v1.3.pdf

DMP, given by a variety of DAS infrastructure activities, represent one of the enablers. The requirement for issuing the DMP will be more easily accepted among research community if has an adequate support provision planned.

Figure 11: Research data management (RDM) policy requirements: Data management plan* by CESSDA Membership status
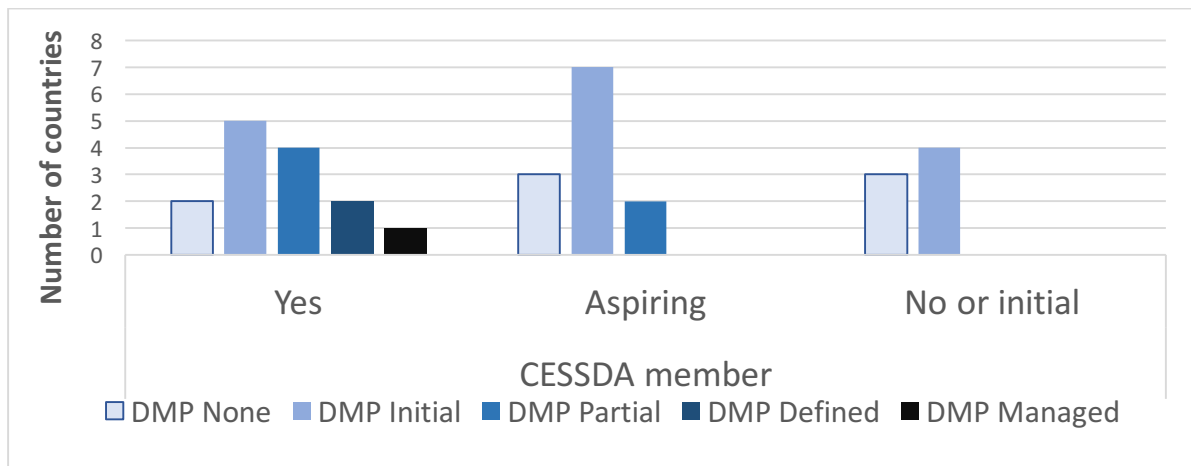
* 2.1.1. How would you characterize the general situation with regard to requirements or recommendations about preparing Data Management Plans (DMPs) as an integral part of on-going project activity, aiming at Open data as the default among public funders of social science research in your country?

1. None
2. Initial: There is growing recognition and awareness of need to require DMP
3. Partial: There is the expectation or recommendation to have DMP in place
4. Defined: Formal requirement, little monitoring and support
5. Managed: DMP is a requirement, clear guidance is issued, support and tools are provided, the content of DMP and exemptions from full open access are defined;

Existence of soft and hard infrastructures that enable sustainable arrangements for data curation and access and that take the responsibility from data producer is one of the enabling functions that motivate researchers to share data. Funders, one of the key recommendations from the EAGDA Report says, 'should agree on clear expectations for study leaders on the use of established repositories with archiving facilities, in order to assure quality and ensure discoverability of data. Such repositories need to be sufficiently and sustainably funded, with clear definition of who is responsible for them.'[34] Following this, the next indicator **(2.1.2)** of funder policy development criteria tests if and how appropriate place of data deposit is defined.

Strong, established, and trusted DAS plays a support role in the data-sharing ecosystem of national disciplinary data centres (such as CESSDA national Service Providers). The results (see Figure 12) show that existence of such a data centre provides support in shaping funders' RDM policies, specifying this as recommended or required place of deposit, and thus helps to enable researchers in preparing the DMP and support final deposit, long-term curation and access to data. More than half of CESSDA member states have reached at least 'Partial' maturity, while 'Aspiring' members and 'No or initial' members countries where no DAS service could be identified, mainly didn't consider those questions yet or are in the initial stage. As a principle, we may conclude that the

---

[34] EAGDA, 2015. Governance of Data Access. https://wellcome.ac.uk/sites/default/files/governance-of-data-access-eagda-jun15.pdf

alignment of data sharing infrastructure development and policy requirements is optimal for a sustainable data-sharing ecosystem.

Figure 12: RDM policy requirements: Appropriate place of data deposit defined* by CESSDA Membership status
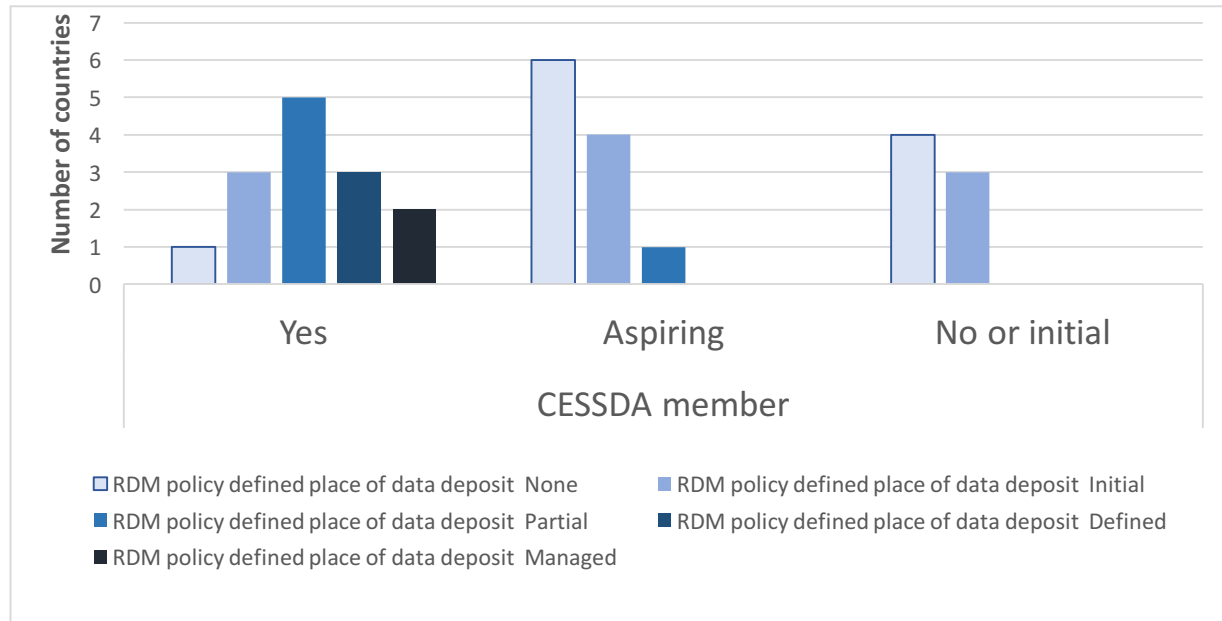
*2.1.2. Overall, in your experience, how thorough going it is that the public research funding organizations operating in your country have issued requirements or recommendations about quality-assured social science research data with associated metadata? Offering or depositing data in an appropriate disciplinary repository.
None
Initial: There is growing recognition and awareness of the need to have disciplinary specific place of deposit and support services
Partial: There is the expectation or recommendation to offer or deposit data in an appropriate disciplinary repository or equivalent data archive service
Defined: Formal requirement, little or no monitoring
Managed: Formal requirement, sanctions for not complying with regulation are in place (such as reduced payment, etc.), full support and guidance is provided

Better incentive mechanisms to promote data-sharing practices amongst researchers are needed. This is also one of the recommendations of a recent expert review: 'Reward mechanisms that are currently under discussion include widespread use of data set citation and/or proper acknowledgment of open science and data - sharing efforts in career advancement mechanisms, or grant attribution to research teams.' [35] Following this, the indicator was the question if there are provided financial and other incentives for data sharing intended for support RDM activities. Some existing questions from the EAGDA Study[36] were adapted to the context of addressing high level

---

[35] OECD, 2015. Making Open Science a Reality. OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jrs2f963zs1-en
[36] EAGDA, 2014. Establishing incentives and changing cultures to support data access. Annex A: Web survey of researchers and data managers. https://wellcome.ac.uk/sites/default/files/establishing-incentives-and-changing-cultures-to-support-data-access-report-annexes-eagda-may14.pdf

country policies. The KE and Science Europe study finds that Research Funding Organisations contribute to policy development, and half of the responding RFOs implement measures to ensure that RDM-related goals are adhered to by relevant stakeholders.[37]

Figure 13: Incentives for data sharing* by CESSDA Membership status
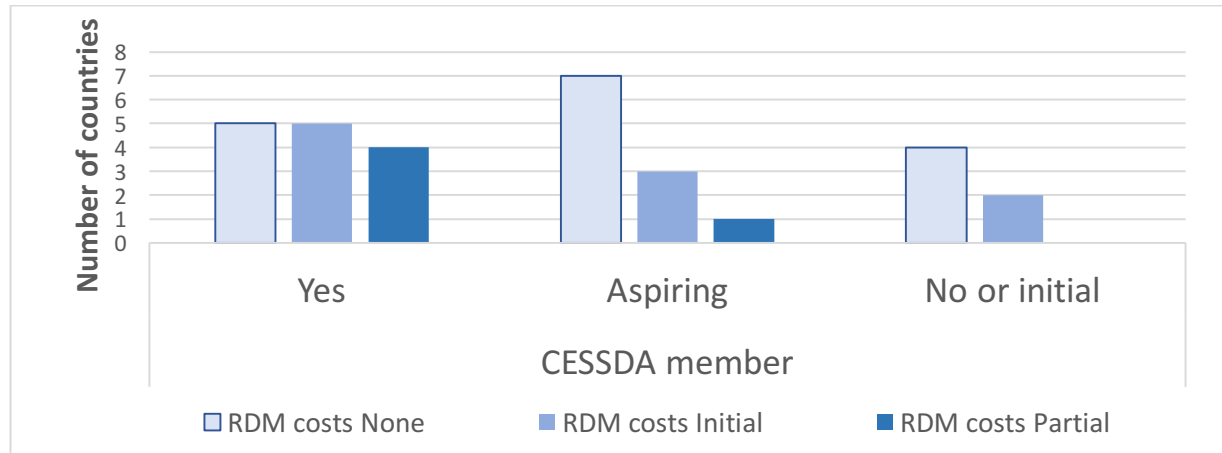
*2.1.4. Overall, in your experience, do the public research funding organizations operating in your country provide the following incentives for sharing research data with associated metadata? Cost for managing the data and preparing it for access are resourced adequately during research project lifetime. Please choose only one of the following:
No
Initial: Cost for managing the data and preparing it for access can be implicitly covered in the overall research project budget
Partial: There is explicit recognition that additional cost for preparing the data for access are legitimate project cost that can partially cover the RDM cost up to a certain limit
Defined: Costs for RDM are fully covered and adequate, based on DMP plan in project documentation

Costs for managing the data resources during the research project lifecycle and preparing them for archiving are only up to 'Partially' covered in any of the countries. The results again show the differences in prevalence according to CESSDA membership status, where member countries are only are having over half positive assessment (*Initital* or *Partial* vs. *None)*. This finding tells us that this area represents a yet unexhausted potential, where public research funding organisations can further incentivise data sharing, which according to the literature, can also have a great impact on changing habits of researchers. One of the common excuses for not sharing data among the researchers is lack of time and resources to prepare the data and documentation according to standards that are suitable for open access.

Sustainability of long-term curation of research data is a recurrent topic in most reports. A recent EC report ERAC Opinion on Open Research Data[38] addresses opportunities and challenges of data sharing. It stresses topics that are specific to open research data and data sharing arrangements: Awareness raising and incentives and reward systems need to be adapted to support sustainable and managed data curation and access. It recommends FAIR (findable, accessible, interoperable

---

[37] Knowledge Exchange Research Data Expert Group and Science Europe Working Group on Research Data, 2016.
[38] https://era.gv.at/object/document/2402

and reusable) data principles and infrastructure arrangements to be followed to enable and increase sustainability.

The RDM briefing paper[39] presents one of the challenges as: Which data should be preserved and for how long? It is important to consider the drivers regarding the benefits and value in relation to 'who pays' and 'who benefits'.[40] The Digital Curation Sustainability Model (DCSM) provides some Example Questions that can be used to get input from all relevant stakeholders.[41] For the indicator we have chosen the question if long-term curation for valuable research data assets, evaluated and selected regarding reuse potential is among funder requirements.

