cessda eric
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Trust Working Group

Workshop: Technical Aspects of Trust

Thursday 27 September 2018
Ljubljana, Slovenia

Hervé L'Hours
Chair: CESSDA Trust Group

Repository & Preservation Manager
UK Data Service, UK Data Archive

Trust Workshop

CESSDA Trust Group

- Herve L'Hours (UK Data Service)
- Mari Kleemola (FSD)
- Ilona von Stein (DANS)
- Maja Dolinar (ADP)
- Jonas Recker (GESIS)
- Birger Jerlehag (SND)

CORE TRUST SEAL

CESSDA TRUST Group

cessda eric

CESSDA Trust Group: Annex 2

Support in compliance with Annex 2 obligations (technical perspective)

- DDI & Metadata compliance
- Single sign-on
- Resource Discovery Metadata Harvesting
- Data via common gateways
- Share data archiving tools

cessda eric

CESSDA Trust: CoreTrustSeal

Support in Achieving CoreTrustSeal certification

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

cessda eric

Perspectives: trusted vs. trustworthy

If I don’t know you, then why should I…

- Trust you with data about me?
- Trust you with data I collected or created?
- Trust you with data I use and cite?

cessda eric

CoreTrustSeal

What you get

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

cessda eric

CoreTrustSeal

What you deliver

- Concise evidence statement for each requirement
  - responding to the requirement
  - addressing guidance (but its not a Q&A)
  - being honest and clear about you organisation
- Supported by documented evidence online
  - evidence statement should identify which part of the evidence is relevant, and why
  - brief English resume of evidence which is not in English

cessda eric

CoreTrustSeal

Community of Reviewers and Board

- Two independent Peer reviewers from CoreTrustSeal organisations
  - Submit comments and recommendation
- Board decision:
  - Applicant is approved for three years
    - May join the pool of reviewers
  - Or receives feedback and can reapply

cessda eric

CoreTrustSeal

Community of Reviewers and Board

- Community of Expertise
  - Developing standards
  - Designing review processes
  - Reviewing and updating
- Community of Practice
  - Public evidence
  - Reviewer Pool
  - Common ground for advancement

cessda eric

CESSDA, CoreTrustSeal & Beyond

A Two-Way Conversation

- Service Providers at different levels
  - Just beginning
  - In progress
  - Renewing and improving
- Expanding to: CESSDA Cloud, FAIR & EOSC
  - Feed back to CoreTrustSeal
  - SSHOC project alignment across SSH ERICs
  - Common ground for advancement and cooperation

cessda eric

Trust Workshop Agenda

- Introductions
- Requirements 0-14: technical aspects
- Requirement 15: technical Infrastructure
- Requirement 16: security
- Joining the dots
- Surgery (no requests yet, let us know by lunchtime)
- Working Group discussion & next steps (all welcome)

cessda eric

Introductions & Collaboration

- Introductions
- Collaborative notes

https://goo.gl/2MWG4H

cessda eric

cessda eric
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Trust Working Group

Workshop: Technical Aspects of Trust

Thursday 27 September 2018
Ljubljana, Slovenia

Maja Dolinar (ADP)
Birger Jerlehag (SND)

Trust Workshop

cessda eric
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

CESSDA ERIC Widening Meeting

Strengthening and Widening of the
European infrastructure of social science data archives

Thursday 27 September 2018
Ljubljana, Slovenia

Maja Dolinar (ADP)
Birger Jerlehag (SND)

Trust Workshop

Overview of Requirements 0-14: Technical aspects

0. Context

Repository Type

- 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)

cessda eric

0. Context

- Designated Community
- 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
- Outsource Partners

cessda eric

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.

cessda eric

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.

Requirement connected with R4 Confidentiality/Ethics.

Here you should think about the type of data you handle – How are sensitive data stored? Who has access and how is it managed? What are the limits on usage environment (safe room, secure remote access) and limits on type of users?

cessda eric

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?

cessda eric

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!

cessda eric

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?

cessda eric

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?

cessda eric

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.

cessda eric

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?

cessda eric

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.

cessda eric

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.

cessda eric

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.)?

cessda eric

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.

cessda eric

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.

cessda eric

cessda eric
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

CESSDA ERIC Widening Meeting

Strengthening and Widening of the
European infrastructure of social science data archives

Thursday 27 September 2018
Ljubljana, Slovenia

Ilona von Stein (DANS)

Trust Workshop

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.

cessda eric

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.

