CESSDA Trust & CTS & FAIR

How to reach? How can we help?

Janez Štebe
CESSDA Trust WG

CESSDA Widening
Skopje 2019

November 5 – 6, 2019

cessda.eu

@CESSDA_Data

cessda

Slovenian Social Science Data Archives (*ADP-Arhiv Družboslovnih Podatkov*)

- Founded in 1997
- Slovenian national research data centre for social sciences
- CESSDA ERIC national service provider
- obtained CoreTrustSeal in beginning of 2018
- involved in different EU, CESSDA and national projects

CORE TRUST SEAL

ADP Social Science Data Archives
http://www.adp.fdv.uni-lj.si/eng/
CTS Certification 2017-2019

Experiences to share in acquiring the CTS

- The archive has been operational for 15 years: expertise gained through collaboration and involvement in CESSDA activities
- Role model: ICPSR, UK DA, DANS, →but then adapted to the size and specificity of the setting: here FSD model helped
- Active at digital preservation expert community nationally*:*
  - *e.g. Lecture at an event "Družbeni in gospodarski vidiki uporabe digitalizirane kulturne dediščine v Sloveniji (posvetovanje), 16. in 17. maj 2019", NUK, Ljubljana (2019)* https://www.adp.fdv.uni-lj.si/publikacije_adp/publikacija/334/
- Institutionalisation of organisational setting: few articles introduced in the Rules of the Faculty of Social Sciences, University of Ljubljana‘ (Host organisation of a national data service ADP)

Key:

- Sustainability guarantee through institutional setting, financial long-term stability of funding, and national membership in CESSDA ERIC infrastructure.

- Learn to speak OAIS language: pre-SIP, SIP, AIP, DIP

- Following the designated community definition in description of processes

- Keeping and regularly updating the written workflow data processing steps

- It‘s more about the processes and roles than about technology: after first submission the only comment was that we need to nominate who are holding the roles: https://www.adp.fdv.uni-lj.si/kontakt/

Trust - CESSDA

https://www.cessda.eu/About/Working-Groups/Trust

Išči

& EVENTS CONTACT TOOLS & SERVICES TRAINING DATA CATALOGUE

CESSDA TRUST Group

The CESSDA Trust Group offers both existing and aspiring service providers guidance and supp meeting a range of issues and standards relating to trusted data and services. CESSDA require adoption of specific criteria such as the internal obligations required from all members (CESS statutes) and the trustworthy digital repository (TDR) requirements set by the CoreTrustSeal.

These goals must be met within the evolving infrastructure (skills, services and technology) of European and international research data science.

CESSDA Trust Group consists of a core of service providers with experience in trust standards a certification and key contacts representing each of the CESSDA members and aspiring membe

The group's goals are met through:

- Guidance, engagement and support to members in understanding, acquiring and maintaining compliance with CESSDA obligations and the requirements of the CoreTrustSeal.
- Monitoring and reviewing compliance at an individual and organisational maturity leve Engaging with trust-related elements of the CESSDA work plan including other working groups and projects.
- Maintaining an overview of the trust landscape including certification standards and th emergence of the FAIR data principles and the requirements of the European Open Scie Cloud (EOSC).

RECENT NEWS

Some Trust WG areas of work

Evidence Alignment

- cases where the CESSDA Service Providers might cooperate on the alignment of evidence for the CoreTrustSeal.

Wider Social Science and Repository Landscape

- European Open Science Cloud (EOSC) and FAIR (Findable, Accessible, Interoperable, Reusable) data work.
- Key areas include EOSC-hub, FAIRsFAIR. FREYA (PIDs) and SSHOC.
- alignment between FAIR and OAIS/TDR.

Trust Status Monitoring

- Of those repositories which have not yet achieved CoreTrustSeal, AUSSDA is in progress, Belgium, Denmark and the UK plan submissions for 2019.

- Slovakia plans to make progress in 2019 and Portugal is actively engaged with the Trust Group but does not yet have sufficient human resources to define a timeframe for application.

- Serbia, which joined CESSDA recently has not yet initiated a self-assessment process.

- The CESSDA Working Group has started to initiate contact with new members North Macedonia and Croatia.