Figure 14: RDM policy requirements: Sustainability and long-term curation* by CESSDA Membership status
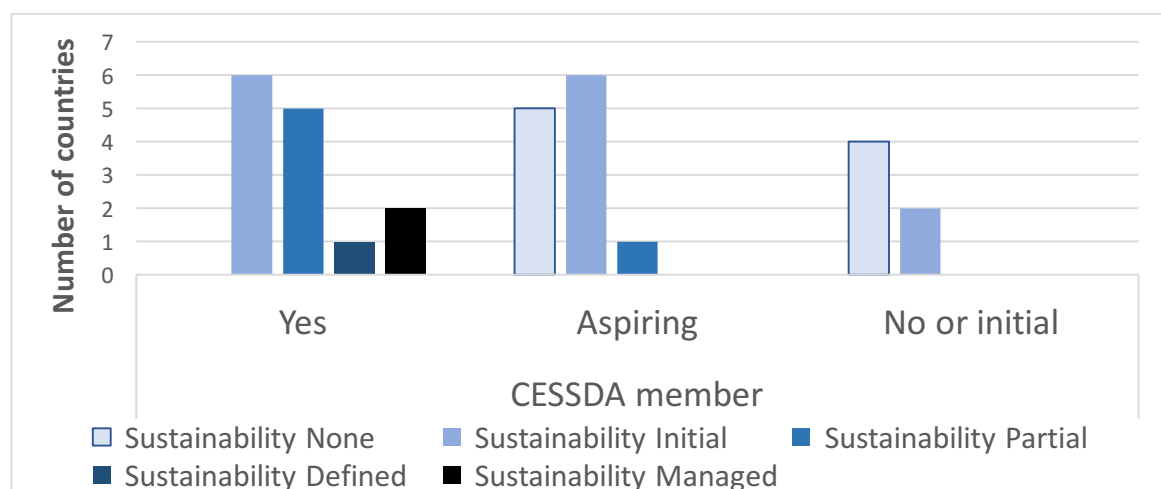
*2.1.3. How thoroughgoing it is that the public research funding organizations operating in social sciences in your country have issued the requirements or recommendations or show awareness about the following aspects of social science research data with associated metadata? Long-term curation for valuable research data assets, evaluated and selected in terms of reuse potential. Please choose only one of the following:
None
Initial: There is growing recognition and awareness about the value of research data produced and about the need for long-term preservation; scarce or no investment and support for long-term curation provided.
Partial: It is expected or recommended to assess the value of research data and resources providers declared their motivations for continuing to invest in sustaining the assets. The FAIR data principle (findable, accessible, interoperable and reusable) are highlighted in these recommendations.
Formal: Requirement to assess research data appropriate for data curation. It is understood that the best use of resources involves making choices based on value judgements and selecting material for curation. Investment and support for long-term curation is in place, based on contractual arrangements.
Managed: Formal requirement, based on contractual collective arrangements of roles and responsibilities among different stakeholders, clear definitions of what data is preserved, how it is documented, and for how long. Funder requires that periodic monitoring is performed if the best use of resources is made for expected amount of benefit.

---

[39]http://www.scienceeurope.org/uploads/PublicDocumentsAndSpeeches/SE-KE_Briefing_Paper_Funding_RDM.pdf
[40]https://www.rd-alliance.org/funding-research-data-management-and-related-infrastructures-science-europe-knowledge-exchange
[41] http://www.4cproject.eu/dcsm; DCSM Appendix 2

'Legal frameworks that explicitly accommodate open science (i.e., that are open science - friendly) are an additional means of promoting open science', says OECD expert report.[42] There is a growing recognition that legal and ethical questions can pose a barrier to sharing data.[43] To make possible the arrangement of research data access that accommodate both legally and ethically sound practices is also one of the responsibilities of funders, in making adaptation to the laws and in providing consultancy. The indicator **2.2.1.** measures if clarification and support on legal and ethical aspects of data sharing (IPR, data protection) is provided.

Figure 15: A description of ethical and legal framework important for data sharing* by CESSDA Membership status
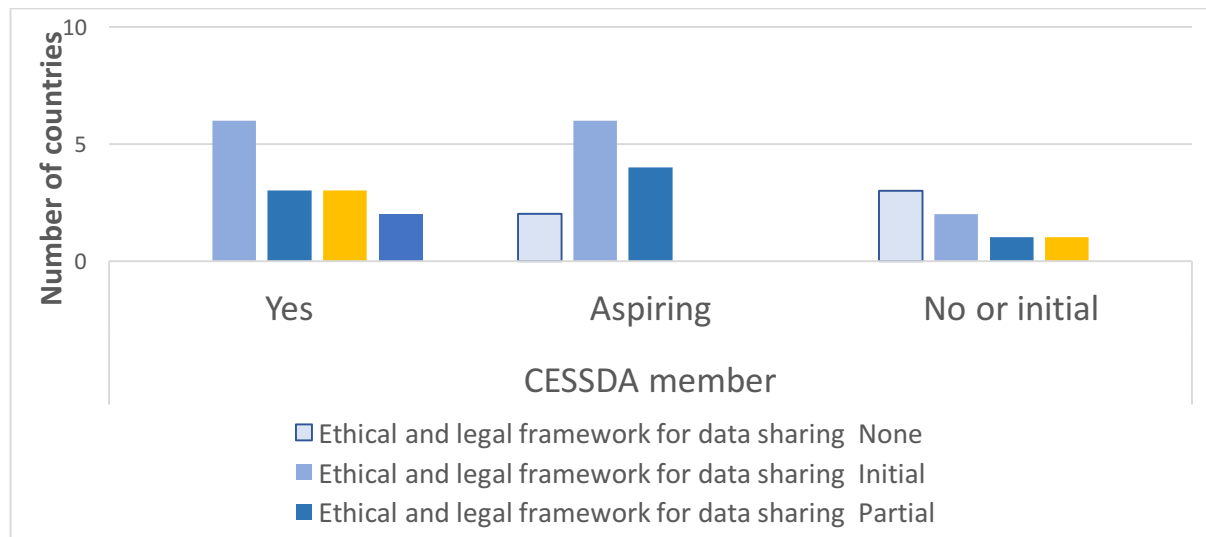
* 2.2.1. How would you characterize the general situation in your country with regard to clarification and support provided on legal and ethical aspects that facilitate social science data sharing (IPR, data protection...)?
Please choose only one of the following:
No awareness
Initial: There is growing awareness about the problem and about the need to provide clarification on legal aspects, scarce or no organised support is given
Partial: There are recommendations and guidance provided on how to respect the legal requirements while sharing data
Defined: Explicit statements about data sharing - ethical and legal aspects are embedded in ethical codes and/or legal documents that govern research and data management activities, little organised support besides guidance and recommendations is given
Managed: Organised services are widely available to support and encourage legally and ethically sound data sharing practice

Recommendation[44] on 'Legal issues' promotes the notion that research data should be open by default: 'as open as possible, as closed as necessary': 'Clear legal frameworks for the sharing of publications and reuse of data sets are needed at the national and international levels. A lack of

---

[42] OECD, 2015. Making Open Science a Reality. OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jrs2f963zs1-en
[43] see also DASISH, 2013. Report about New IPR Challenges: Identifying Ethics and Legal Challenges of SSH Research. Deliverable D6.1. http://dasish.eu/publications/projectreports/D6.1_final.pdf
[44] ERAC Opinion on Open Research Data. http://data.consilium.europa.eu/doc/document/ST-1202-2016-INIT/en/pdf

clarity on the interpretation of national and international legal frameworks may prevent the sharing or reuse of research results.[45]

There is the European Group on Ethics in Science and New Technologies (EGE) [46] with the task to advise the Commission and provide guidance and recommendations on ethical questions, relating to science and new technologies and the wider societal implications of advances in these fields. The impact of the new European General Data Protection Regulation (GDPR) on national data sharing practices needs to be examined.

Our data shows (Figure 15) that most of the countries have at least some 'Initial' awareness about the importance of clarifying the legal and ethical framework for data sharing. Some of the CESSDA members are more advanced and have reached the Defined or Managed maturity level.

3.2.2 Summary results

In summary, the conclusion is that CESSDA members on average show more advanced policy arrangements, while the 'Aspiring' and 'Initial membership countries' both categories are also more at the initial stage of acquiring awareness about the articulation of the RDM policy requirements, support and incentives. Costs for data sharing are the least articulated topic of the RDM policy areas, regardless of the CESSDA membership status.

Table 4: Funders data management strategy and legal/ethical framework (average of the 1 through 5 on the self-assessment scores by CESSDA Membership status

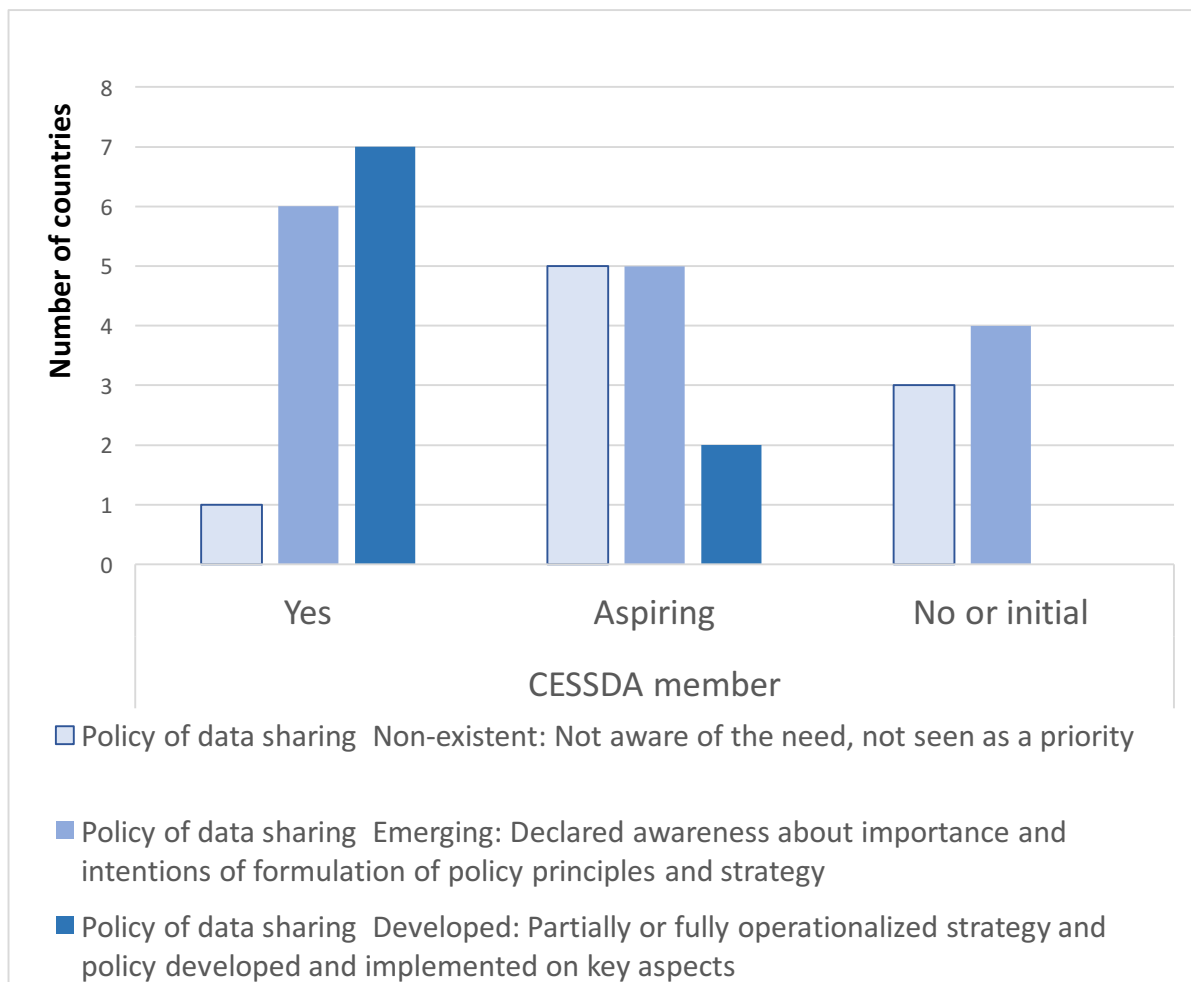| CESSDA member	Research data management (RDM) policy requirements: Data management plan	Appropriate place of data deposit defined	Sustainability and long-term curation	Incentives for data sharing: Cost for managing the data resources	A description of ethical and legal framework important for data sharing
Yes	2.6	3.1	2.9	1.9	3.1
Aspiring	1.9	1.5	1.7	1.5	2.2
No or initial	1.6	1.4	1.3	1.3	2.0
Total	2.2	2.2	2.2	1.6	2.5

---

[45] see also https://rd-alliance.org/making-open-science-reality-oecd-publishes-analysis-open-science-progress.html
[46] https://ec.europa.eu/research/ege/

Summary categorisation of the **Funders data management and sharing strategy and/or policy** (*Overarching strategy and policy to enable sustainable data access and sharing of publicly financed social science research data*) into **0 - Non-existent:** *Not aware of the need, not seen as a priority*, **1 - Emerging:** *Declared awareness about importance and intentions of formulation of policy principles and strategy supporting data sharing motivation*, and **2 - Developed:** *Partially or fully operationalized strategy and policy developed and implemented in calls on key aspects enabling data sharing*, is based on corresponding indicators described above, with median values of original scale (1 None; 2 Initial; 3 Partial; 4 Defined; 5 Managed) divided into heading concept categories following the rule: (1=0) (1.5 thru 2=1) (2.5 thru 5=2):

In summary, the majority of CESSDA members are on the Developed level regarding both the RDM policy requirements and legal and ethical framework support, with larger differences compared to the other types of countries on the general policy area. What can be observed and put into the agenda for future actions is that more than half of the countries, regardless of the CESSDA membership status, have room for improvement to reach the Developed level, where some of the countries that already reached that level can also lead by example, and share their good practices and experiences gained so far.

