



CESSDA Strategy 2018-2022

Status: Public
Author: Ron Dekker, Director CESSDA
Date: 28 September 2018
Document: CESSDA Strategy 2018-2022
Version: Final



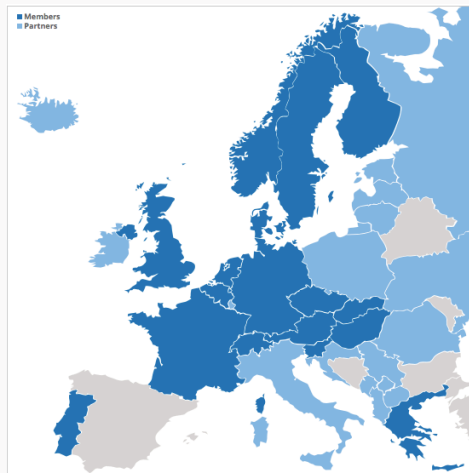
Table of Contents

1	Introduction – a new strategy	3
2	Background, Mission and Vision.....	4
2.1	Background on CESSDA	4
2.2	Mission.....	5
2.3	Vision.....	5
3	Environment and Stakeholder Analysis.....	7
3.1	SWOT.....	7
3.2	Stakeholder Analysis	11
4	Goals and Strategy.....	12
4.1	Strategic Goals.....	12
4.2	Strategic choices	13
5	Tactical and Operational Level.....	14
Annex 1	Establishing the strategy.....	16
Annex 2	Overview of CESSDA projects 2015-2018	17



1 Introduction – a new strategy

The Consortium of Social Science Data Archives (CESSDA) is a distributed research infrastructure in the social sciences domain. CESSDA has been on the ESFRI¹ Roadmap since 2006 and as an ESFRI Landmark since 2016. As of June 2017, CESSDA is an ERIC – a European Research Infrastructure Consortium – a European legal entity.



Member States of the European Union, associated countries, third countries other than associated countries, and intergovernmental organisations can join CESSDA as a member or an observer. Membership implies that a country assigns and supports a national service provider that will provide CESSDA services in their country and across Europe.

At the start of CESSDA ERIC in June 2017 there were 14 members and 1 observer: Austria, Belgium, the Czech Republic, Denmark, France, Germany, Greece, Hungary, the Netherlands, Norway, Slovakia, Slovenia, Sweden, Switzerland (observer) and the United Kingdom. By the end of 2017, Portugal and Finland had also become members.

This strategy covers the period 2018-2022 and should complete the transition from starting up the ERIC in 2017, towards a mature high-quality social science infrastructure, well embedded in the European Open Science Cloud, with CESSDA's Service Providers performing as trusted repositories, and heavily used by researchers and professionals in their roles of data producers and data users.

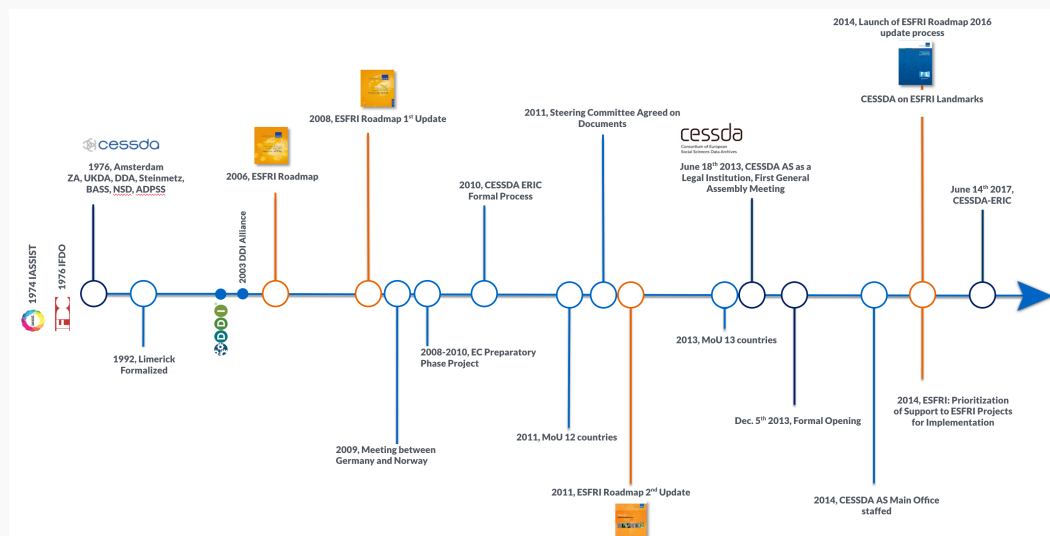
¹ European Strategy Forum on Research Infrastructures