CESSDA Trust Group

Hervé L'Hours (UK Data Service)
Mari Kleemola (FSD)
René van Horik (DANS)
Maja Dolinar, Janez Štebe (ADP)
Jonas Recker (GESIS)
Birger Jerlehag (SND)

CESSDA Trust: CoreTrustSeal

- Self-assessment
- Internal Peer-review
- Comments and recommendations
- Repeat as necessary
- Apply for CoreTrustSeal

CESSDA Trust: CoreTrustSeal

- Questions about Context
- 16 Requirements
- Additional Guidance
- Glossary

Workshops, reports

- Trust workshop 1,2, Bergen and Cologne 2013
- CESSDA **Expert Seminar 2015, The Hague**
- Zagreb SEEDS Workshop, 2017
- SaW Workshops, the Hague and Zagreb, 2016, 2017
- Trust Workshops, Milan, Ljubljana, Berlin, 2018
- Trust WP Workshops, Paris, Skopje 2019

Current events November 2019

- November 13:
  - Webinar-Trusted-Repository-Certification-and-ICPSR
  - WORKSHOP: „CERTIFICATION WORKSHOP ON FAIR-ALIGNED REPOSITORIES IN AUSTRIA“
- November 26, Cologne
  - CESSDA Trust Workshop

Resources for preparing for CTS

- SaW policy alignment iteration: CESSDA SaW D4.3: Report Overview of Data Management Policies in Social Science Data Archives

- Existing applications as a source: alignment happens by borrowing elements from each other

- B. Lavoie. (2014). The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2ndEdition), Digital Preservation Coalition, [Online]. Available: https://www.dpconline. org/docs/technology-watch-reports/1359dpctw14-02/file

- Core Trustworthy Data Repositories Requirements: Glossary https://drive.google.com/file/d/0B4qnUFYMgSc-REpsNVQwWDVfSkU/view

EYE ON CORE TRUST SEAL:

Recommendations for Criterion R0 from Digital Preservation and Research Data Management Perspectives (Michelle Lindlar, Pia Rudnik): iPres2019_paper_143.pdf

Criterion R0 From CTS:

Designated Community

Repository Type:

- Domain or subject-based repository
- Institutional repository
- National repository system, including governmental
- Publication repository
- Library/Museum/Archives
- Research project repository
- Other (Please describe)

Lindlar and Rudnik findings

- Based on analysis of 40 CTS Assesment reports available on January 2019
- Data is available at:
  - Lindlar, Michelle, & Rudnik, Pia. (2019). Eye on Core Trust Seal - Data Set (Version 1.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3267690

Definitions of the terms

- Archive: the terms "archive", "data centre", and "service provider" refer to any organisations "...that intends to preserve information for access and use by a Designated Community" for the long term,

- where the „Designated Community“ is an identified group of potential consumers, or users, who should be able to understand a particular set of information

- based on Knowledge base.