Figure 16: Summary RDM policy requirements by CESSDA Membership status

□ Policy of data sharing Non-existent: Not aware of the need, not seen as a priority

■ Policy of data sharing Emerging: Declared awareness about importance and intentions of formulation of policy principles and strategy

■ Policy of data sharing Developed: Partially or fully operationalized strategy and policy developed and implemented on key aspects

Finally, individual country maturity self-assessments are shown in Table 5.

Table 5: Funders data management strategy and legal/ethical framework (1'no' thru 5 'Managed') by Country

Country	Research data management (RDM) policy requirements: Data management plan	RDM policy requirements: Appropriate place of data deposit defined	RDM policy requirements: Sustainability and long-term curation	Incentives for data sharing: Cost for managing the data resources	A description of ethical and legal framework important for data sharing
Albania	2	1	1	2	2
Belgium	2	2	2	2	2
BiH	1	1	1	1	1
Bulgaria	2	2	2	.	4
Croatia	2	2	2	2	2
Cyprus	1	1	1	1	3
Czech Republic	1	1	2	1	2
Denmark	2	3	2	1	2
Estonia	2	2	2	2	3
Finland	4	4	3	3	3
Germany	2	2	2	1	2
Greece	1	3	3	1	4
Hungary	3	3	3	1	3
Ireland	1	1	2	1	3
Israel	2	2	2	1	3
Italy	2	1	1	1	2
Kosovo	2	2	2	1	1
Latvia	2	1	2	1	2
Lithuania	2	2	2	2	2
Macedonia	1	1	1	1	1
Montenegro	2	2	.	2	2
Netherlands	4	4	4	3	4
Norway	3	5	5	3	5
Poland	2	2	1	3	2
Portugal	3	3	3	2	2
Romania	1	1	1	1	2
Russia	3	.	2	.	1
Serbia	2	1	1	1	1
Slovakia	1	1	1	1	3
Slovenia	3	3	3	2	3
Sweden	2	3	2	2	4
Switzerland	3	4	3	2	2
United Kingdom	5	5	5	3	5
Total	2	2	2	2	3

3.3   Data sharing culture and enablers for data sharing

The aim of the section is to describe the scientific community dimension of the data sharing culture in relation to data archives and services. In particular, this includes the following aspects of **data sharing culture**:

- General data sharing practices that exist in research in general and in the social sciences in particular, and that therefore influence the availability of data for reuse,

- Related routines and researcher attitudes, considering real and perceived barriers to data sharing.

Further on, it includes assessments about **enablers for data sharing**:

- Career progression in academia as a motive for data sharing,

- Awareness, knowledge and quality of RDM practices, and data documentation and preservation practices,

- Available RDM support services and tools for data sharing and reuse.

For this section, in particular the self-assessment was difficult for some of the informants due to lack of information or hard to answer subjective evaluations, which in turn lead to about a quarter answering 'Unable to estimate', a majority of those in the 'No or initial' CESSDA membership status. The interpretation of the obtained answers should thus be taken with caution.

One or several experts who filled in the self-assessment web form or answered interview questions mainly provided the data. When possible, the information provided by the experts was complemented, validated or substituted with aggregated results of past studies which were based on samples of members of research communities. Specific notes about the situation and sources of information in each of the countries are described in CESSDA SaW D3.2 Country report on development potentials 1.

3.3.1   Data sharing practice, routines and attitudes

The prevalence of data sharing and reuse was measured as whether or not researchers provide (share) data to the community and can access the data they need[47]. All but 10 countries provide answers to at least one question. Out of the countries not providing any answer, 1 was a CESSDA member, 3 – aspiring countries, and 6 countries with no or initial data services.

The culture of data sharing, assessed through the reported proportion of researchers sharing their own data, is mainly assessed as 'Low' (*10% of less researchers sharing data they produce*) in aspiring CESSDA member countries (7 out of 9) and countries with no or initial DAS activities (4 out of 6) (see Table 6). In CESSDA member countries the level of sharing data reported is somewhat higher - more than a half have indicated that level of sharing is at least medium or high.

---

[47] See Appendix 3, SaW D3.2 report: indicators 3.1.1 and 3.1.2.

The difference between CESSDA member countries and the other countries is even more profound in 'Data access' assessments. As expected, the more established data services[48] have a role in making more data available in a transparent manner. Thus, most of the CESSDA member countries, 9 out of 11, have estimated the proportion of researchers having access to data they need as medium or high, while in other countries it is about half.

Table 6: Prevalence of data sharing and reuse by CESSDA Membership status (Countries count)

CESSDA member	Proportion of social sciences researchers sharing data they produce*			Proportion of social sciences researchers that can access data they need**		
	Low	Medium	High	Low	Medium	High
Yes	4	4	2	2	4	5
Aspiring	7	2	0	3	1	2
No or initial	4	2	0	3	1	1
Total	15	8	2	8	6	8

* 3.1.1. What proportion of social science researchers in your country have shared the research data they produced in the period between 2011 and 2016? Please give an estimate of the proportion of researchers, based on experience in your institution and published reports on data sharing (if available). Please choose only one of the following:

    low (0-10%)

    medium (10-30%)

    high (>30%)

    Unable to provide estimate for 2011-2016

**3.1.2. What proportion of social science researchers in your country have been able to access existing third party data they need in the period between 2011 and 2016? Please give an estimate of the proportion of researchers, based on experience in your institution and published reports on data access and reuse (if available). Please choose only one of the following:

    low (0-10%)

    medium (10-30%)

    high (>30%)

    Unable to provide estimate

Transparent and formalised data sharing routines are another aspect, characterising a developed data sharing culture. Five data sharing routines of social sciences researchers were ranked by preference, based on expert evaluations and previous survey data, and data was available for 26 countries. In general, more informal and less transparent data sharing routines, like project and personal websites and informal contacts are ranked as first more often than formal and transparent routines.

Difference among CESSDA members and other countries is visible in the first selection of ranking existing common routines or channels via which data are made available (indicator 3.1.3, Table 7). Data archive or repository is ranked first in 5 countries, and these are exclusively CESSDA member countries. More informal and less transparent routines, like project and personal websites and

---

[48] As we already show in Chapter **Error! Reference source not found. Error! Reference source not found.**, the higher maturity of DAS is also more of the characteristics of CESSDA member countries.

informal contacts are in general ranked first more often, and a division between CESSDA members, aspiring countries and countries with no or initial DAS is about the same. This indicates that if a formalised and transparent channel for sharing data is available, it is also preferably more used. What is shown here is only an answer about the rank 1 of the data sharing routines.

We should bear in mind that most of the data services assess and select the data for inclusion based on the quality and reuse potential. In return, they offer different support services and are adapting continuously to meet the needs of social science research community. Some DAS also provide lightweight options for self-archiving, in case that data does not fulfil the highest standards. There are other options available in some countries, such as institutional repositories. Therefore, we may expect the preferred options for data sharing (formal and transparent routines – data archives and supplementary data in journals) gradually to gain more popularity, again if supported with the appropriate incentives and removed barriers.

Table 7: Rank 1 routines of data sharing by CESSDA Membership status

	Routines for sharing social science* data_Rank1				
CESSDA member	Data archive or repository	Supplementary data in a journal	Via project or personal websites	Via informal contacts	Other
Yes	5	0	2	3	1
Aspiring	0	1	2	5	0
No or initial	0	2	2	3	0
Total	5	3	6	11	1

* 3.1.3. Can you rank the following routines for data sharing to reflect the preferences of the social science research community in your country between 2011 and 2016?
Please base the ranking on experiences in your institution and published reports on data sharing, access and reuse (if available).
Please number each box in order of preference from 1 to 6
Data archive or repository
Supplementary data in a journal (alongside paper)
Via project or personal websites
Via informal contacts (peers and colleagues)
*Other (please, describe in the field below)*
Unable to provide estimate for 2011-2016

The separate analysis also revealed that transparent and formal data sharing routines are related to a more mature policy framework for data sharing, and that maturity of policy framework in general is positively related to the estimated level of data access, which might reflect the focus on data with high potential for reuse in situations with limited resources.

Researchers' attitudes to data sharing (indicator 3.1.4) in the social science research community in the country reflect a variety of perceived enablers and barriers of data sharing. In this audit, the attitudes have been estimated by the country experts on a 5-points scale from True to False. Answers to the questions were based on recent reports (if available), and recent experiences of the reporting institution or expert. In the Figure 17 below Positive and Negative (or neutral) attitudes estimates are shown. The propensity of a prevailing positive attitude is only for the 'Data sharing

has no benefits at all' and the 'Reuse of existing data can answer new research questions and facilitate advancement of science' statements. This result shows that regardless of the advancement of countries in other areas of the data sharing ecosystem, the researchers' attitudes tend to be slowly progressing towards the recognition of data sharing.

Figure 17: Researchers' attitudes to data sharing (+) = (Positive); (-) = (Negative or Neutral)
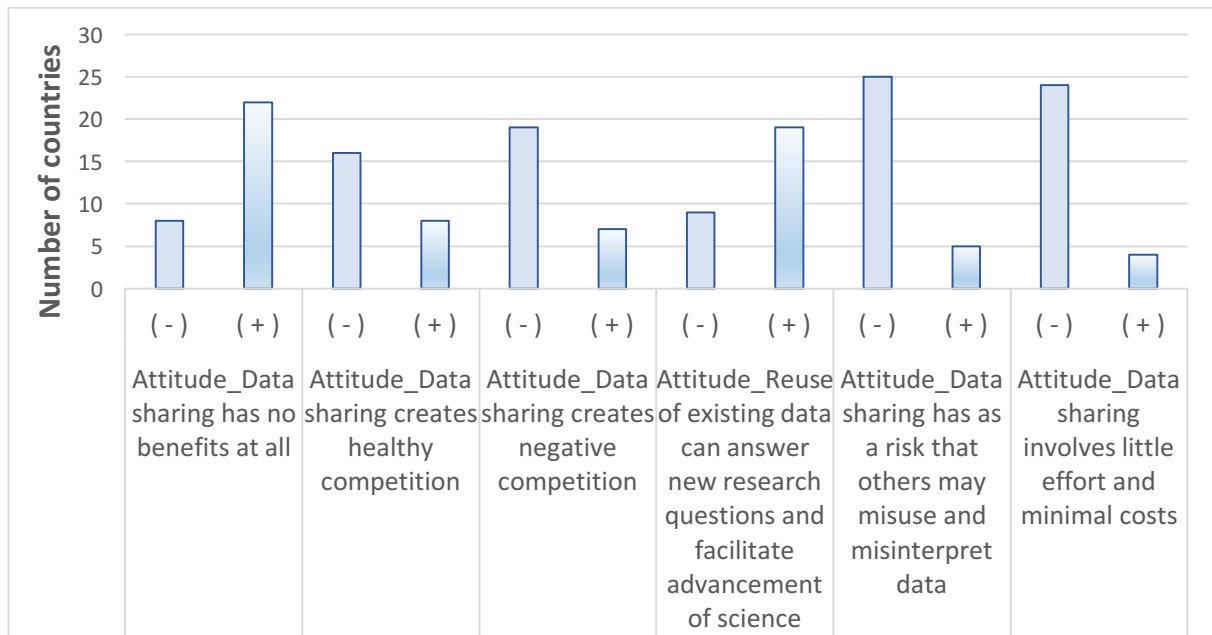
3.1.4. Please score the statements below to best match the overall attitudes of social science researchers in your country, based on experience in your institution and previously published reports, in the period from 2011 to 2016, on a five point scale from 5 - true, to 1 - false.

Please choose the appropriate response for each item: TRUE FALSE *Unable to provide estimate for 2011-2016*

Data sharing has no benefits at all.
Data sharing creates healthy competition in research.
Data sharing creates negative competition (for example, being scooped and therefore reduced publication opportunities) for the researcher.
Reuse of existing data can answer new research questions and facilitate advancement of science.
Data sharing has as a risk that others may misuse and misinterpret data.
Data sharing involves little effort and minimal costs.

