

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

Wednesday 6 June 2018  
Milan, Italy

Hervé L'Hours

Chair: CESSDA Trust Group

Repository & Preservation Manager  
UK Data Service, UK Data Archive

## Trust Seminar



# Seminar Themes

- » Open Roundtable of Evidence for Requirements
  - » What do we have vs what do we need, and opportunities for alignment

## Additional Topics

- » Big Data (Volume, Variety, Velocity, Veracity, Value)
- » New and Novel Forms of Data
  - » Social Media
  - » Administrative Data
  - » Cross-disciplinary
- » Appraisal & Selection
- » Workflows
- » Versioning

# CESSDA Trust Group

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



# CESSDA Trust Group

Support in Achieving CoreTrustSeal certification

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

# 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?

# CoreTrustSeal

What you get

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

# CoreTrustSeal

## What you deliver

- Concise evidence statement for each requirement
  - responding to the requirement
  - addressing guidance (but it's not a Q&A)
  - being honest and clear about your 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

# CoreTrustSeal

Community of Reviewers and Board

- Two independent Peer reviews 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



# CoreTrustSeal

Assembly 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

# 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

# 0. Context

- 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)

# 0. Context

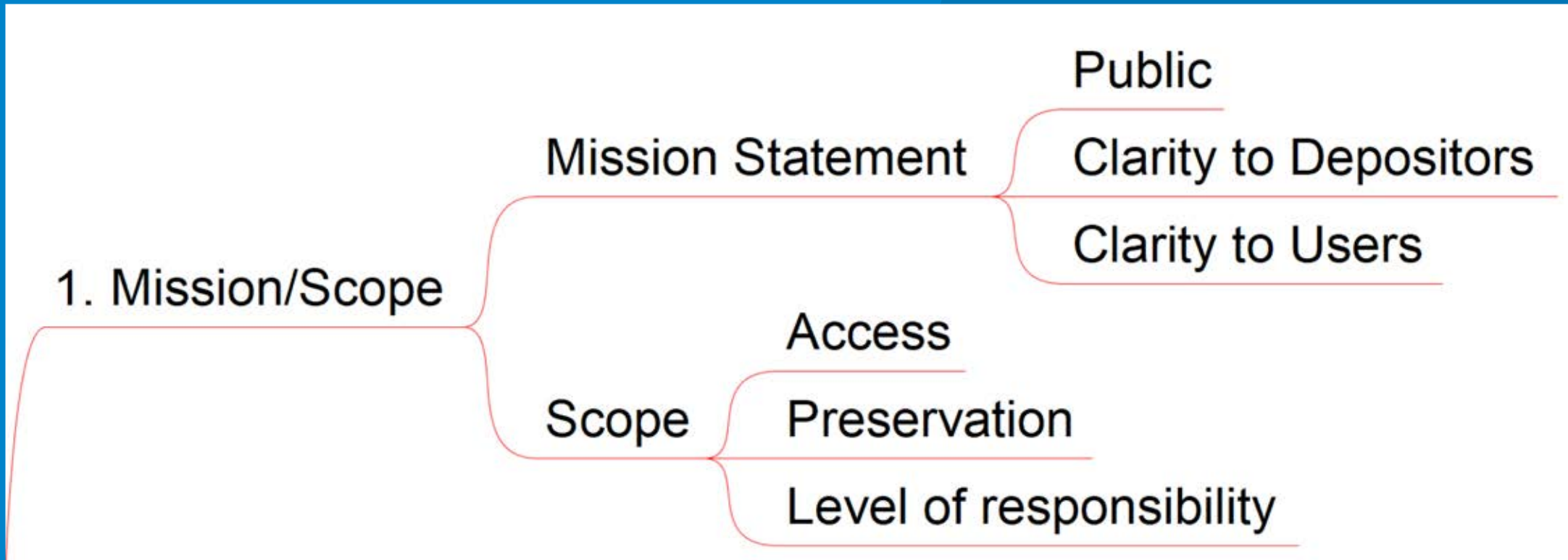
- Designated Community
- Level of Curation Performed
  - A. Content distributed as deposited
  - B. Basic curation – e.g., brief checking, addition of basic metadata or documentation
  - C. Enhanced curation – e.g., conversion to new formats, enhancement of documentation
  - D. Data-level curation –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