2 Background, Mission and Vision

2.1 Background on CESSDA

The idea of establishing CESSDA dates to the mid-seventies. Seven social data archives started a Council of European Social Science Data Archives. In the beginning, it was an informal gathering of staff from the archives. CESSDA became more formalised in 1992 and reached ESFRI status in 2006.



The decision to become an ERIC was taken in 2010, backed by twelve countries. Three years later, in 2013, a legal entity was established with thirteen countries committed, and the Council became a Consortium. This change in name revealed that CESSDA became more structured. It also stressed the involvement of governments alongside the important role played by the national data service providers.

In 2016, CESSDA received the ESFRI Landmark status – reflecting the sustainability of CESSDA as a research infrastructure. In June 2017, the formal phase of becoming an ERIC was finalised. Its goal, governance and organisation were acknowledged by the European Commission. Reaching the ERIC status marked an important milestone: CESSDA was transformed from an informal gathering of data archives into a well-structured and sustainable European research infrastructure, with member countries agreeing on the goals of CESSDA, meeting obligations to establish and maintain national service providers and contributing financially to the consortium.



2.2 Mission

Member countries seek to increase the scientific excellence and efficacy of European research in the social sciences, as well as to expand easy access to data and metadata regardless of borders. They want to provide a research infrastructure for their researchers and join forces among their (national) data service providers.

The mission of CESSDA, as phrased in the statutes, is:

1. to provide a distributed and sustainable research infrastructure,
 - enabling the research community to conduct high-quality research in the social sciences,
 - contributing to the production of effective solutions to the major challenges facing society today;
2. to facilitate teaching and learning in the social sciences.

2.3 Vision

We believe and foresee that:

- Science will open.
This is good for science, for the economy and for society. Open Science was the main outcome of the European Commission consultation in 2014 and has become one of the major goals in science policy.
- The importance of data will grow.
In our information society, data is the new oil, but unlike oil, data are reusable, instant, perfect & free². Time Magazine calculated that in 2017 the value of data assets is bigger than the value of oil assets³.
- Platforms will become increasingly important for the economy and for society.
Platforms induce a shift to the demand side and stress that innovation happens in networks where companies and customers – or producers and users – meet. Instead of ownership and linear production modelling, the key feature will be value-creating interactions between producers & users⁴.

² Benkler (2006), *The Wealth of Networks*, Yale University Press; McAfee & Brynjolfsson (2017), *Machine, Platform, Crowd: Harnessing our Digital Future*, ISBN 978-0-393-25429-7.

³ Time Magazine, May 2017.

⁴ Geoffrey G. Parker, Marshall W. Van Alstyne, Sangeet Paul Choudary (2016), *Platform Revolution: How Networked Markets Are Transforming the Economy - and How to Make Them Work for You*, W.W. Norton & Company Inc., New York.



Hence, we live in an information society where science will open, the importance of data will grow, and platform economies will prevail.

In research policy, we notice an urge to better connect science with society and innovation⁵ and the European Commission (EC) is planning 'mission-oriented' research and innovation⁶ for the new Framework Programme⁷. Hence, as with platforms, we see a shift in European research policy to the demand side, which stresses a problem-oriented approach, attention to Sustainable Development Goals and grand challenges. This will require more interdisciplinary research to solve complex questions. It also implies that data need to be connected – either afterwards, by using new semantic techniques, or beforehand⁸.