Based on the estimated attitudes towards the six above-mentioned items, the attitudes were classified as mainly negative towards data sharing, neutral or mixed, and mainly positive. The CESSDA member countries tend to somewhat more often report mainly negative perceived attitudes towards data sharing in social sciences communities (4 out of 12 countries) or neutral or mixed attitudes (6 out of 12 countries). Aspiring members report mainly neutral or mixed attitudes (7 out of 11 countries), and countries with no or initial DAS tend to report only mainly positive attitudes (all 7 countries). This affirms the results presented in the previous paragraph, that the CESSDA countries report more down-to-earth attitudes related to practical experience, while countries with no or initial DAS see everything more hypothetically positively.

Figure 18: Attitudes towards data sharing by CESSDA Membership
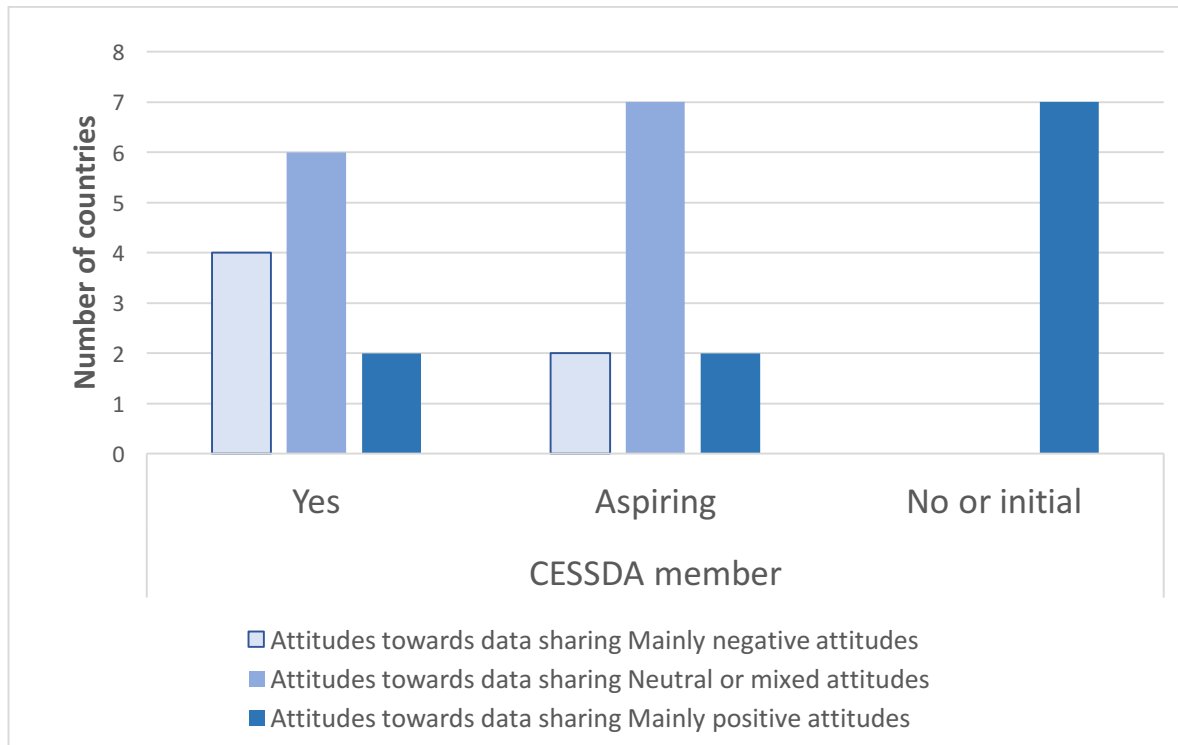
The aggregation of **Data sharing culture**, defined as *Prevalence of data sharing and reuse, existing routines and attitudes*, is based on respective values of the majority of indicators characterizing the dimension, data sharing and access, transparency and formality of the data sharing routines, and attitudes towards data sharing. This leads to the maturity heading concept categorisation as:

> **0 - Underdeveloped:** *Data sharing and reuse is rare or not existing in the social sciences research community; with no existing routines for sharing and negative, indifferent attitudes;*

> **1 - Developing:** *Data sharing and reuse is not that common, with largely informal, non-transparent routines, and indifferent attitudes to sharing culture;*

> **2 - Developed:** *Data sharing and reuse is very common, there are formal and transparent routines for data sharing, attitudes positive.*

If the majority of the indicators have no meaningful score assigned (due to lack of information or resources to acquire information via alternative sources), the maturity level for a particular country was not estimated[49].

Altogether, data sharing culture could be considered as developed in 7 countries, developing in 19 and underdeveloped in 7. The differences according to the CESSDA membership status are small,

---

[49] Details are described in **Appendix 4**, CESSDA SaW D3.2 Country report on development potentials 1: Classification protocols for assigning development level.

with all type of countries being prevailingly estimated as in the middle (developing) category (see Figure 19).

The summary scores again show that perceived attitudes of CESSDA member countries' social scientists on average looks as even more critical than in other counties. Perhaps the presence of developed DAS increase the expectations about data sharing activities, whereas the majority of research community is lagging behind. Data sharing culture is an area where we can expect only gradual changes, and even the established data services need to be actively promoted among the science community to foster real change. Simultaneously, these services should be further developed and adapted to meet the needs of research community. Similar results of gradual changes in data sharing culture among scientific community in developed countries were also observed in other studies, with respect of differences among different disciplinary communities[50].

Yet, the present results need to be further validated and the situation in each of the countries assessed separately, using national reports and studies, where available. These results might indicate a bias in the measurement as well, because estimated proportions of researchers sharing and being able to access data in the audit has been very much reliant on subjective expert evaluations, and included often experts coming from DAS. Where surveys are used as a base for judgement, the bias stems from self-selective samples of web surveys of researchers, where it is suspected that those with strong opinions, more interested and active in data sharing and access, are answering more often.

One of the explanations to this clustering in the middle and no clear-cut relationship between CESSDA membership data sharing culture indicators might be that CESSDA membership could be seen as a result of favourable conditions of a broader data sharing ecosystem. Data sharing culture is an integral part of this ecosystem. However, it should be seen as result of different enablers, incentives and removed barriers. We may expect that stricter policy recommendation and funders' requirements, legal and ethical framework that foster open data sharing etc., will have stronger and more direct impact towards establishing a data sharing culture, when combined with the established DAS and related availability and variety of support services. Implementation of funders policies is needed that will address the specific barriers and provide a consistent package of different incentives and enablers (mature DAS being one of them) for data sharing. After this is realised, we may expect more visible change in data sharing habits occur also among the scientific community. Changing data sharing culture involves parallel efforts to improve quality of data, change publication culture as well as practice regarding evaluation of scientific impact.

---

[50] Youngseek and Stanton (2012) found data sharing being seen as critical for new science by biological, chemical and ecological scientists, whilst computer scientists, engineers, mathematicians and physical oncologists stated the opposite. RECODE (2013), by exploring attitudes towards open data, found data sharing to be limited to particle (astro)physics (despite the collaborative nature of research), health and clinical research (due to ethical constraints), and archaeology; and prevalent in bioengineering and in environmental research. Significant variations in attitudes and practices were found within each discipline as well (RECODE, 2014).

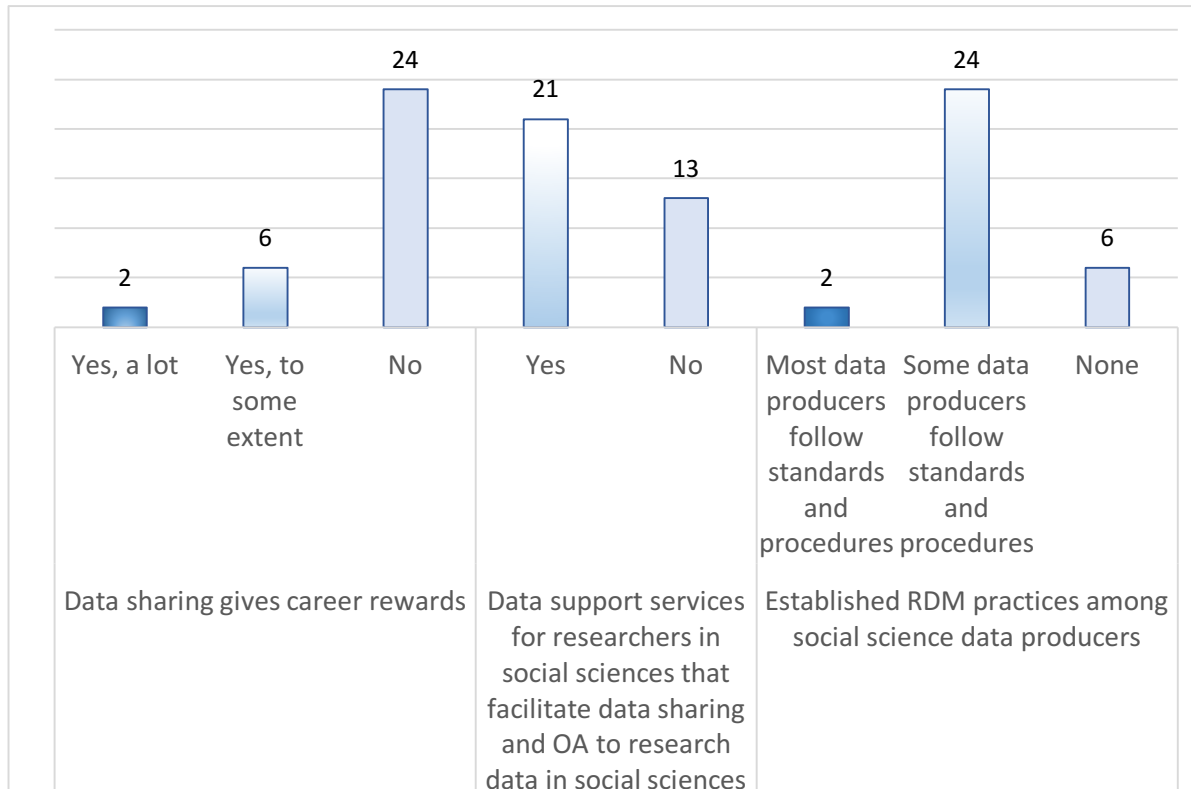*Figure 19: Overall data sharing culture estimate by CESSDA Membership status*
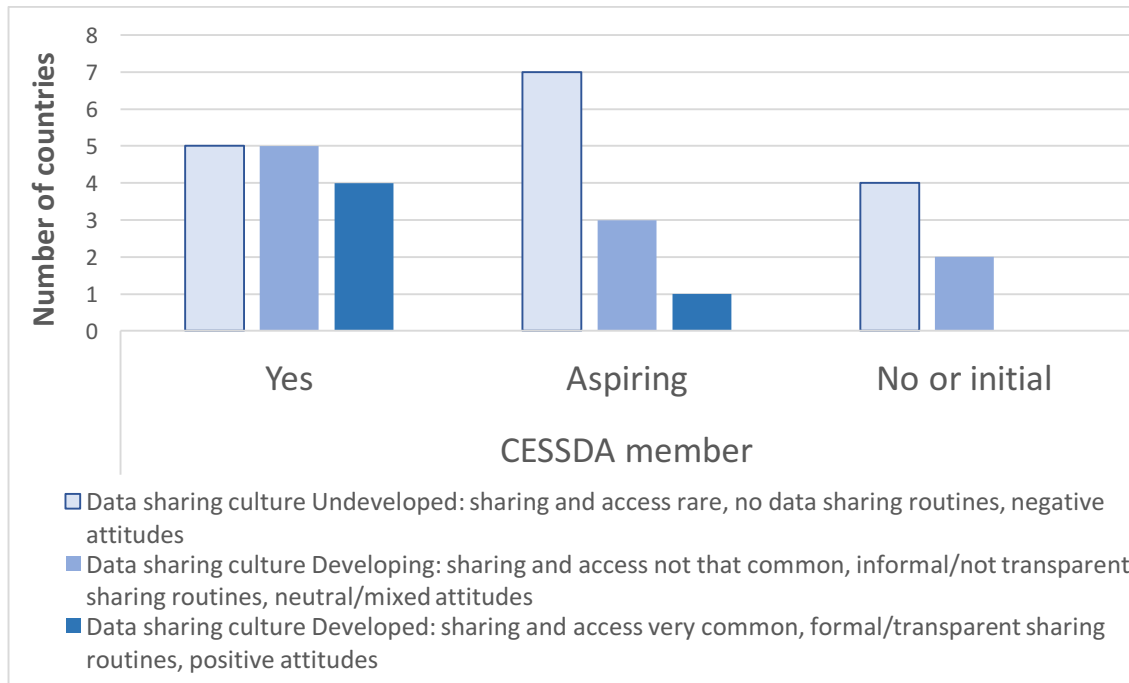
The impact of data support services and career incentives for a data sharing culture is analysed in the next subchapter.