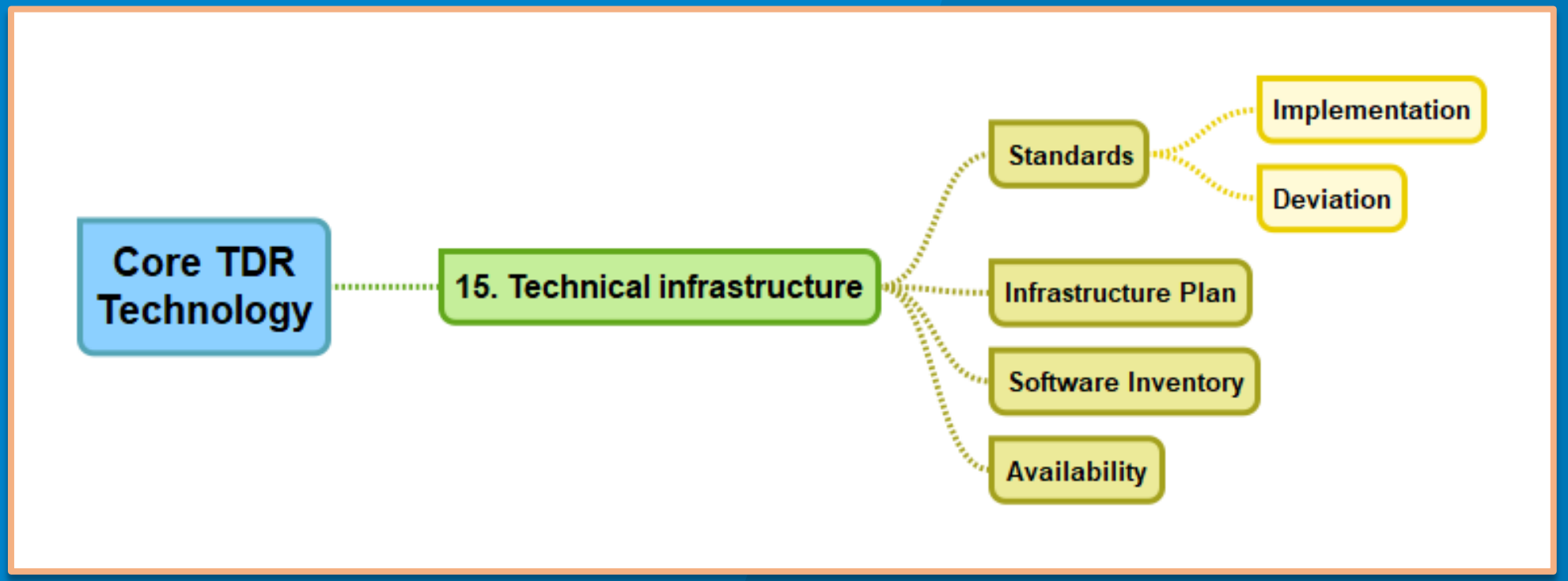
- Reviewer is looking for evidence that the applicant understands the wider ecosystem of standards, tools and technologies available for research data management and curation

- Understand your own technical infrastructure : what technical activities the repository is doing itself, and what is outsourced (and who is responsible)

cessda eric

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.



cessda eric

cessda eric
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

CESSDA ERIC Widening Meeting

Strengthening and Widening of the
European infrastructure of social science data archives

Thursday 27 September 2018
Ljubljana, Slovenia

Mari Kleemola (FSD)

Trust Workshop

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 techical risks and that you hace mechanisms in place to respond to security incidents.
- Focus on technical infrastructure rather than on managerial aspects of business continuity.

cessda eric

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?