Alongside surveys and other traditional research data collections, new data modes will arise, such as registries and social data. Registry data are combined with surveys and social data with geographical and health data. Hence, we are witnessing an amalgamation of data types and content. From a user's perspective, data need to be clustered around missions or big societal questions, like the sustainable development goals.

To increase the reuse of data, we must be able to ensure the quality of data, to transfer knowledge on these complex data, as well as to provide safe & secure access to sensitive data. This may imply that data remain where they are and that platforms are provided to securely access the data. Trust in the ecosystem is vital. This is important for science and society: to increase the scientific excellence and efficacy of European research in the social sciences and to understand the major challenges facing society today.

Our vision is that the provision of access to social science data and metadata is vital – for both science and society. For this we must offer services to data producers to easily describe and store their data – if needed in a secured environment. We will adhere to the FAIR data principles to make data findable and provide information about the data, where they are and how they can be

⁵ Directorate-General for Research and Innovation (European Commission) (2017), Europe's Future: Open Innovation, Open Science, Open to the World, reflections of the Research, Innovation and Science Policy Experts (RISE) High Level Group, DOI 10.2777/79895.

⁶ Directorate-General for Research and Innovation (European Commission) (2018), Mission-Oriented Research & Innovation in the European Union - A problem-solving approach to fuel innovation-led growth, DOI 10.2777/360325.

⁷ EC (2018), Horizon Europe, https://ec.europa.eu/info/designing-next-research-and-innovation-framework-programme/what-shapes-next-framework-programme_en

⁸ Cf. large international surveys like the European Social Survey, European Values Study, SHARE - Survey of Health, Ageing and Retirement in Europe, and GGP, Generations & Gender Programme.



accessed. We will focus on providing training and other modes to transfer expertise and share knowledge about the data, the rules and regulations.

CESSDA wants to be a consortium of trusted repositories with full European coverage, offering a platform with tools and services to both data producers and data re-users. It wants to be part of the European Open Science Cloud – especially on social science data. By 2022, CESSDA should be leading in data discovery.

3 Environment and Stakeholder Analysis

What is happening in our environment, what are the circumstances and trends that CESSDA is confronted with?

3.1 SWOT

A Strengths-Weaknesses-Opportunities-Threats (SWOT) analysis describes the environment and opportunities of CESSDA. The strengths and weaknesses focus on the current situation and position of CESSDA, while the opportunities and threats describe the future environment of CESSDA.

Strengths	Weaknesses
<ul style="list-style-type: none"> • Political attention <ul style="list-style-type: none"> ◦ European Open Science Agenda⁹ ◦ National and EC / ESFRI Roadmaps • Data Management <ul style="list-style-type: none"> ◦ Protocols by EC & Research Councils ◦ Continuous work on certification • Expertise & facilities by Service Providers • Wide range of cooperation: <ul style="list-style-type: none"> ◦ support to new and smaller archives within the consortium; 	<ul style="list-style-type: none"> • Lack of incentives to share data • Hardly any viable business models for research data services • Lack of EU Data Industry & Services • Lack of human capacity (data experts) • Lack of standards, incl. implementation • Scattered landscape <ul style="list-style-type: none"> ◦ Many players in the research infrastructure landscape ◦ Pipelines instead of Platforms ◦ Fragmentation in public sector

⁹ Important features are: Expert Groups on EOSC and FAIR data, Open Science Policy Platform, FAIR data principles, Grant Regulations on Open Science, OpenAire, Foster and other EC projects on Open Science.