3.3.2   Enablers for data sharing

This section focuses on the questions about incentives and career rewards, related to data sharing in the social science community in a country, as well as skills, RDM practices of researchers, and available data support services and tools for data sharing and reuse.

Figure 20: Enablers for data sharing*

Yes, a lot	Yes, to some extent	No	Yes	No	Most data producers follow standards and procedures	Some data producers follow standards and procedures	None
2	6	24	21	13	2	24	6
Data sharing gives career rewards			Data support services for researchers in social sciences that facilitate data sharing and OA to research data in social sciences		Established RDM practices among social science data producers		

* 3.2.2. Would you say that there are career rewards related to data sharing, if you consider social science researchers in your country? This can include any kind of career rewards, e.g. influence on career progression within institution or community or due to government rules, higher success rate in obtaining research funding, better standing within the research community and other. Please choose only one of the following:

Yes, a lot

Yes, to some extent

No

3.2.3. According to your experience, are there data support services provided to social science researchers in your country, that facilitate data sharing and/or Open Access to research data (regarding for example, data management plans, data preservation, and data access)? Examples of services: web guidance, trainings, workshops, webinars, online reference materials, helpdesk or contact and info point, metadata creation and publishing tools (Nesstar, NADA, DataVerse...), linkages between papers and data, support to data management planning, support to long-term preservation of data, access to data. Please choose only one of the following:

Yes

No

3.2.4. Would you estimate that social science data producers in your country follow data management and data documentation standards and procedures that facilitate data reuse? Please choose only one of the following:

Most data producers follow data management and data documentation standards and procedures.

Some data producers follow data management and data documentation standards and procedures.

None

Out of 33 countries (Figure 20), providing an answer to the questions related to enablers and barriers of data sharing, a majority – 24 indicated there are no career rewards, while the opposite is true with regards to data support services: 21 countries have some data support services for researchers in the social sciences that facilitate data sharing and OA to research data, while 12

have indicated that there is none. Finally, most have indicated that there are some researchers following established RDM practices.

The enabling effect of data support services and RDM practices for data sharing and access, identified in our literature review, can be observed also on a country level. Established RDM practices among social science data producers in a country are also related to a higher estimated proportion of researchers being able to access data they need (Figure 21).

Figure 21: Proportion of researchers being able to access data they need by established RDM practices
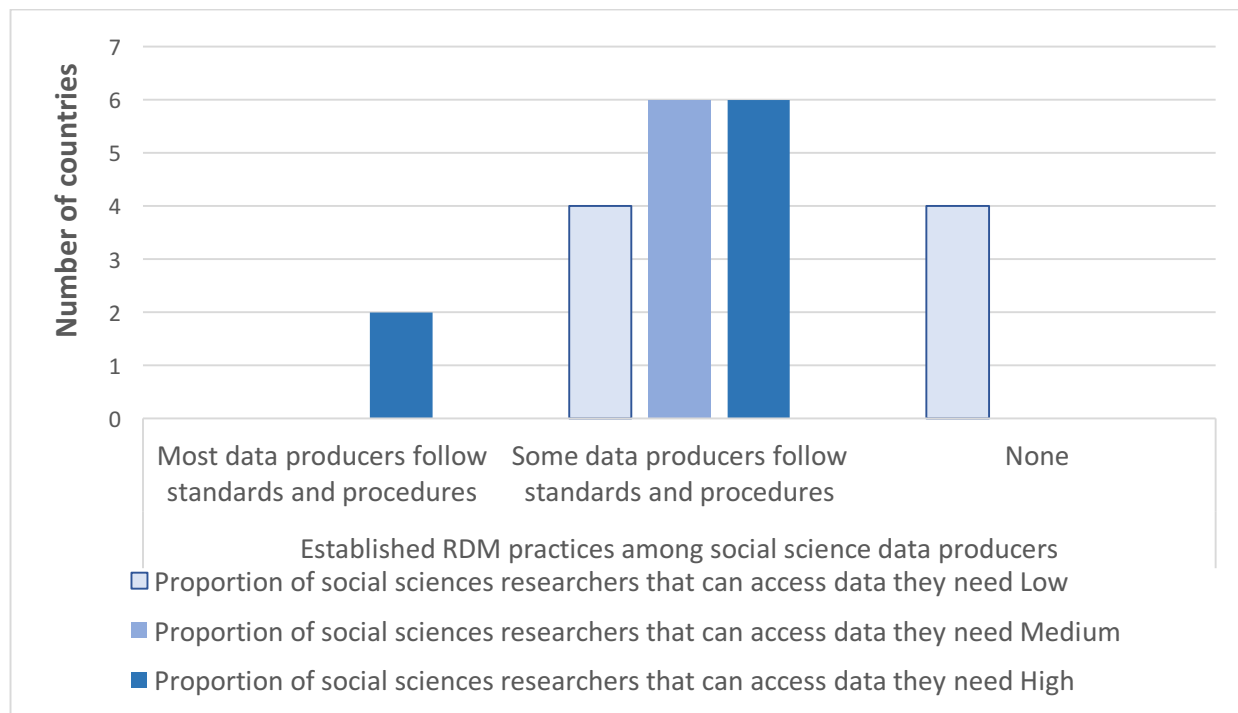
Categorisation of **Enablers for data sharing** (*Enablers for data sharing: incentives, rewards, skills, support services and tools)* to summary score below was carried out based on the sum of positive answers (Yes or partially yes), where the Underdeveloped category represents the absence of enablers ('No' or 'None'), the Developing category represents 1 or 2 positive answers, and the Developed category represents a situation when all three enablers are assessed positively[51]:

> **0 - Underdeveloped:** *No enablers for data sharing*;
>
> **1 - Developing:** *Some enablers for data sharing*;
>
> **2 - Developed:** *Many enablers for data sharing.*

Existing and aspiring CESSDA member countries both show the highest presence of some enablers (Developing). In most of the countries, at least some incentives, rewards, skills, and support services are present, which is a promising ground for further development also in the area of

---

[51] Details of heading concept categorisation are described in **Appendix 4,** D3.2 Country report.

general data sharing practices and attitudes. Data support services are available in most CESSDA member countries and aspiring member countries and even in some countries with initial or no DAS. However, there is still a space in most of the countries for further supplying of incentives. Other studies also emphasize[52] that the reward mechanisms and support in RDM would be better accepted rather than exposing researchers to sanctions. This relates to the fine-tuning of the policy of OA to research data, which funders prescribe as mandatory for all publicly funded projects, to put an emphasis on enablers and incentives while promoting the policy among the research community.

Figure 22: Overall enablers for data sharing estimate by CESSDA Membership status

_____
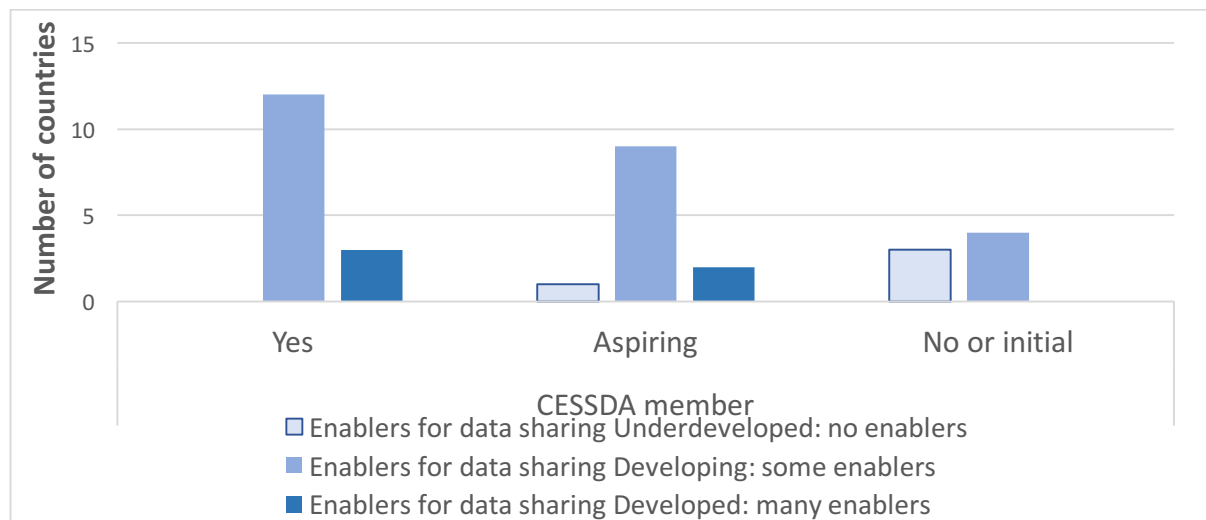
[52] See EAGDA, 2014; RECODE, 2014; Costas et al, 2013; Force 11, 2013 mentioned in the introduction.

3.4   Summary of the broader data sharing ecosystem

The Table 8 (below) provide a summary of the data sharing ecosystem heading concepts by country. Sorted from left to right and from highest to lowest value this can give a glimpse of which countries are similar in their development pattern. It is important to bear in mind that data for all of the countries were not available while constructing heading concept values, where some countries even at the heading level has some empty values. Inevitably, there is some arbitrariness in construction of the heading concepts, and subjective and other bias when comparing countries. The purpose is simply to show where there potentially some gaps exists, which can be further evaluated and addressed by different stakeholders in each of the countries.

The first group of countries, starting with Netherlands and including Switzerland, characterise a comparatively high-level of social science sector development, a progressive funders policy and a strategy orientation towards enabling data access. Probably at least Germany and UK would fall into this group of the most highly developed data sharing ecosystem as well, if there would not be some arbitrariness in how the indicators perform, they would appear in the second group of countries.

A second group of countries, characterised with mixed results regarding the SSH sector development, can be further divided according to the presence of strong enablers and support in some other ecosystem components. Strongly enabling and supportively assessed is research data policy reported in Hungry, United Kingdom and Russia, while some of the data sharing culture and enablers in scientific community are assessed in highest developing level in Poland, Germany, Lithuania, Italy, Albania and Greece. For any of the countries in this group, some of the data sharing ecosystem components are exceptional, and some are in a transitory developing level.

The last group of countries, not counting France (lack of sufficient data), are generally just starting to build their data sharing support ecosystems. Any of those can quickly reach the two other groups, in particular as a result of an advocacy and expertise demonstrated during participation in various widening project activities, such as SERSCIDA, SEEDS and now CESSDA SaW. The overview of activities and potentials towards establishing the national DAS in some of those countries will be described in next section.

Table 8: Data sharing ecosystem (0 – Underdeveloped; 1 – Developing; 2 – Developed) by Country

Country	Development of research data production in SSH	Development of social science sector	Overarching strategy and policy to enable sustainable data access and sharing of publicly financed social science research data	A description of ethical and legal framework important for data sharing	Data sharing culture: proportion sharing and accessing data, formality and transparency of routines of, and attitudes to data	Enablers for data sharing
Netherlands	2	2	2	2	1	2
Slovenia	2	2	2	2	1	2
Finland	2	2	2	2	1	1
Norway	2	2	2	2	1	1
Portugal	2	2	2	1	0	1
Estonia	2	2	1	2	1	1
Sweden	2	2	1	2		1
Denmark	2	2	1	1	1	1
Switzerland	2	1	2	1	0	1
Belgium	2	1	1	1	1	1
Latvia	2	1	1	1	1	1
Hungary	2	0	2	2	1	1
Russia	2	0	2	0	0	2
Bulgaria	2	0	1	2	1	1
Poland	2	0	1	1	2	2
Germany	2	0	1	1	2	1
Czech Republic	2	0	0	1	1	1
United Kingdom	1	2	2	2	1	1
Lithuania	1	2	1	1	2	1
Italy	1	2	0	1	2	1
Croatia	1	1	1	1	1	1
Albania	1	0	1	1	2	1
Ireland	1	0	0	2	1	1
Romania	1	0	0	1	0	0
Israel	0	2	1	2	1	1
Greece	0	2	1	2	0	2
Cyprus	0	2	0	2	1	0
Serbia	0	2	0	0	0	1
Montenegro	0	1	1	1	2	1
Kosovo	0	1	1	0	1	0
Slovakia	0	1	0	2	1	1
Macedonia	0	0	0	0	2	1
BiH	0	0	0	0	1	0
France		2			0	1

We conclude the overview by showing the heading concepts' values averages by CESSDA membership status. Reaching CESSDA membership is mostly related to the 'Overarching strategy and policy to enable sustainable access and sharing of social science data' – the biggest differences between members and non-members occur here. As expected, it is the funders' strategic decision to recognize the need to support strong and sustainable DAS. This gives an incentive to the science community to follow the policy recommendations. Advanced policy recommendations and strong DAS can lead to a sustainable data sharing ecosystem.