# *I. Mission/Scope*

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

# I. Mission/Scope

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

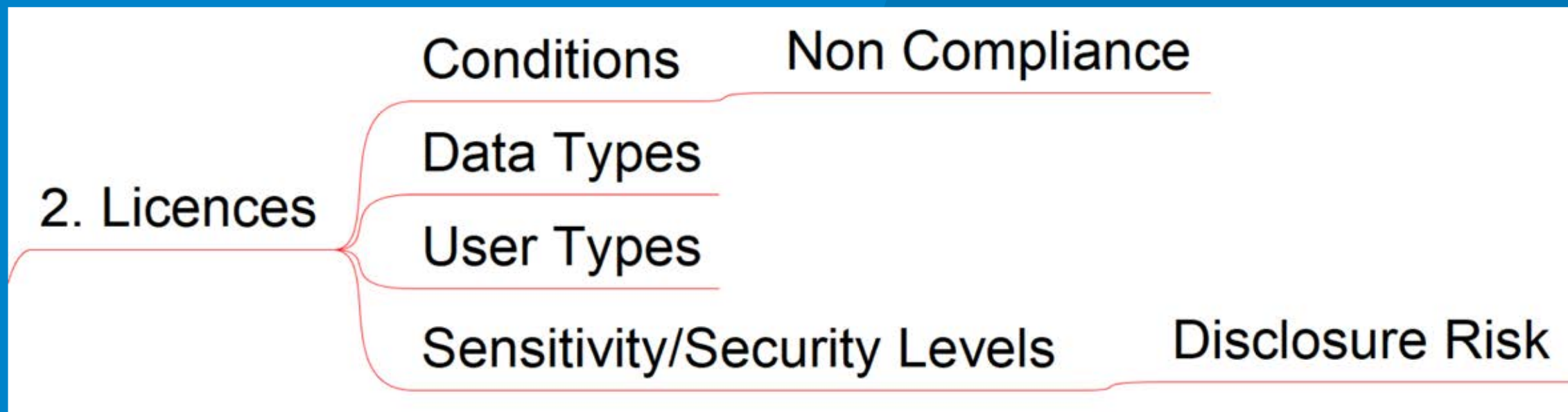


## *II. Licenses*

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

## II. Licenses

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



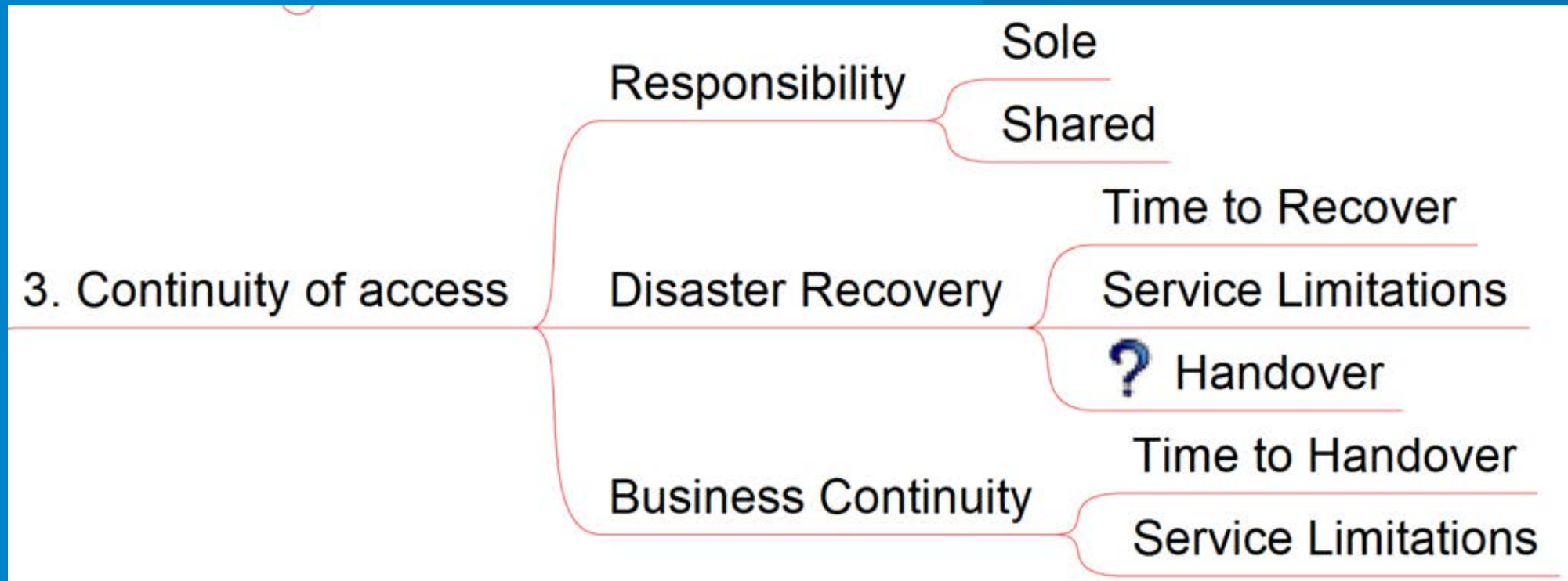


# *III. Continuity of access*

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

# III. Continuity of access

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

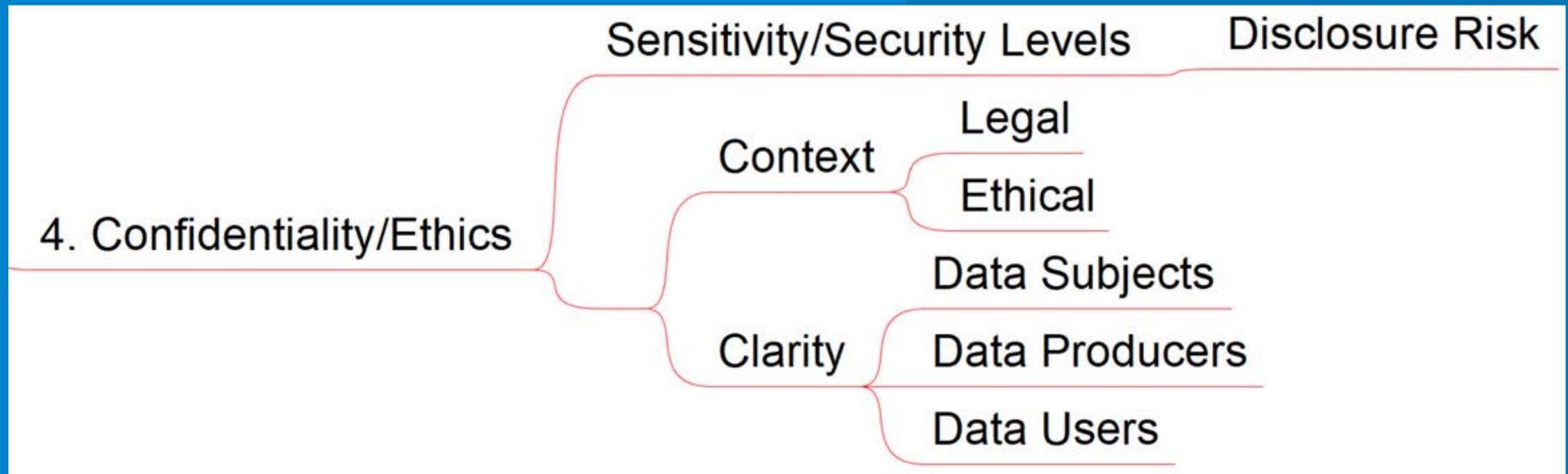


## *IV. Confidentiality/Ethics*

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