<p>advancement of more mature ones</p> <ul style="list-style-type: none"> ○ extensive contacts with potential partners; support through the membership acquiring process; support to aspiring archives 	
<p>Opportunities</p> <ul style="list-style-type: none"> • Growing importance of data <ul style="list-style-type: none"> ○ Data-driven, reproducible data ○ Better connecting science and society ○ Mission-oriented science & innovation • Need for data content <ul style="list-style-type: none"> ○ Combining data from multiple disciplines ○ Linking and connecting data, ○ Big Data & Artificial Intelligence. • Platform economies <ul style="list-style-type: none"> ○ Focus on value-creating interactions between producers & users ○ Shift from owning to sharing and stimulating interactions 	<p>Threats</p> <ul style="list-style-type: none"> • Complexity in Legislation • Increasing complexity of data, new types of data • New competitors <ul style="list-style-type: none"> ○ New data providers – next to national service providers ○ Commercial companies taking over – making Service Providers superfluous • Security conditions for data <ul style="list-style-type: none"> ○ Legal issues ○ Need for trust ○ Data must stay where they are

In this SWOT-analysis we see the shift towards the demand side as indicated by platform economies: more focus on re-use of data, including integration or combination of data. Data services providers must provide trust on this platform – towards data producers in making their data shareable, and towards users to ensure quality and relevance of the data. Due to legislation, type of data and security, the data will stay where they are, and users can access these data via secured platforms. Complexity, size, and variety of data ask for extensive training on describing as well as using data. Ultimately, sharing and reusing data will go beyond science: professionals, citizen scientists and other users will also seek access to relevant data.

Explanation of the SWOT-items

There is a political tailwind on open science, including attention to optimal reuse of research data. The EC set up an Open Science Agenda and has made its European Open Science Cloud a priority.



Research funders and governments have developed – or are developing – research infrastructure roadmaps. The ESFRI-process contributed significantly to developing policies on research infrastructures.

Awareness about reuse is increased as more funders follow the Wellcome Trust and EC policy to make data management plans obligatory for receiving grants. Still, one of the major barriers for sharing data is not resolved: a rewards and incentives system for researchers that share their data. Consequently, many researchers regard ‘their’ data as assets and inputs for publishing. It is beyond the scope of CESSDA to change this incentive system for researchers to share their data. However, CESSDA can take away barriers and create the necessary conditions for data sharing, e.g. on providing persistent identifiers for data citation and support initiatives that connect researchers, articles, data and other research data output.

We see a growing importance of data in information society and growing technological support to reuse data. There are, however, many players, resulting in fragmentation and risks of reinventing the wheel. Research infrastructures face difficulties in retaining sustainability as they depend on grants and long-term funding¹⁰. An underlying complication is that the one who pays for data collection or invests time in preparing the data is different from the ones who will reuse and benefit from these data. In research, paying for access to existing data is a no-go, although there seems to be room to develop business models for data services. A threat is that big conglomerates will offer data services for free to attract and bind researchers to their platform and pay for (publishing) services later in the research process. On data complexity and legal issues, we should help data producers and data users. For example, by providing explanations, sharing expertise and providing data-discovery training on thematic data clusters.

In Europe, we do not have a viable data industry (cf. USA with Amazon, IBM, Microsoft, Google and Facebook as major players). The current landscape of e-infrastructures (storage, networks, computing) still tends to be in silos, whereas many policy studies hint at e-infrastructures as a service to researchers and other professional users. We do have standards for technology and metadata, but implementation is sometimes lacking, thus obstructing interoperability, as well as describing and exchanging data.

CESSDA Service Providers have almost fifty years of experience in data archiving and providing services to data producers and data users. A key focus during this

¹⁰ OECD Global Science Forum (2017), Business models for sustainable research data repositories, DSTI/STP/GSF (2017)1/FINAL.



period was reliability and ensuring sustainability. Joining forces within CESSDA creates a critical mass and offers the sharing of expertise. Based on this legacy and strengthening cooperation, CESSDA's core competences are:

- **Expertise**
Expertise in data management, curation and archiving to preserve the initial investment in collecting data; expertise in standardising metadata, providing tools & services for the reuse of data.
- **Trust**
CESSDA is a network of trusted repositories and has ongoing efforts to raise trustworthiness – using internal quality assessment – and improve data processes within its Service Providers.
- **Cooperation**
CESSDA focuses on strong international cooperation through pan-European membership and a wide network of partners in Europe and beyond – e.g. CESSDA is a member of the Research Data Alliance and participates in the ERIC-Forum.