Repository? Is not equal to ‚archive‘

Narrow or broad, Disciplinary – specific or general

OAIS concept not elaborated in CTS Glossary

0. Context

Level of Curation Performed

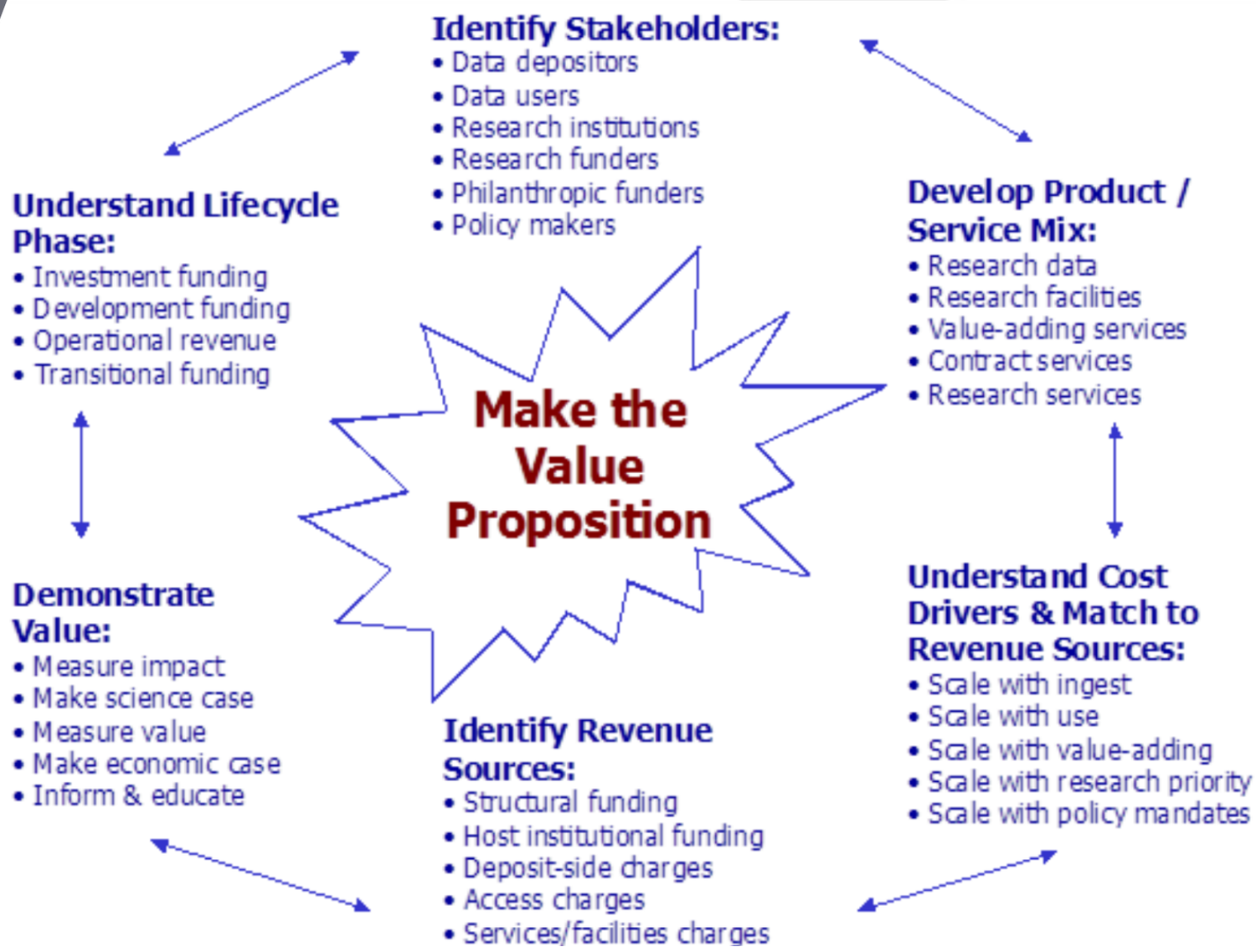
- A. Content distributed as deposited
- B. Basic curation – e.g., brief checking, addition of basic metadata or documentation
- C. Enhanced curation – e.g., conversion to new formats, enhancement of documentation
- D. Data-level curation – as in C above, but with additional editing of deposited data for accuracy

Lack of guidance on (Lindlar and Rudnik, 2019)

- How the two communities (Digital Preservation and Research Data Management) understand data lifecycle activities ,conversion to new formats: digital preservation vs. Curation: normalisation vs. Migration
- Conditions for Levels Applied: depositor agreement, external requirement/funding, technical suitability

RDA resource to address the issue of funding and sustainability

OECD Global Science Forum 2017, RDA - World Data System (WDS) collaboration on survey of data repositories:
BUSINESS MODELS FOR SUSTAINABLE RESEARCH DATA REPOSITORIES
http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/STP/GSF(2017)1/FINAL&docLanguage=En



„not all research data can or should be made broadly available. „ (OECD 2017)

„Many datasets are not of requisite quality, are not adequately documented or organised, or are of insufficient (or no) interest for use by others.„ (OECD 2017)

„The reality is that research institutions frequently have no idea how many datasets are held in ad hoc systems or how they are preserved. Many if not most of the digital data created or used in research over the last century have been lost because no long-term repository or other safeguards existed.„ (OECD 2017)

CTS and FAIR principles combined vs. COSTS

CTS Appraisal R8.
The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users. * Does the repository use a collection development policy? * Does the repository have quality control checks to ensure the completeness and understandability of data? * Does the repository have procedures in place to determine that the metadata required to interpret and use the data are provided? …

RDA FAIR data maturity model WG Collaborative sheet (Draft):
R1. meta(data) are richly described with a plurality of accurate and relevant attributes

- R1.1. (meta)data are released with a clear and accessible data usage licence
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards

RDA resources

- **RDA/WDS Certification of Digital Repositories IG**
- **Social Science Research Data IG just established**
- **FAIR Data Maturity Model WG (integration effort)**
- **FAIRSharing Registry: connecting data policies, standards & databases WG**
- **RDA for the social sciences -** Ricarda Braukmann from Data Archiving and Networked Services (DANS) - the national RDA node for the Netherlands and RDA ambassador for the social sciences
- **The RDA CoreTrustSeal adoption story across domains and regions**

Questions?