# 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.*

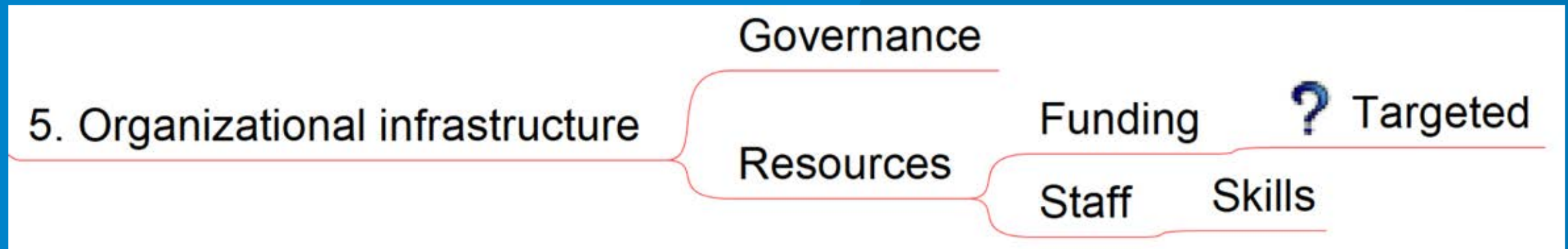


# *V. Organizational infrastructure*

*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.*

# 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.*

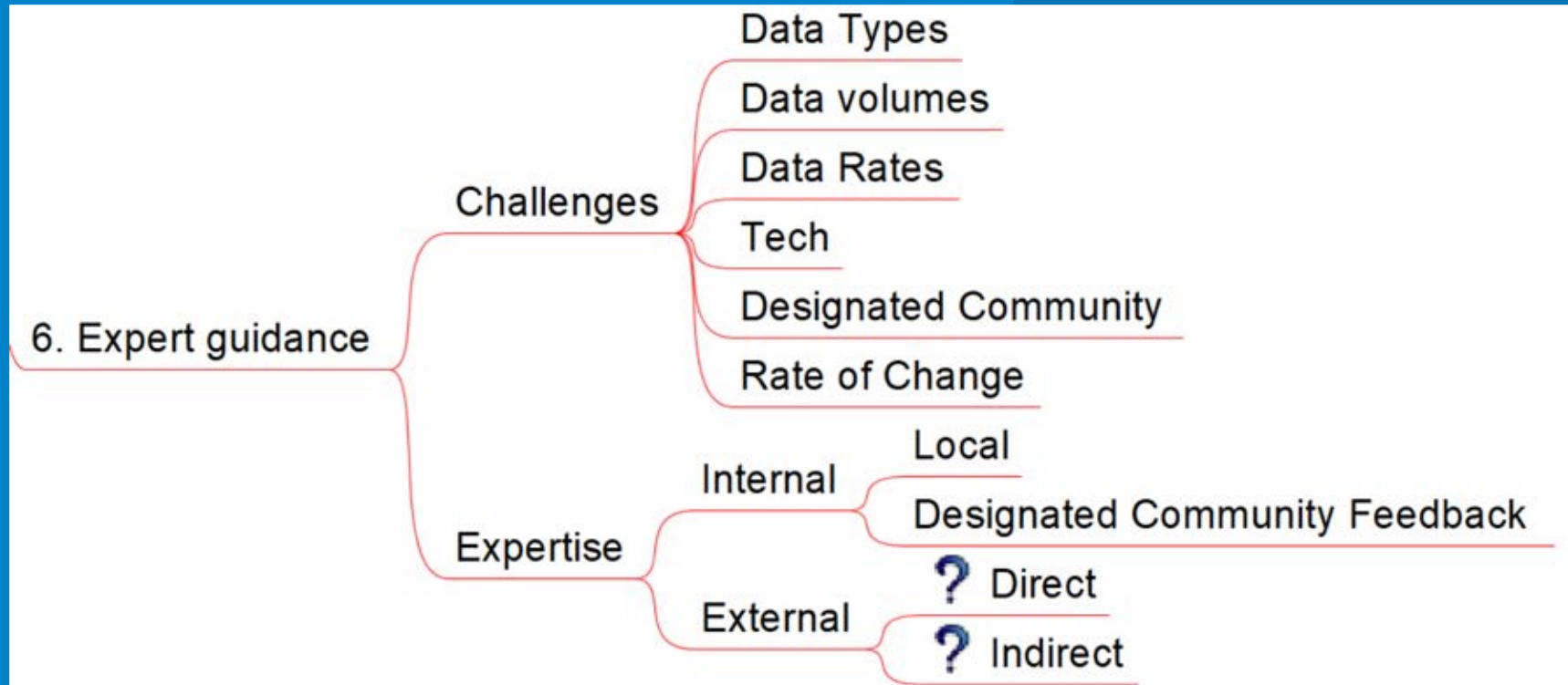


## VI. *Expert guidance*

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

# 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).





# *VII. Data integrity and authenticity*

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

# VII. Data integrity and authenticity

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