The shift in the data world to the demand side is reflected in the importance of the FAIR data principles, stressing the relevance of a data catalogue for findability, easy access – even to sensitive data, interoperability and linking of data for societal challenges. Plans by the EC for the new framework programme Horizon Europe focus on the demand side by setting missions for research – such as the UN Sustainable Development Goals. This demand-orientation implies more focus on the reuse of data, and on creating a match between supply and demand. Due to legal requirements, security, and the sheer size of data, the data will stay where they are – decentralised, yet findable via a central catalogue. This emphasises the need for platforms where users can access sensitive data in a secured way.

Key features of platforms are full coverage and trust. Hence CESSDA should continue to strive for full European coverage, make data easily findable and available, and share its knowledge and expertise. Given the sensitivity of social data we need to provide secure and safe ways to use these data. Interactions – i.e. reuse of data – will only flourish if there is trust between the data producer and the user. If neither the data user nor data producer have access to all the information about the data, this will hamper their interaction. Here the role of CESSDA national Service Providers as intermediate organisations becomes important¹¹.

¹¹ Cf. Airbnb, Uber providing quality checks on supply and demand, taking away risks by offering home insurance, fixed pricing, ratings, etc. If there is lack of trust, participants will leave the platform – cf. reputation damage for Facebook in the Cambridge Analytics case.



Irrespective of the system and its technology, it will be the human factor that is decisive for success. For the reuse of data, new users need to be trained, and data producers must be serviced when they describe and provide their data.

3.2 Stakeholder Analysis

The main stakeholders targeted by CESSDA are:

- Members
- Service Providers
- Data Producers
- Data Users

Members: we focus on Ministries and Research Councils who fund the consortium and establish national service providers, complemented by the European Commission.

Service Providers: we focus on the national service providers of CESSDA members. This could be expanded in a later phase to service providers from non-members and commercial and non-commercial providers of social science tools and services.

The two groups of Data Producers and Data Users can become complex. CESSDA is aware of this and will focus on both groups of researchers at universities and research institutes. At a later stage, CESSDA could expand these groups and consider including professionals and citizen scientists as data re-users, and national statistical offices, private companies, governments as data producers.

Value Propositions¹²

To Members

- Increase the scientific excellence and efficacy of research in the social sciences
 - Expand easy access to data and metadata regardless of borders
 - Train researchers and data experts
- Provide a European research infrastructure for sharing social science data
 - Increase the reuse of these data
 - Join forces and share expertise and technology among data service providers
 - Participate in the social science data part of the European Open Science Cloud.

To Service Providers

¹² Based on Simone Cicero, Platform Design Toolkit, <http://platformdesign toolkit.com>.



- Share their expertise
- Be more efficient and provide better services by
 - Setting up a common technology
 - Developing and using standards
 - Eligibility and better access to European funding.

To Data Producers

- Easy & safe deposit of research data
- Providing visibility for sharing data and credits on reuse
- Compliance with funder requirements – on GDPR and Research Data Management
- Training – esp. on data management and sharing data.

To Data Users

- A common data catalogue to find and access relevant data (CESSDA Data Catalogue)
- Secured access to sensitive data
- Tools and projects to make data Interoperable and easy to Reuse
- Training and sharing expertise on reusing data.

4 Goals and Strategy

4.1 Strategic Goals

In Europe, we see today a tendency to integrate. This is reflected in open science policy, in data and other research output, in infrastructures such as the ESFRI-clusters, and in the EOSC, with an integrated catalogue of services, software and data core. These integration processes require trust from all stakeholders in the ecosystem of data, infrastructures, services, research, innovation, etc.

CESSDA's goal is to increase the reuse of data, or to quote the European Commission¹³: the goal is "moving from the current fragmentation to a situation where data is easy to store, find and re-use". Reuse requires transparency and quality of data and research processes, but also security and to be in control on what's happening with the data.