Comments?!

Sugesstions!

- Thank you!

University of Ljubljana

Faculty of Social Sciences

Social Science Data Archive

Kardeljeva ploščad 5

1000 Ljubljana

Slovenia

www.adp.fdv.uni-lj.si

arhiv.podatkov@fdv.uni-lj.si

Arhiv.Druzboslovnih.Podatkov

@ArhivPodatkov

Univerza v Ljubljani

cessda eric

SSHOC social sciences & humanities open cloud

seriss SYNERGIES FOR EUROPE'S RESEARCH INFRASTRUCTURES IN THE SOCIAL SCIENCES

Thanks & Goodbye

cessda.eu

@CESSDA_Data

cessda

I. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Guidance:

Repositories take responsibility for stewardship of digital objects, and for ensuring that materials are held in the appropriate environment for appropriate periods of time. Depositors and users must be clear that preservation of and continued access to the data is an explicit role of the repository.

II. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Guidance:

Repositories must maintain all applicable licenses covering data access and use, communicate about them with users, and monitor compliance. This Requirement relates to the access regulations and applicable licenses set by the data repository itself, as well as any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information. Reviewers will be seeking evidence that the repository has sufficient controls in place according to the access criteria of their data holdings, as well as evidence that any relevant licences or processes are well managed.

III. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Guidance:

This Requirement covers the measures in place to ensure access to and availability of data holdings, both currently and in the future. Reviewers are seeking evidence that preparations are in place to address the risks inherent in changing circumstances.

Evidence for this Requirement should relate more to governance than to the technical information that is needed in R10 (Preservation plan) and R14 (Data reuse).

Who will take over the responsibility of the data holdings, and how will they be accessible in the future?

IV. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Guidance:
Adherence to ethical norms is critical to responsible science. Disclosure risk—for example, the risk that an individual who participated in a survey can be identified or that the precise location of an endangered species can be pinpointed—is a concern that many repositories must address. Evidence sought is concerned with not only having good practices for data with disclosure risks, but also the necessity to maintain the trust of those agreeing to have personal/sensitive data stored in the repository.

Requirement connected with R2 Licences.
How do you handle data with disclosure risk? Are data with disclosure risk stored appropriately to limit access?
Are there any special procedures applied to manage data with disclosure risk?
Evidence: documented procedures!

V. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Guidance:
Repositories need funding to carry out their responsibilities, along with a competent staff who have expertise in data archiving. However, it is also understood that continuity of funding is seldom guaranteed, and this must be balanced with the need for stability.

Does the repository have sufficient technical resources to fulfil the mission?
Does the repository have technical staff with the right competences?
Are there sufficient ongoing technical training to ensure skills and competences are maintained?

VI. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).

Guidance:
An effective repository strives to accommodate evolutions in data types, data volumes, and data rates, as well as to adopt the most effective new technologies in order to remain valuable to its Designated Community. Given the rapid pace of change in the research data environment, it is therefore advisable for a repository to secure the advice and feedback of expert users on a regular basis to ensure its continued relevance and improvement.

Does the repository have any objective technical expert advice beyond its own skilled staff?
How do you keep up with the most effective new technologies?

VII. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Guidance:

The repository should provide evidence to show that it operates a data and metadata management system suitable for ensuring integrity and authenticity during the processes of ingest, archival storage, and data access.

Integrity ensures that changes to data and metadata are documented and can be traced to the rationale and originator of the change.

Authenticity covers the degree of reliability of the original deposited data and its provenance, including the relationship between the original data and that disseminated, and whether or not existing relationships between datasets and/or metadata are maintained.

VIII. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Guidance:

The appraisal function is critical in determining whether data meet all criteria for inclusion in the collection and in establishing appropriate management for their preservation. Care must be taken to ensure that the data are relevant and understandable to the Designated Community served by the repository.

How do you deal with data that are deposited in non-preferred formats?
Do you use any special software for format transformations? How do you document transformations?