cessda eric

XVI. Security

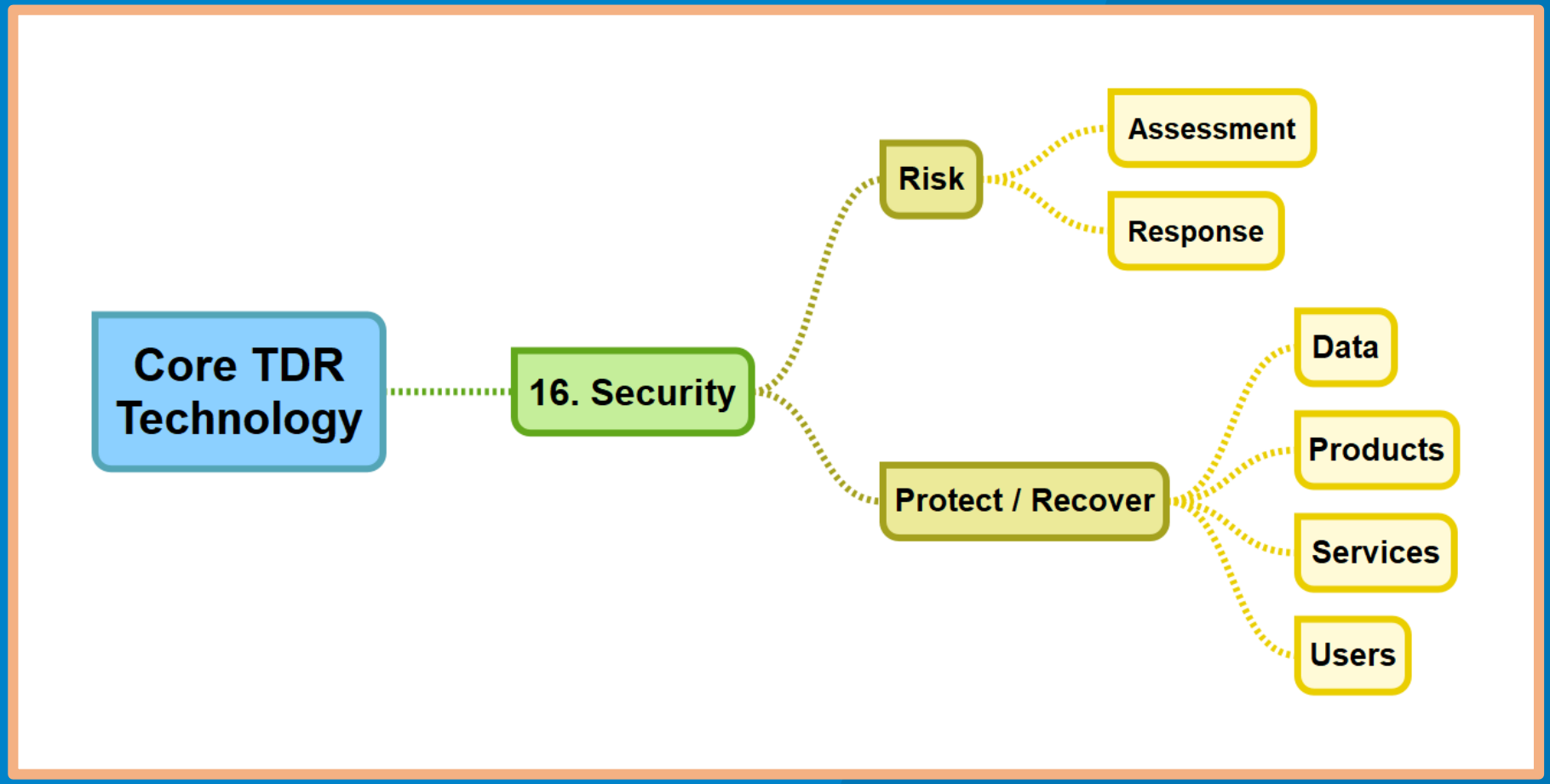
R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

- 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?

cessda eric

XVI. Security

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



cessda eric

cessda eric
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Trust Working Group