In our strategy, we will align with the European Open Science Agenda, especially on the European Open Science Cloud, FAIR data principles and on skills. CESSDA

¹³ EC Staff Working Document 2018 (83) p. 1.



wants to provide seamless access to social science data to researchers. To do so, CESSDA aims to engage and enhance existing assets (people, knowledge, awareness) to make open science a default. In short, CESSDA wants to realise a platform for social data that is part of the European Open Science Cloud.

4.2 Strategic choices

In the previous strategy, the emphasis was on establishing the governance model and preparing the technological backbone of CESSDA. Technological support remains important, though the current strategy focusses on tools & services, training and on providing trust in CESSDA.

To achieve our goals, we will work along four strategic lines:

1. Technology: Platform & Cloud
 - building the platform infrastructure on which service providers can offer tools and services covering the whole research data life cycle
 - fostering and supporting standards on both technology and (meta) data
2. Tools & Services: as a service for data producers and data re-users
 - providing tools & services for data producers to curate & publish their data
 - providing tools & services for data re-users to find, access & reuse research data and metadata according to the FAIR principles and regardless of borders
 - with a flanking policy on software adoption procedures, on technology risk analysis, and on ways for e-infrastructure services.
3. Trust: Scope & Quality of Service Providers
 - striving for full European coverage
 - sharing expertise among service providers (strengthening)
 - performing quality assessments of CESSDA national service providers, amongst others, by elaborating a Quality Assessment and KPIs for CESSDA and its service providers.
4. Training: Awareness on Open Science and Increasing Data Skills of Researchers
 - providing training & educational materials for strategic partners and service providers to serve the research community throughout the whole research life cycle.
 - as there are many providers at a local and national level, the niche for CESSDA is expert-level training and provide training materials for other trainers. Training activities for individual researchers can be offered by national service providers, research libraries and other third parties.



For each of these four strategic lines, we have Working Groups led by and with participants from CESSDA's service providers. The service providers remain at the core of CESSDA and will contribute to activities along these strategic lines, while the CESSDA Main Office will coordinate membership, acquisition, the portfolio of services, standards and the technology for the common platform.

CESSDA should not intervene or duplicate activities carried out by national service providers or other parties in the EOSC ecosystem. To increase efficiency and efficacy, CESSDA should look for global partners and cooperation with third parties, that is: *be cooperative instead of competitive*.

Examples of strategic cooperation are:

- on training with FOSTER+, OpenAIRE;
- on e-infrastructures with the EOSC-hub and EGI (European Grid Initiative);
- on CESSDA projects with RDA (Research Data Alliance);
- on secured access with EUROSTAT, ILO, and National Statistical Institutes;
- on fostering social sciences with EASSH, EUA, LERU, CESAER;
- on sustainability of infrastructures with EC, ESFRI-clusters, ERIC-forum – especially within the Social Sciences and Humanities cluster;
- on metadata and strategic development with ICPSR and other global social science data infrastructures;
- on developing tools & services with YEAR (Young European Associated Researchers) and thematic data sets (e.g. Migration, Election Studies, etc.).

5 Tactical and Operational Level

CESSDA will work along these four strategic lines to develop and maintain a platform for data tools & services for our stakeholders, ensure trust in this platform by stimulating and guarding the quality of its service providers, strive for full European coverage, increase awareness on open science, and provide training facilities and content. Project deliverables and activities must be SMART – specific, measurable, achievable, relevant and time-bound.

Based on our strategy and the projects of the last three years, the following tools & services are available or must become available within 2-3 years.