IX. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Guidance:

- Repositories need to store data and metadata from the point of deposit, through the ingest process, to the point of access. Repositories with a preservation remit must also offer ‘archival storage’ in OAIS terms.
- How are relevant processes and procedures documented and managed?
- What levels of security are required, and how are these supported?
- How is data storage addressed by the preservation policy?
- Does the repository have a strategy for backup/multiple copies? If so, what is it?
- Are data recovery provisions in place? What are they?
- Are risk management techniques used to inform the strategy?
- What checks are in place to ensure consistency across archival copies?
- How is deterioration of storage media handled and monitored?
- This requirement needs both input and close cooperation between data managers, technical staff and management

X. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Guidance:

The repository, data depositors, and Designated Community need to understand the level of responsibility undertaken for each deposited item in the repository. The repository must have the legal rights to undertake these responsibilities. Procedures must be documented and their completion assured.

The preservation plan should be managed to ensure that changes to data technology and user requirements are handled in a stable and timely manner.

XI. Data quality

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality related evaluations.

Guidance:
Repositories must work in concert with depositors to ensure that there is enough available information about the data such that the Designated Community can assess the substantive quality of the data. Such quality assessment becomes increasingly relevant when the Designated Community is multidisciplinary, where researchers may not have the personal experience to make an evaluation of quality from the data alone. Repositories must also be able to evaluate the technical quality of data deposits in terms of the completeness and quality of the materials provided, and the quality of the metadata.
Data, or associated metadata, may have quality issues relevant to their research value, but this does not preclude their use in science if a user can make a well-informed decision on their suitability through provided documentation.

XII. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Guidance:

To ensure the consistency of practices across datasets and services and to avoid ad hoc and reactive activities, archival workflows should be documented, and provisions for managed change should be in place. The procedure should be adapted to the repository mission and activities, and procedural documentation for archiving data should be clear.

Evidence should include levels of security at different steps within the workflow.
How does the type of data managed impact the workflow (technical aspect - data transformation, handling of sensitive data etc.)?

XIII. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Guidance:

Effective data discovery is key to data sharing, and most repositories provide searchable catalogues describing their holdings such that potential users can evaluate data to see if they meet their needs. Once discovered, datasets should be referenceable through full citations to the data, including persistent identifiers to ensure that data can be accessed into the future. Citations also provide credit and attribution to individuals who contributed to the creation of the dataset.

Give advice on technical solutions to enhance usability.
Technical aspects of data discovery and identification for both man and machine.
Extended searchability of the catalogue (elastic) + metadata harvesting.

XIV. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Guidance:

Repositories must ensure that data can be understood and used effectively into the future despite changes in technology. This requirement evaluates the measures taken to ensure that data are reusable.

XV. Technical infrastructure

R15. The repository functions on well-documented operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

- Repositories need to operate on reliable and stable core infrastructures
- Also, hardware and software used should be relevant and appropriate to:

the Designated Community

the functions it fulfils

- If possible, repository functions should be described by using standards, such as the OAIS
- specifies the functions of a repository in meeting user needs.

XV. Technical infrastructure

R15. The repository functions on well-documented operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

- Reviewer is looking for evidence that the applicant understands the wider ecosystem of standards, tools and technologies available for research data management and curation
- Understand your own technical infrastructure: what technical activities the repository is doing itself, and what is outsourced (and who is responsible)

XVI. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

- "The repository should analyze potential threats, assess risks, and create a consistent security system."

- Describe your arrangements to provide swift recovery of essential services in the event of an outage. Describe your disaster plan and risk analysis methods.

- Evidence is needed that you understand the technical risks and that you have mechanisms in place to respond to security incidents.

- Focus on technical infrastructure rather than on managerial aspects of business continuity.

XVI. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

The repository should:

- analyze potential threats
- assess risks
- create a consistent security system

Think about the damage scenarios:

- What are the malicious actions, human errors, or technical failures that pose a threat to the repository and its data, products, services, and users?
- What is the likelihood and impact of such scenarios?
- Which risk levels are acceptable?
- Which measures should be taken to counter the threats?