Figure 23: Heading concepts (0 – Underdeveloped thru 2 – Developed) by CESSDA membership
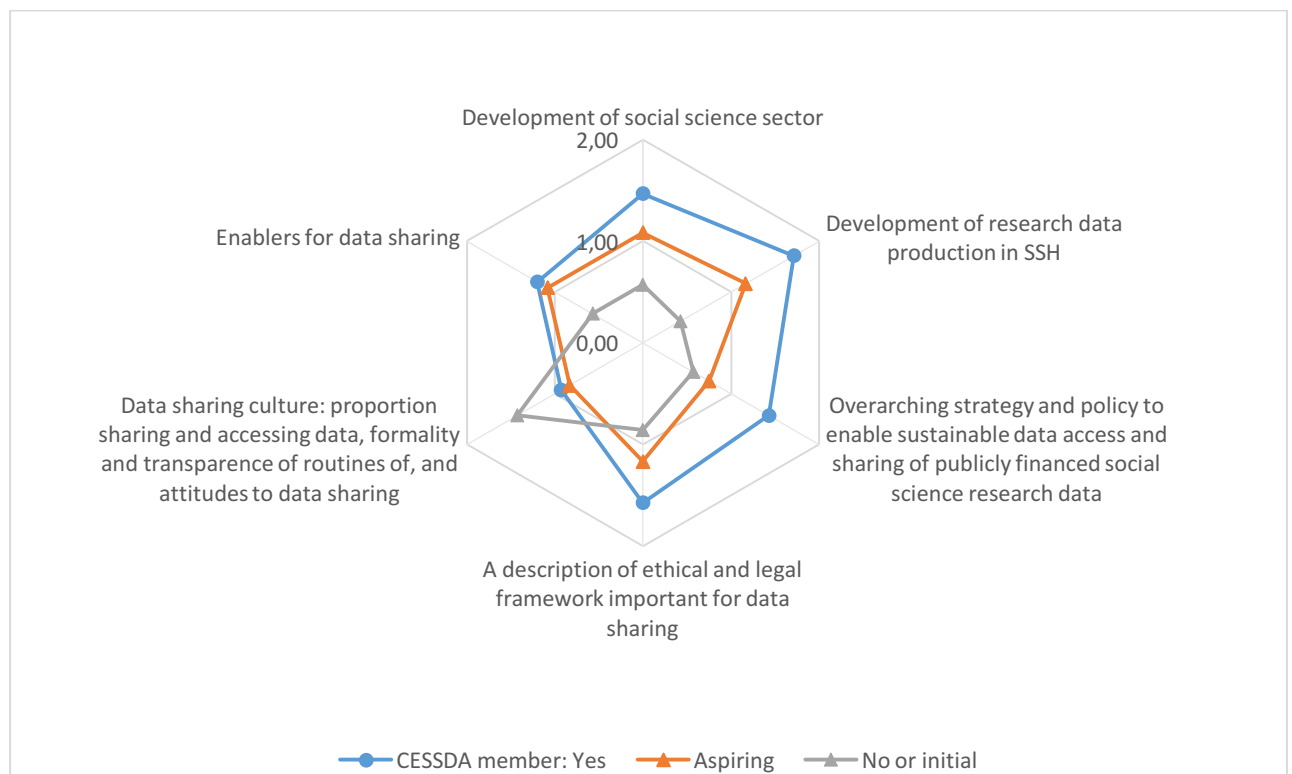
The established traditions in research data production also provide the support environment for a more robust data sharing support infrastructure: there is a supply and demand for enhanced data management and data access services. Aspiring countries are also close to the existing members on that second dimension, which means that there is only a small step to reach the membership status, however such a decision needs to be further supported together with an upgrade of the other components of enablers and incentives, in order to make the whole system sustainable in the end.

The differences between types of countries are the smallest in the area of Enablers of data sharing. Even a reverse relationship can be observed at the Data sharing culture. There are only 7 countries among those responding, that are in the Initial membership category. The result can be thus also affected by the few outliers. Fully described in the country reports, some of the values of self-assessment by informant from countries are inconsistent with the results of the desk research of previous studies. But there is also a process of gradual building of data sharing culture that is based on the formal, transparent and trusted data sharing channels. There the expectations about the participation in data sharing are set higher as a goal, which still needs to be achieved in the future.

4.   DAS proto-activities results

Focus of the description of situation for the countries that do not have national DAS was put on exploring the conditions for establishing a data service that could in the future obtain the role of CESSDA national service provider, labelled as DAS proto-activities. Some of the European countries have a long tradition of research data management (RDM) and data archiving in social sciences, while others are at the very beginning. With the launch of [Open Research Data Pilot](#) ([Horizon 2020](#)), [European Open Science Cloud](#) and the adoption of [Digital Single Markets](#) strategy, the RDM strategies and implementations are becoming an important factor in research infrastructure development in countries where no formal data archive service (DAS) exists or where RDM support infrastructure is not integrated.

We identified pioneers in these countries and the key players involved in DAS-related activities, and described their current expertise, the level of technical infrastructure development, as well as their overall activities and potential regarding open science issues in general. Thus, the goal of this part was to assess data archive proto-activities and open access (OA) support activities in countries where no formal DAS exists yet. Identifying those activities allowed us to detect actors and institutions that could play a key role in the elaboration of new national DAS.

In order to identify DAS proto-activities and OA support activities, and to detect key actors and institutions, we used the following indicators:

1.   **Determinant factors of DAS proto-activities include:**

- Availability of technical infrastructure

- Organisational first steps;

- Availability of trainings and capacity development for (future) collaborators in the main DAS activities: RDM, data preservation, and data access.

These factors allowed us to assess the overall DAS activities implementation type of a given country. According to the results, we grouped the countries into three types of DAS activities implementation:

0 - No DAS activities: No DAS exists, and no activities – related to technical infrastructure, organisation or capacity building – are provided on a national and/or institutional level;

1 - Basic DAS activities: No DAS exists, but basic activities – related to technical infrastructure, organisation or capacity building – are provided on a national and/or institutional level;

2 - Advanced DAS activities: No DAS exists, but advanced activities – related to technical infrastructure, organisation or capacity building – are provided on a national and/or institutional level.

2.   **Determinant factors of OA support activities include:**

- The availability of OA support for researchers from institutions like university libraries;

- OA projects or initiatives that can be enablers.

These factors allow us to assess the OA support activities implementation type of a given country. According to the results, we also grouped the countries within three types of OA support activities:

0 - No OA support activities: Support activities to encourage and facilitate OA are rare or not existing in the social science research community;

1 - Basic OA support activities: Some support activities to encourage and facilitate OA exist, but not on a regular basis and only in few institutions;

2 - Advanced OA support activities: Support activities to encourage and facilitate OA are well-established (common and in most institutions), known and used by the social science research community.

Three channels of information were used. First a literature review and desk research were conducted. Then, when representatives of a future DAS were known (e.g. relevant partners of CESSDA SaW, SERSCIDA or SEEDS projects), they completed the self-assessment survey. Otherwise (or to supplement the previous information), the self-assessment survey was sent to informants who have a particularly good view of the social science research community in their respective countries. Those informants were identified during the desk research (e.g. Open Access Representatives (OpenAIRE National Open Access Desks) or relevant researchers, policy makers, institutions, etc.

4.1   Results

Twenty European countries were targeted in this section – those that the established DAS could not be determined for them. Information was gathered for seventeen of them, leaving out Iceland, Malta and Turkey. Very little information was found for Luxembourg, Moldova, and Spain, where only 1 or 2 indicators out of 6 where filled out.

Overall results (see Figure 24) show that some countries with no established national DAS for the social sciences are already far along with the activities towards their establishment of a national DAS (e.g. Croatia, Israel, and Serbia), while others are in the beginning stage. The OA support activities are particularly advanced across institutions in Croatia, Israel and Moldova.

Figure 24: Overall DAS proto-activities by country (Scale values from 0 - no activities to 2 - advanced activities)
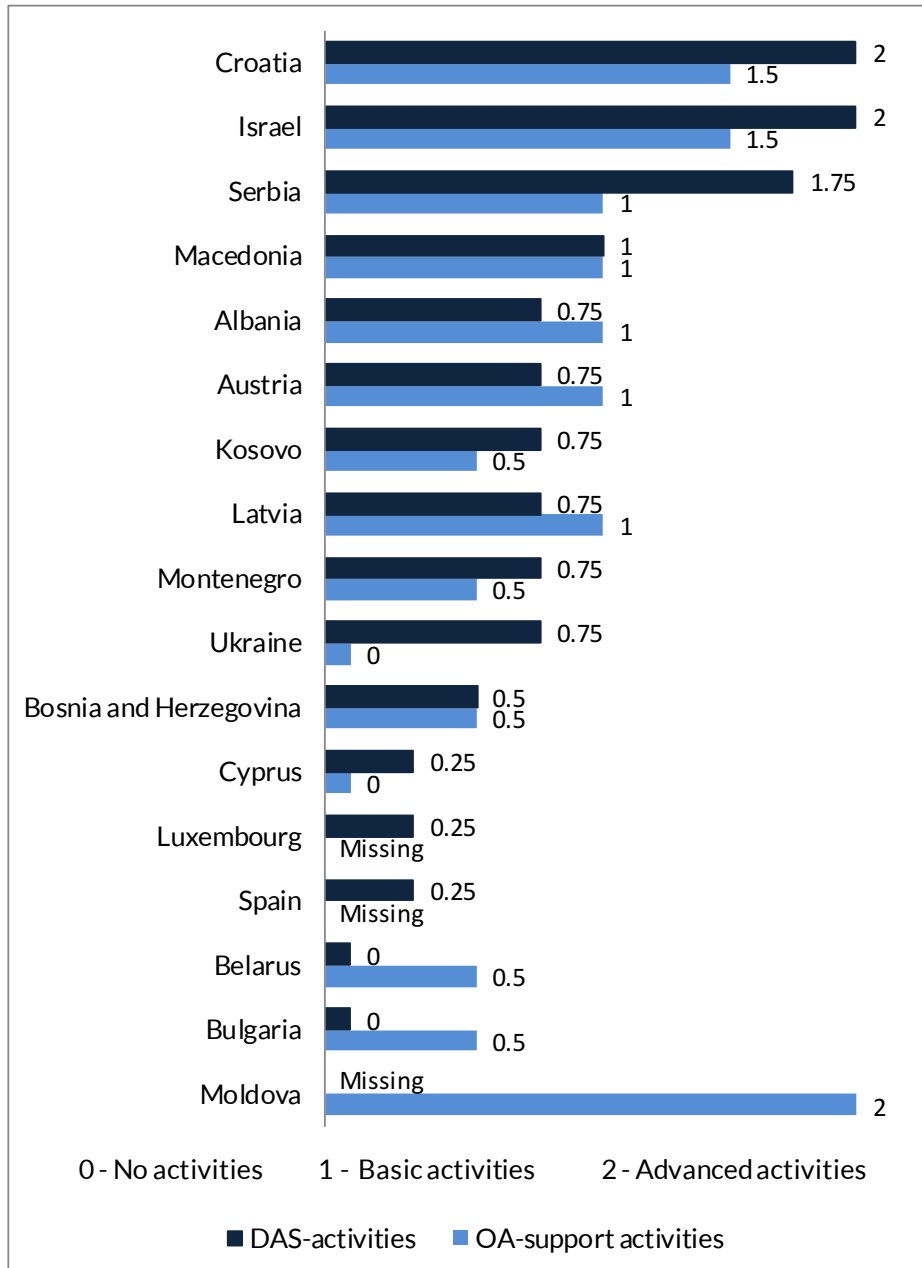
Looking further into each indicator, we first notice that most of the countries have no support services that facilitate data sharing and OA to research data in the social sciences, even though there are OA initiatives in almost every country. However, in nearly every country, respondents could name at least one potential hosting institution. Only few countries have advanced organisational activities, technical infrastructure, and required skills and training. The indicator with the greatest number of missing values relates to "Capacity building and training". This shows that the identification of currently available expertise initiatives or initiatives to develop appropriate knowledge and skills for a DAS is difficult.