	TOOLS & SERVICES	Data Users	Data Producers	Service Providers ¹⁴	Members
Available	CESSDA Data Catalogue	X			
	CESSDA Metadata Harvester			X	
	CESSDA Metadata Profile			X	
	CESSDA RDM Training Module		X		
	CESSDA Knowledge Exchange (Intranet)			X	
2018	Social Sciences Common Vocabulary Manager			X	
	Social Sciences Multilingual Thesaurus			X	
	Social Sciences Multilingual Question Bank	X	X		
2019	CESSDA Quality Assurance				X
	CESSDA Self-Archiving		X		
	Single Sign On	X	X	X	
	CESSDA Access to Sensitive Data	X			
2020	Social Sciences Data Discovery Tools	X			
	Social Sciences Research Data Management		X		
	CESSDA Data Tags		X	X	

To help realise this strategy, CESSDA will initiate and finance projects that are aligned with this strategy and its goals. CESSDA has an annual Work Plan that comprises approximately 10-12 projects. In addition, CESSDA will submit proposals to the European Commission and other research funding organisations.

¹⁴ This may include Third Party Developers.



Annex 1 Establishing the strategy

The discussion on this new strategy started in mid-2017 at the CESSDA Main Office in Bergen. The period covered by the previous strategy was coming to an end and new direction was needed.

The mission of CESSDA, trends in research data infrastructures, and a first draft of goals and strategic lines was set up based on the previous strategy.

A first draft was discussed at the Service Providers' Forum in October 2017. An important result was to extend the existing strategic pillars (Technology, Trust, Training) and add a fourth pillar: Tools & Services.

In November 2017, the General Assembly discussed the updated concept. Major remarks by the General Assembly were to:

- connect with other research infrastructures,
- pay attention to the human factor (acceptance, awareness, training),
- clarify the added value and impact of CESSDA,
- provide a high quality of services and high level of trust,
- create awareness on diminishing boundaries in research, e.g. between data management and analysis and between disciplines,
- increase efforts in widening (European coverage) and in sharing expertise.

In March 2018, the CESSDA Director met with the Scientific Advisory Board. First this board stressed that a strategy is about setting priorities and bringing focus to activities – hence, that a strategy is about making choices. CESSDA's role is twofold; what should be done at the level of consortium, and what could be done by partners or service providers at national level. Hence, CESSDA's consortium role is about economies of scale, providing critical mass, setting standards, fostering innovation, solving common (IT, legal) problems, offering tools & services and data access on a transnational scale.

In April 2018, a full concept text was presented to the Service Providers' Forum, and in early June at the CESSDA Widening Meeting with participants from affiliated (non-member) data service providers and government representatives. The final version was discussed at the General Assembly of 21 June 2018.



Annex 2 Overview of CESSDA projects 2015-2018

	Projects	2015	2016	2017	2018
Policy	Data Access Policy	X			
	PID Policy		X	X	X
Technology (incl. back office tools)	Technical Working Group		X	X	X
	Technological Framework	X	X	X	X
	Meta Data Harvester	X			
	CESSDA Metadata Manager	X	X	X	X
	Multilingual Thesaurus (incl. ELLST)		X	X	X
	Common Vocabulary Manager			X	X
Training	Training Working Group		X	X	X
	CESSDA Expert Seminars	X	X	X	X
	Data Discovery Training		X	X	X
	CESSDA RDM Training Module		X	X	
	Internal Knowledge Platform			X	X
Trust	Trust Working Group		X	X	X
	Widening Activities		X	X	X
	Quality Assurance			X	X
	GDPR Code of Conduct				X
	Access to Sensitive Data				X
Tools & Services	Tools & Services Working Group				X
	CESSDA Data Catalogue		X	X	X
	CESSDA Euro Question Bank	X	X	X	X
	CESSDA Self-Archiving (DataVerseEU)				X



The table includes Policy Studies that were carried out either as pathfinders for projects or to set CESSDA policies.

Working Groups are established and ended by the CESSDA Director. These groups cover long term activities, with participation and coordination by service providers' staff.

In this new strategy, we wish to strengthen the role of the CESSDA Working Groups within CESSDA's ongoing projects. There are regular virtual meetings of the Working Group Leaders and the CESSDA Chief Operations Officer, who is formally responsible for project deliverables within the set time and budget.