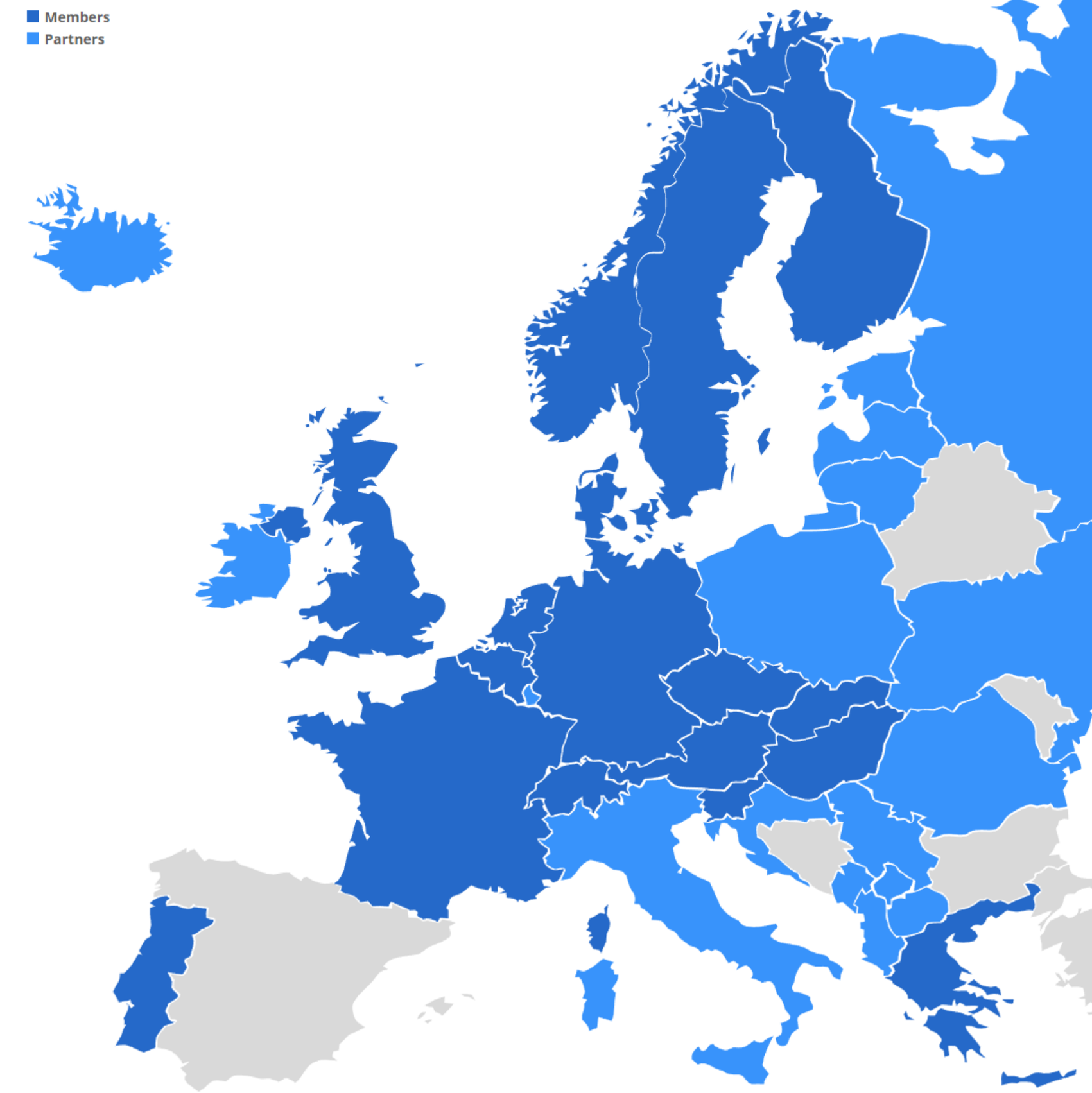
XVI. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

- Describe your arrangements to provide swift recovery of essential services in the event of an outage.
- Describe your IT security system, disaster plan and risk analysis methods.
- Evidence is needed that you understand the technical risks and that you have mechanisms in place to respond to security incidents.
- Focus on technical infrastructure rather than on managerial aspects of business continuity.
- If technical infrastructure is outsourced: how do you control that the arrangements are sufficient to guarantee the long-term preservation of and/or access to the data holdings?

Members

- Austria
- Belgium
- Czech Republic
- Denmark
- France
- Finland
- Germany
- Greece
- Hungary
- Netherlands
- Norway
- Portugal
- Slovakia
- Slovenia
- Sweden
- Switzerland (Observer)
- UK



cessda