7. Data Integrity - Authenticity

Integrity

? Versions

Authenticity

# *VIII. Appraisal*

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

# VIII. Appraisal

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

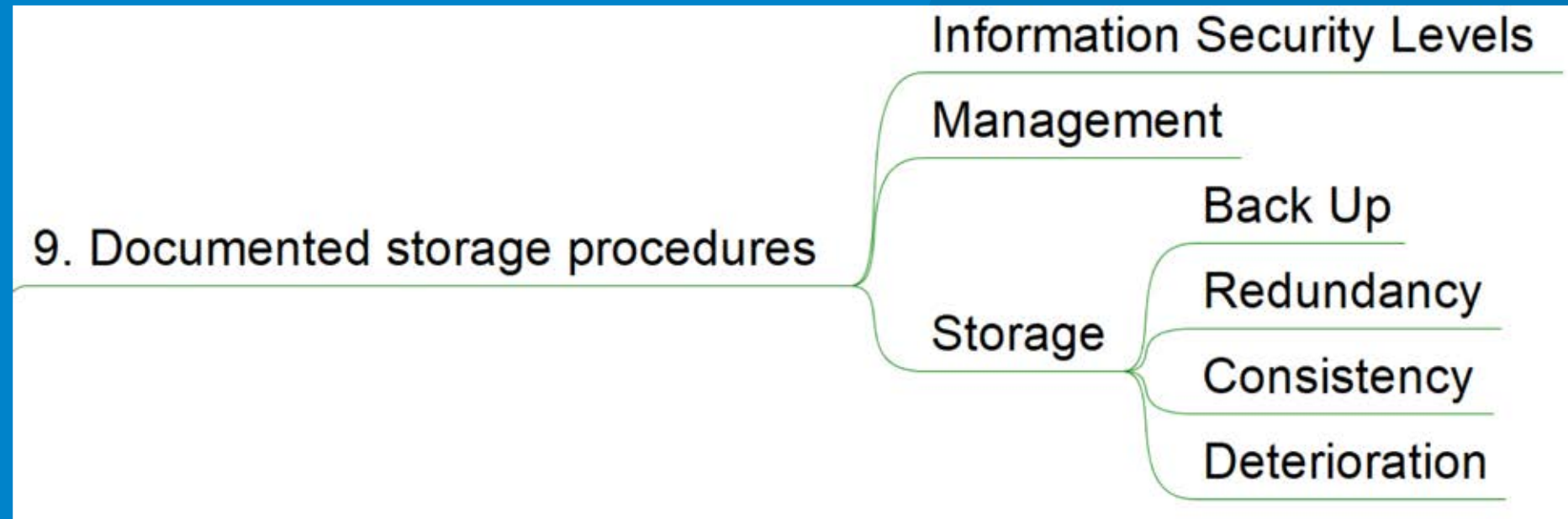


# *IX. Documented storage procedures*

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

# IX. Documented storage procedures

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



# *X. Preservation plan*

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

# X. Preservation plan

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

	Responsibility	Rights