Workshop: Technical Aspects of Trust

Thursday 27 September 2018
Ljubljana, Slovenia

Hervé L'Hours
Chair: CESSDA Trust Group

Repository & Preservation Manager
UK Data Service, UK Data Archive

Trust Workshop

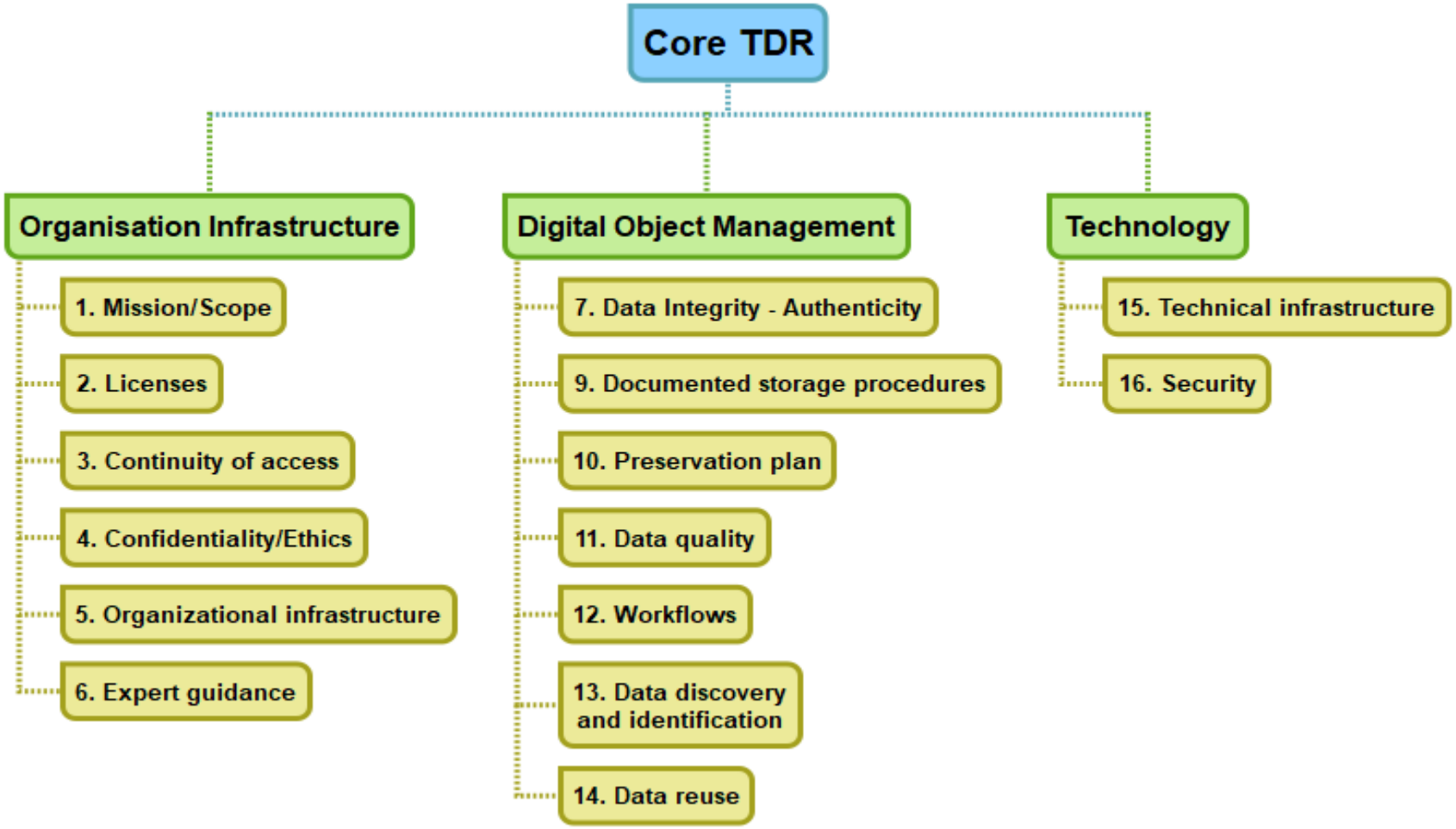
Joining the Dots

CoreTrustSeal

- What have we got
- What do we need
- How can we help

cessda eric

Joining the Dots



cessda eric

Compliance Levels

CoreTrustSeal

- 0 – Not applicable
- 1 – The repository has not considered this yet
- 2 – The repository has a theoretical concept
- 3 – The repository is in the implementation phase
- 4 – The guideline has been fully implemented

cessda eric

cessda eric
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Trust Working Group

Workshop: Technical Aspects of Trust

Thursday 27 September 2018
Ljubljana, Slovenia

Trust Workshop