Figure 25: Number of countries by indicator, according to their DAS activities' and OA support activities' implementation type*
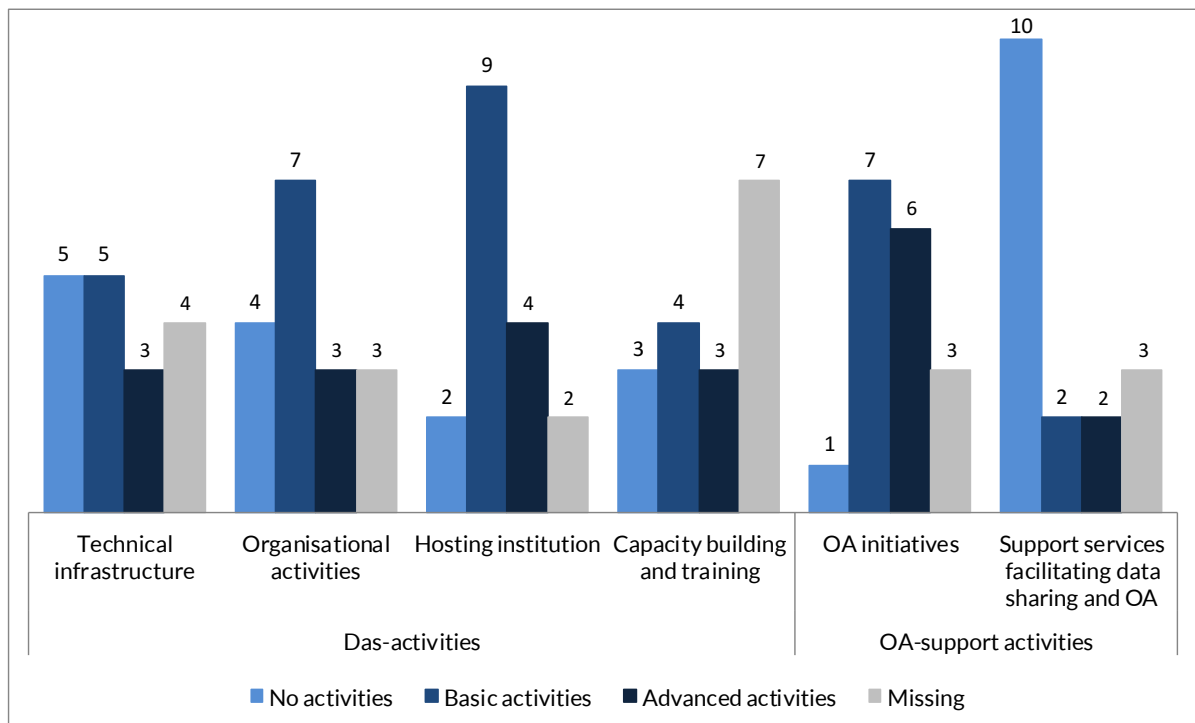
See questions 3.2.3, 4.1.1 to 4.1.4 and 4.2.1 in the Appendix 3, CESSDA SaW D3.2 Country report on development potentials 1

Figure 26 shows which kind of DAS activities are undertaken in each country and how developed they are: not at all (0), basic (1) or well under way (2). Missing information has been treated in this graph as *no activities* (0) to facilitate legibility. As seen before, Croatia, Israel and Serbia are the most advanced countries in each DAS activities indicator, while apparently no or very little activity is currently carried out in Belarus, Bulgaria, Cyprus, Luxembourg, Moldova and Spain.

Figure 26: Detailed DAS activities indicators (values from 0 - not at all, 1 - basic to 2 - well under way, by country
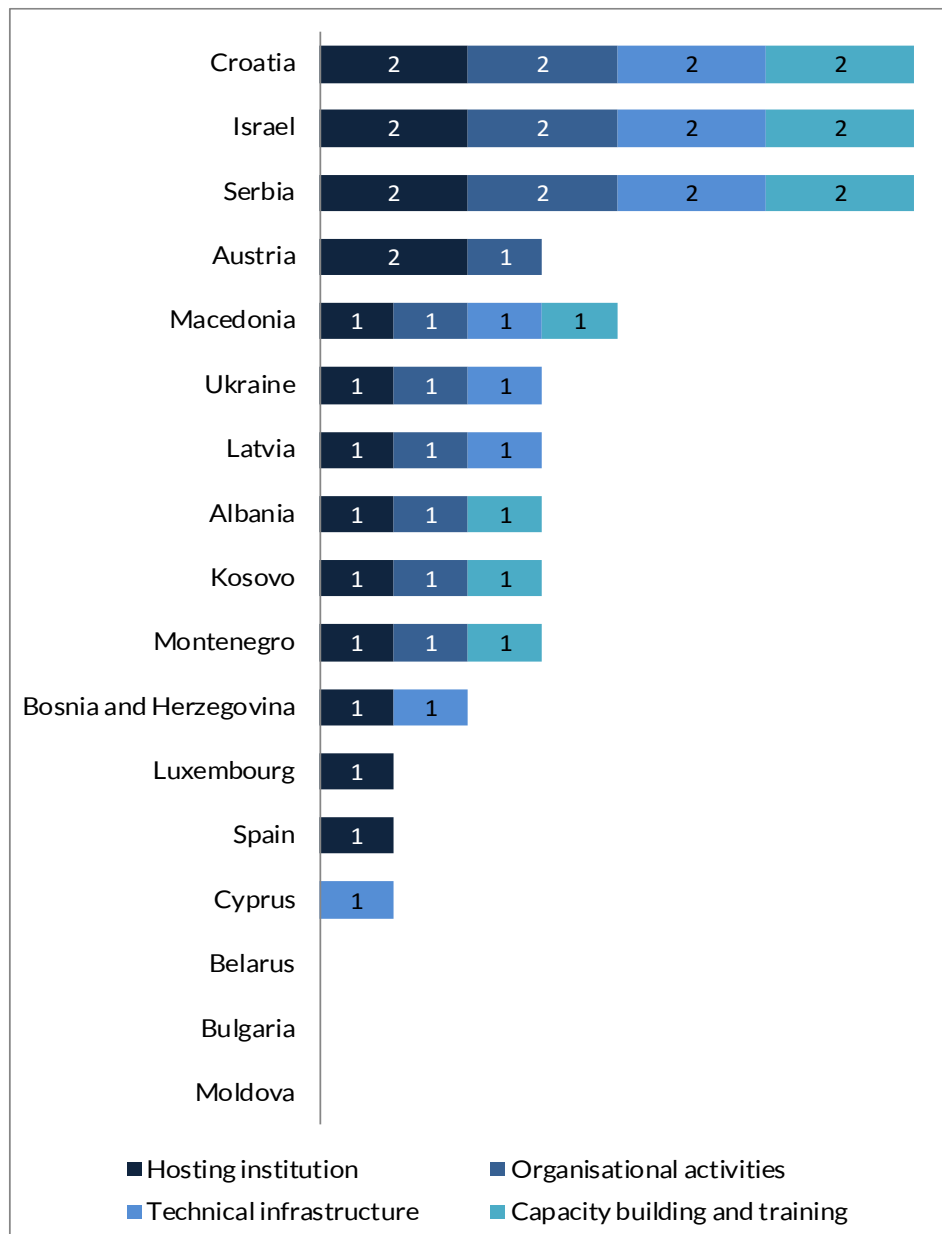
Figure 27 shows the OA support activities performed in each country and how developed they are: not at all (0), basic (1) or well under way (2). Missing information has also been treated in this graph as *no activities* (0) to facilitate legibility. As mentioned before, support services for data sharing and OA are almost never developed, except in Moldova, Croatia, Israel, and Serbia. The countries that provide little or no DAS activities are also the ones that provide little or no OA support activities, with the exception of Moldova.

Figure 27: Detailed results of OA support activities indicators, by country
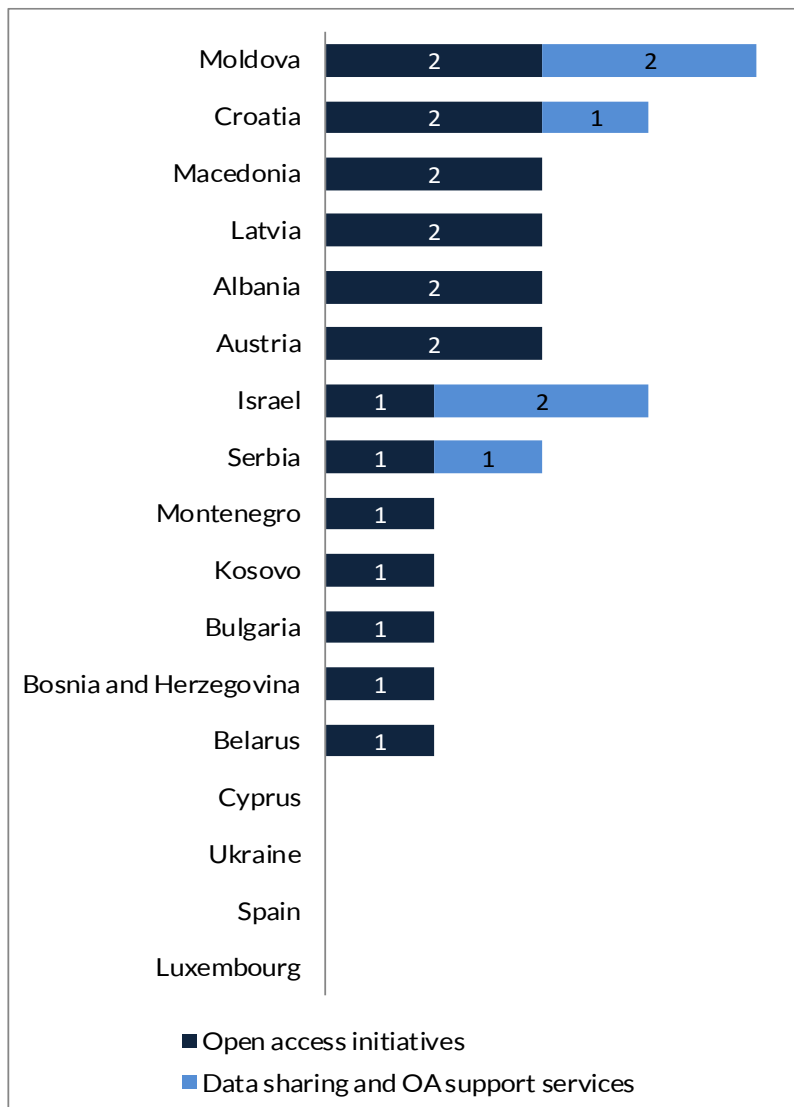
Generally, countries that earlier had been targeted by programmes aiming to develop activities and skills in order to establish social science data services are already much advanced. Programmes such as SERSCIDA (2012-2014, including Bosnia and Herzegovina, Croatia and Serbia) and SEEDS (2015-2017, including Albania, Croatia, Kosovo, Macedonia, Montenegro and Serbia) clearly has been successful in this respect. Detailed information about each national situation can be found in the Deliverable 3.2. of the CESSDA SaW project.

5.   Concluding remarks

Compromises were needed with respect to length of SaW T3.2 Web form collection instrument regarding topics included, that still could reflect the diversity of situations in different countries, and serves multipurpose – also for CESSDA to orient its activities in the long run. In particular, regarding the use of CESSDA-CDM it was decided to be pragmatic and to apply a simplified and shortened version of the model. In some of the sections, the self-assessment was difficult due to lack of information or hard to answer subjective evaluations. The interpretation of results should acknowledge these limitations.

Digital preservation activities are still not followed on the highest professional level in all of the organisations. Yet, we can expect this to improve in the following years. A dedicated CESSDA Trust group support is giving support there, and activities of different other tasks and work packages of the CESSDA SaW are addressing those issues as well.

In future development planning the differences among organisations in sustainability arrangements (Continuity of access) and technological features of digital preservation infrastructure need to be addressed among other things. Small organisations cannot afford professionalization and keeping track of the knowledge in various fields of the data services. CESSDA membership proves as one of the circumstances that lead to the improved sustainability of services. In future, more collaborative projects should be proposed that address current gaps that are present among groups of CESSDA members lagging from the top-quality service delivery. Analysis show that some of the elements of contemporary digital research data ecosystem, such as citations management, are discriminating the CESSDA members and aspiring DAS. CESSDA ERIC may use this observation in setting the priorities among areas, taking into account where the gaps are largest among more developed and less developed services, and where the common standardised solutions can be proposed and designed to apply to all DAS. The comprehensive metadata and persistent identifiers services that are being proposed by CESSDA Metadata Management portfolio 1.0[53] and Persistent identifier policy respectively, can as examples lead such development.

In the countries, where no formal DAS exist, we identified the key players involved in DAS-related activities, and described their current expertise, the level of technical infrastructure development, as well as their overall activities and potential regarding open science issues in general. Our review shows that those countries, in which interested institutions have been already taking part in international programmes and projects aiming to develop skills for establishing DAS, are already more advanced.