10. Preservation plan	Documented	Managed
	Audit	

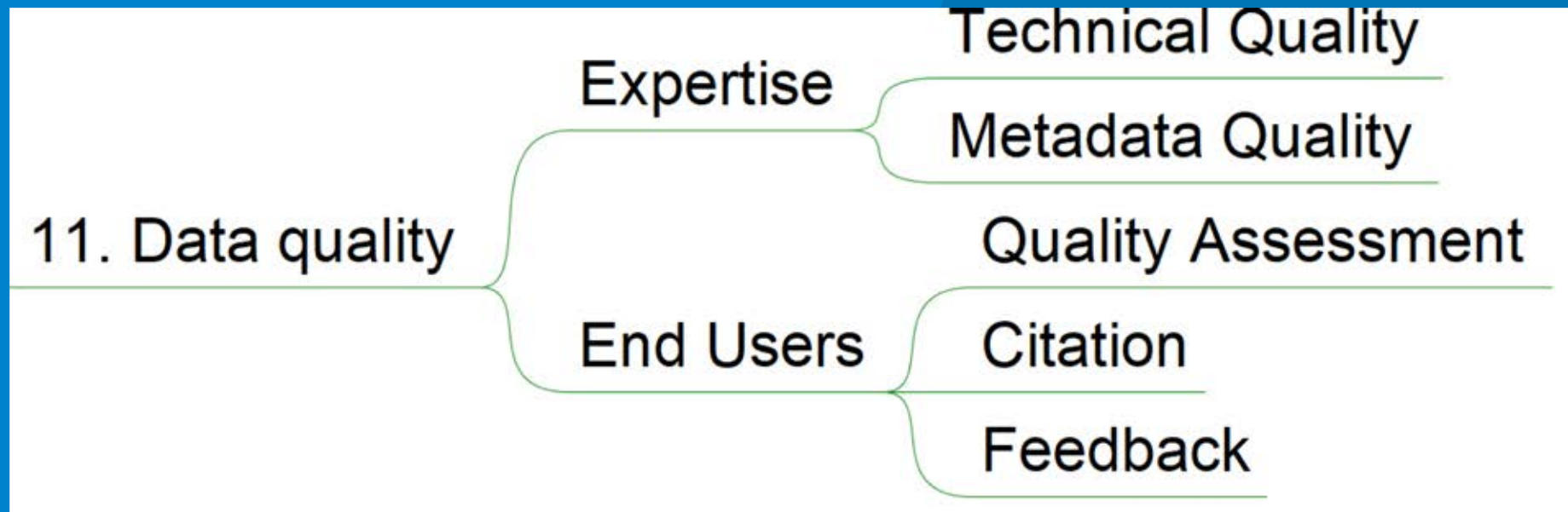


# *XI. Data quality*

*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.*

# 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.*

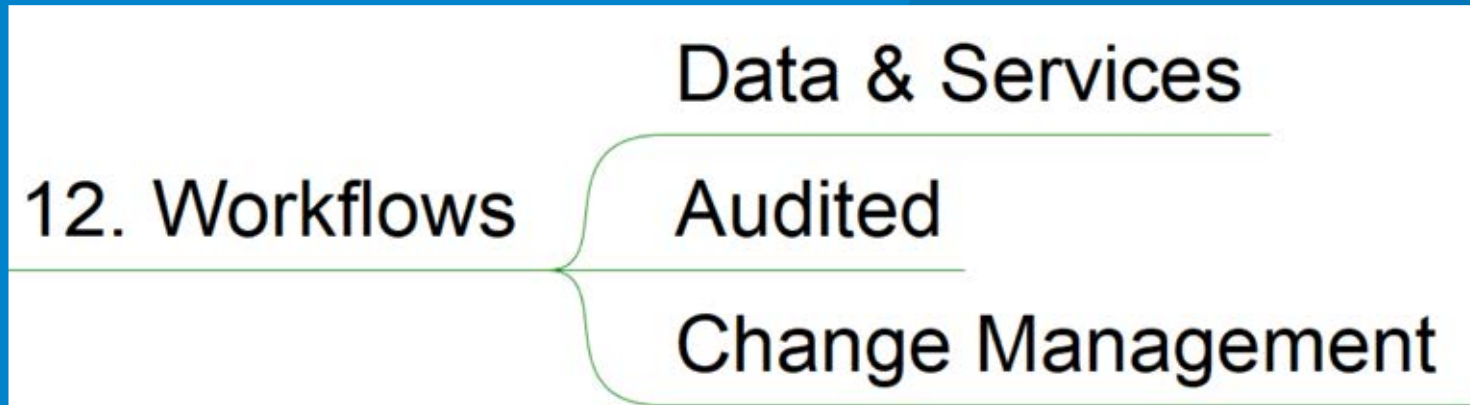


# *XII. Workflows*

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

# XII. Workflows

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

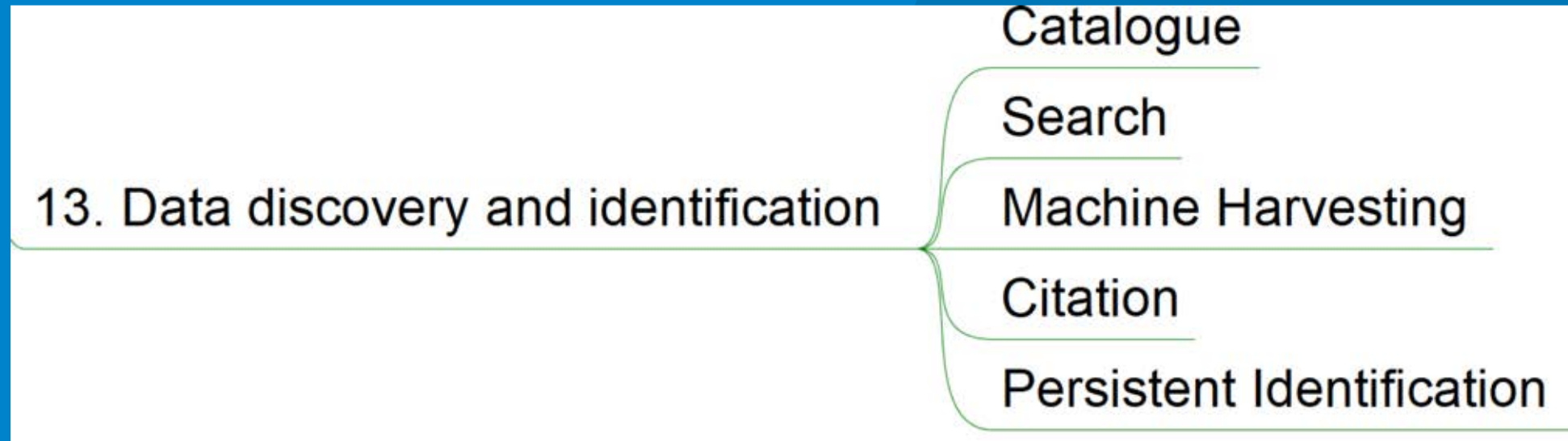


# *XIII. Data discovery and identification*

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

# 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.*

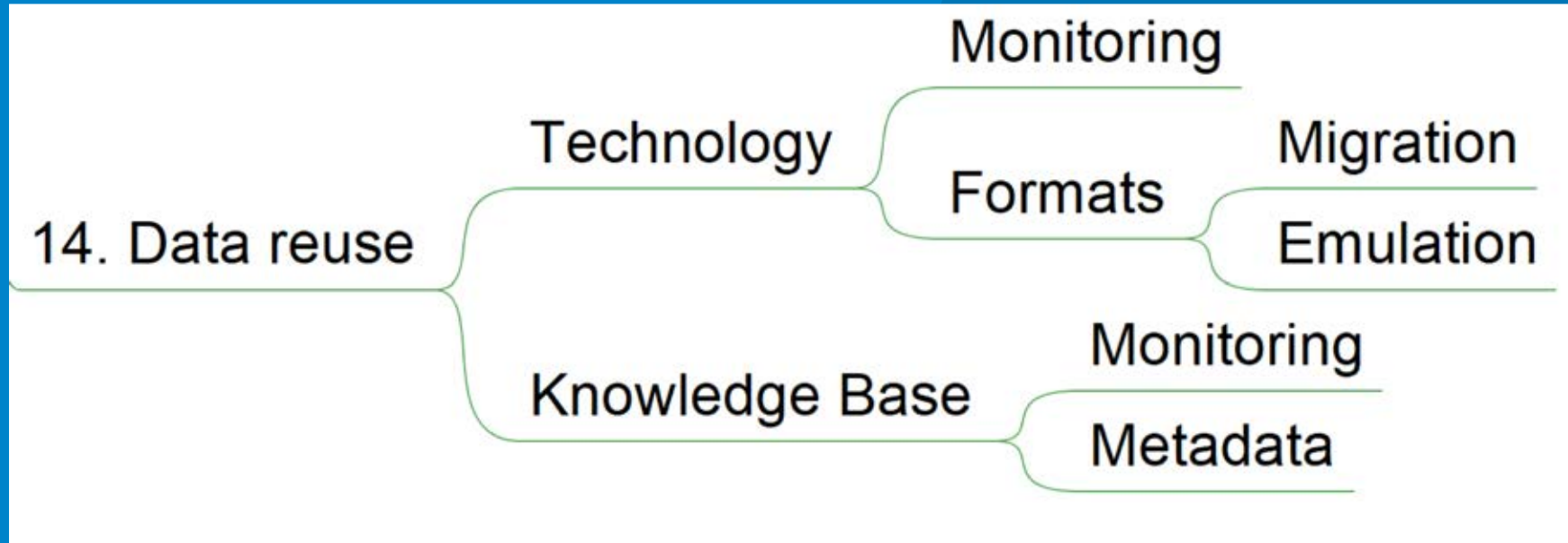