---

[53] See https://drive.google.com/file/d/0B2CMCTB5AE7Cak1WcDBGU0NmRWM/view.

List of Figures

List of Tables

Appendix 1: Review of existing studies for desk research

Title and bibliographic reference/ Link	Content area/ Type of countries / Which stakeholders are referred?/ Which disciplines?/ Relevance for task 3.2 / Other comments	Method (self-assessment, questionnaire, interviewing, official expert report, etc.) / Who is providing information?	Geographical scope (international, regional, country case study)/ Reference period
Van den Eynden, V. and Bishop, L. (2014). *Sowing the seed: Incentives and Motivations for Sharing Research Data, a researcher's perspective.* Knowledge Exchange.	data sharing practices in research, benefits of data sharing, influence of data policies, status of data policies, recommendations for stakeholders, multidisciplinary	qualitative interviews with researchers	Europe: UK, Denmark, Germany, Netherlands, Finland
Expert Advisory Group on Data Access (2014), Establishing incentives and changing cultures to support data access. http://wellcome.ac.uk/stellent/groups/corporatesite/@mshpeda/documents/webdocument/wtp056495.pdf	social and medical data; examine the extent to which the costs to researchers and research teams of providing access to datasets are balanced by the benefits (in terms of recognition, rewards and career advancement) that they receive. The goal was to identify possible areas in which additional incentives might be required, in order to foster a culture that supports researchers in making high quality data available and gives due recognition to those who do it well.	interviews with key stakeholders (research funders, senior academic managers, postdoctoral researchers, a chair of a Research Excellence Framework panel and a senior data manager) and a web survey with researchers and data managers	international
Sayogo, D.S. and Pardo, T.A. (2013), Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. Government Information Quarterly, 30(1): 19-31. doi 10.1016/j.giq.2012.06.011	multidisciplinary	survey of 1,329 researchers by DataONE	international

Tenopir, C., Allard, S., Douglass, K., Aydinoglu ,A.U,, Wu, L., Read, E., Manoff, M., and Frame, M. (2011), Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 6. doi:10.1371/journal.p one.0021101plosone. org/article/info:doi/10 .1371/journal.pone.00 21101	multidisciplinary; exploring current data sharing practices and perceptions of the barriers and enablers of data sharing.	survey with 1329 scientists	international
SERSCIDA (Support for Establishment of National/Regional Social Sciences Data Archives) (2012-2014): http://www.serscida.e u/en/deliverables	Analysis of existing potentials for the establishment of a social sciences digitised data service in **Bosnia and Herzegovina**; Analysis of existing potentials for the establishment of a social sciences digital data base archive in **Croatia**; Analysis of existing potentials for the establishment of a social sciences digital data base archive in **Serbia**	3 groups: researchers: *Surv ey on Gathering, Preserving and Use of Research Data*; *Survey on Infrastructure for Data Preservation*; *Policy, Legal, and Financing Provisions for Science*	international
Data Service Infrastructure for the Social Sciences and Humanities (DASISH, Deliverable: D4.3)	The goal has been to allow the selection and promotion of high-quality deposit services for researchers in the Social Sciences and Humanities (SSH) and to make suggestions for service improvements	A survey was sent to 89 persons working at existing and developing data archives services (DASs) in Europe. Additionally in-depth interviews have been conducted.	
CESSDA PPP: Summary of organisational expertise for activities set out in the ERIC workplan (D3.1b)	This document provides information from PPP partners who were all invited to express their interest in working in clusters or in taking responsibility for activities set out in the PPP work plan.	Descriptions of expertise are all self-assessed.	international
CESSDA PPP: WP6 Final report: Strengthening the CESSDA RI (D6.1)	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.	During the May and June of 2008 an online survey was conducted among CESSDA member	international

		archives regarding their modes of operations.	
CESSDA PPP: WP7 Final report: Widening CESSDA: Inclusiveness and Comprehensiveness of the upgraded European research infrastructure (D7.1)	This is the final report of work package 7 (WP7), which focused on plans for widening an upgraded CESSDA. The first part deals with strategic measures to bring in new national data resources into an upgraded Research Infrastructure. The second part focuses on strategies for the inclusion of data resources that currently reside in organisations and repositories outside of the existing CESSDA network. A list of relevant resources is provided and a specific programme for interoperating with identified agencies and organisations is recommended.	Methods of information collection: Country reports and workshops; Onsite visits;	Latvia, Lithuania, Estonia, Poland, Slovakia, Bulgaria, Croatia, Serbia, Macedonia, Belarus, Russia, Ukraine; Belgium and Portugal
DwB: Metadata Standards – usage and needs in NSIs and Data Archives (DELIVERABLE D7.1)	To get an overview on the current usage of metadata standards at European NSIs and DAs for the work in the DwB project, WP7 and WP8 have carried out surveys12 to collect information on the subject. A presentation and an analysis of the survey responses can be found in Chapter 6.	Two surveys: one Eurostat report on monitoring metadata systems at European NSIs from 2009 (see Section 6.1.1) and one survey conducted by WP8 in DwB (D8.3, see Section 6.1.2).	international
COAR Roadmap – Future Directions for Repository Interoperability. https://www.coar-repositories.org/files/Roadmap_final_formatted_20150203.pdf	Promoting greater visibility and application of research through global networks of Open Access repositories. Repositories, Open Access	Roadmap process, information, publishing and repository community	international
Parse.Insight Report Insight into digital preservation of research output in Europe (D3.4) - Survey report	This report (deliverable 3.4 of PARSE.Insight) describes the results of the surveys conducted by PARSE.Insight to gain insight into research in Europe. Major surveys were held within three stake-holder domains: research, publishing and data management. In total, almost 2,000 people responded; they provided us with interesting insights in the current state of affairs in digital preservation of digital	In total 1,840 people responded: 1,389 responses on the researchers' survey; the data managers' surveys yielded 273 responses, and the 178	international

	research data (including publications), the outlook of data preservation, data sharing, roles & responsibilities of stakeholders in research and funding of research.	publishers started the publishers survey. The funders survey gained only a few responses.	
Parse.Insight Report Insight into digital preservation of research output in Europe (D3.3) - Case studies report	This document summarises the results of the in-depth case studies on digital preservation in particular scientific disciplines carried out in the framework of the PARSE.Insight project. The case studies are intended as a supplement to the general surveys that were also conducted by the project. The investigations which are described provide new insight into the current state, needs and aspirations for long-term preservation of digital data in these disciplines, and contribute to the roadmap for the future infrastructure to support this aim.	The case studies addressed two disciplines in "hard sciences", High Energy Physics and Earth Observation, and two within social sciences and humanities, Psycholinguistics and Book Studies.	
The Research and Innovation Observatory (RIO) https://rio.jrc.ec.europa.eu/en. The latest available report is from the year 2014. The RIO Country Report 2014 builds on the experience of the ERAWATCH project.	Could be useful for identifying relevant policies and funding perspectives in EU Member States. Supports policy-making in Europe by providing practical information and a breakdown by Member State. The annual RIO Country Report analyses and assesses the development and performance of the national research and innovation system and related policies in the perspective of EU strategy and goals. The report also assesses the match between national policy priorities and the structural challenges of the research and innovation system. It includes chapter titled "National progress towards realisation of ERA", organized around ERA priorities one of which is "ERA priority 5: Optimal circulation and access to scientific knowledge" where e-infrastructures and OA to publications and data policies are described.	official reports / Horizon 2020 Policy Support Facility (PSF), developed jointly by DG Joint Research Centre (JRC) and DG Research & Innovation (RTD)	EU Member States
Strategic Priorities, Funding and Pan-European Co-operation for Research Infrastructures in Europe, Survey Report	In this study, Science Europe seeks to extend the understanding of the role of RIs in the ERA by documenting and analysing the decision-making processes of Science Europe Member Organisations, which underpin the decisions needed to build and	Survey questionnaire aimed to identify the situation at national, regional and organisational level answered by	international

	operate RIs at the national, European or global scale.	26 organisations from 19 countries	
Engelhardt, C., S. StrathmOverall, the survey received feedback from 44 countriesann, K. McDadden. (2012) Report and analysis of the survey of Training Needs. DigCurV; http://www.digcur-education.org/eng/Resources/Report-and-analysis-on-the-training-needs-survey	Could be used for assessment of human resources skills and needs, for how to provide training programme and capacity building, outsource service needs etc. This report describes the conception and execution of the survey on training needs in digital preservation and curation carried out in the context of the DigCurV project. It summarises the main findings of the survey together with those from a series of focus group meetings held in the partner countries and also an analysis of job advertisements which have appeared since January 2011 when the project began.	Survey among preservation staff in different types of organisations.	44 countries worldwide
A. White (2016) Final Results from the DCC RDM 2014 Survey	Could be used for organisational profile (staff number, activities current and planned). Several months ago we carried out our 2014 survey of senior managers in UK Higher Education Institutions who are involved in decision-making about their institution's Research Data Management services. - See more at: http://www.dcc.ac.uk/blog/rdm-2014-survey#sthash.PIjc0Y1f.dpuf	senior managers in UK Higher Education Institutions	UK

Appendix 2: Glossary

Authentication and Authorization Infrastructure (AAI): An AAI is an infrastructure to verify a user's identity (authentication) and to verify that a user has the rights to access the service the user has requested (authorisation) [DASISH].

Data access: The OAIS entity that contains the services and functions which make the archival information holdings and related services visible to Consumers.

Data Documentation Initiative (DDI): The DDI is an international standard for describing statistical and social science data. Documenting data with DDI facilitates interpretation and understanding, both by humans and computers [http://www.ddialliance.org/].

Data Management Plan (**DMP):** Data Management Plan is part of grant application or research project delivery that consider essential properties of RDM throughout the project, aiming at Open data as the default.

Data preservation: or more specifically, digital data preservation refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. This broad definition of data preservation refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technological change. Long-term preservation can be defined as the ability to provide continued access to digital materials, or at least to the information contained in them, indefinitely. (Source: IFDO, Data preservation, http://ifdo.org/wordpress/preservation/)

Data Seal of Approval (DSA): Self-assessment for trusted digital repositories. The DSA is granted to repositories that are committed to archiving and providing access to research data in a sustainable way.

Designated Community: An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time [OAIS].

Digital Object Identifier (DOI): A DOI name is an identifier (not a location) of an entity on digital networks. It provides a system for persistent and actionable identification and interoperable exchange of managed information on digital networks [www.doi.org].

GERD - Gross domestic Expenditure on Research and Development is total intramural expenditure on research and development performed on the national territory during a given period [https://stats.oecd.org/glossary/detail.asp?ID=1162].

Ingest: The preservation functional entity that contains the services and functions that accept data and metadata from data producers/depositors, prepares data and metadata for storage, and ensures that the information becomes established within the archive.

Intellectual Property Right (IPR): IPR is grouping or class of several different legal regimes that generally concerns creations of the human mind. Copyright is one of the legal regimes that fall under the umbrella of intellectual property. [http://corecopyright.org/2009/12/03/copyright_ip/]

Maturity Level: a Maturity Level estimates the level of maturity for an Activity. Maturity Levels are defined in a scale ranging from 0 to 5: N/A (0), Initial (1), Repeatable (2), Defined (3), Managed (4), and Optimised (5).

Open Archival Information System (OAIS): An Archive, consisting of an organization, which may be part of a larger organization, of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community. The term **OAIS** also refers, by extension, to the ISO OAIS Reference Model for *an* OAIS. This reference model is defined by recommendation CCSDS 650.0-B-1 of the Consultative Committee for Space Data Systems; this text is identical to ISO 14721:2003 which is superseded by ISO 14721:2012. It meets a set of responsibilities, as defined in section 4, that allows an OAIS Archive to be distinguished from other uses of the term 'Archive'. The term 'Open' in OAIS is used to imply that this Recommendation and future related Recommendations and standards are developed in open forums, and it does not imply that access to the Archive is unrestricted [OAIS].

Open data as the default: '(E)xpectation that all government data be published openly by default, (…) while recognizing that there are legitimate reasons why some data cannot be released.' **(See https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex)**

Persistent Identifier (PID): A maintainable identifier that allows one to refer to and have reliable access to a resource or object over long periods. A PID must always be resolvable through a resolution system [APARSEN].

Research Data Management (RDM): 'Research Data Management comprises the different components of the research data lifecycle, from data creation to data preserving, sharing and re-use.' 'RDM is an integral part of the wider research process, contributing the standards and principles of research, and applicable not just to the research data lifecycle, but throughout the lifecycle of research projects as a whole.' (See http://repository.jisc.ac.uk/6379/16/Training_for_RDM_-_Comparative_european_approaches_May_2016.pdf)