## *XIV. Data reuse*

*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.*

# 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.*



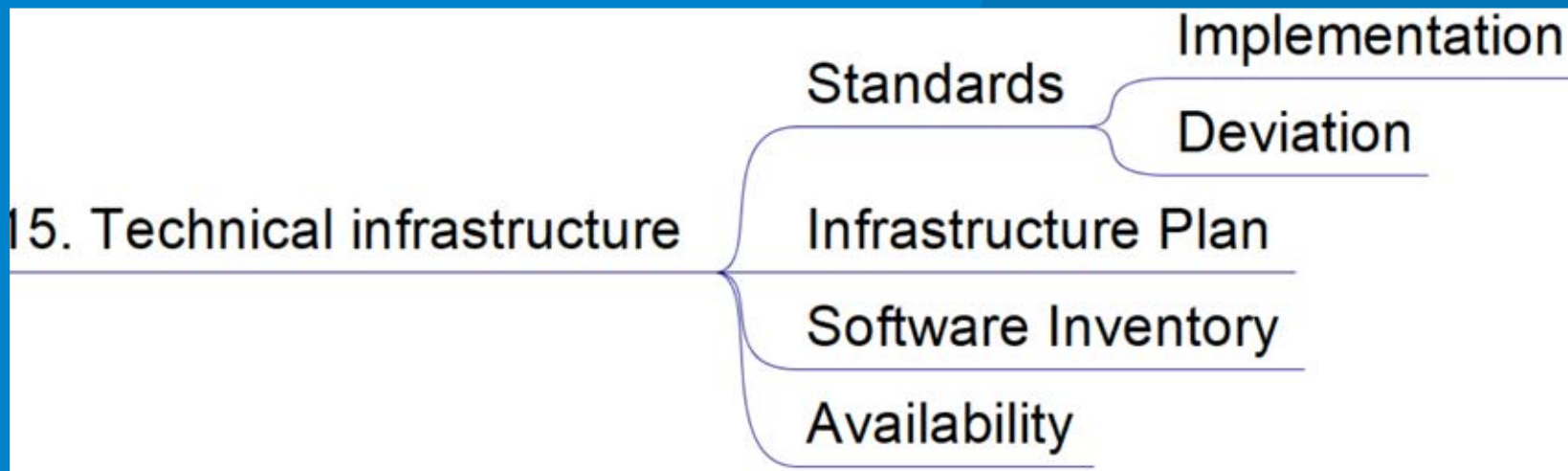


# *XV. Technical infrastructure*

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

# XV. Technical infrastructure

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

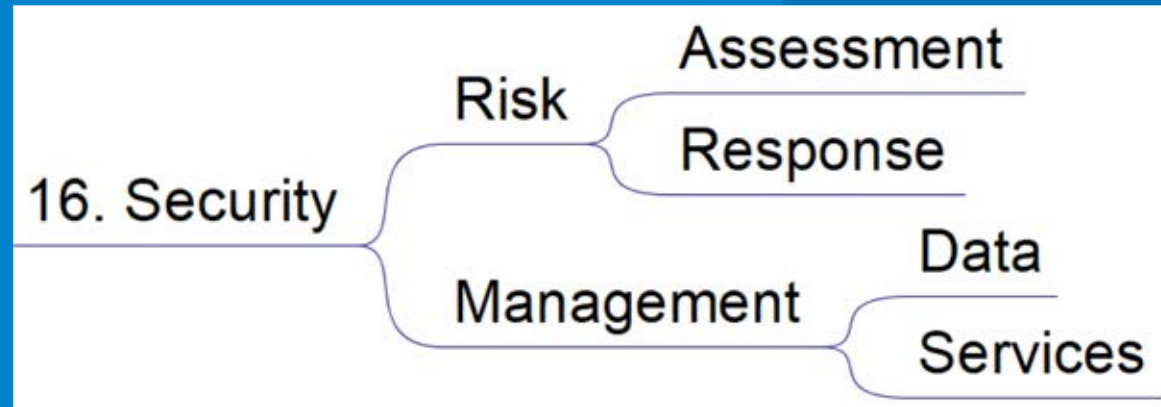


# *XVI. Security*

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

# XVI. Security

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



# Seminar Themes

- » Open Roundtable of Evidence for Requirements
  - » What do we have vs what do we need, and opportunities for alignment

## Additional Topics

- » Big Data (Volume, Variety, Velocity, Veracity, Value)
- » New and Novel Forms of Data
  - » Social Media
  - » Administrative Data
  - » Cross-disciplinary
- » Appraisal & Selection
- » Workflows
- » Versioning

cessda eric

cessda eric

cessda eric



# I. Mission/Scope

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

- If data management is not referred to in the mission statement, then this requirement, as a rule, cannot have a compliance level of 3 or higher.

## II. Licenses

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

- Access and use conditions could be set differently: either as standard terms and conditions, or as differentiated for particular depositors or datasets. These could cover the level of curation, what is the liability level, the level of responsibility taken for the data, limitations on use, limits on usage environment (safe room, secure remote access), limits on types of users (approved researcher, has received training, etc.).
- The consequences if noncompliance is detected (e.g., sanctions on current or future access/use of data) should be made clear. Ideally, repositories should have a public policy in place for noncompliance.
- The minimum compliance level should be 4, if the applicant is currently providing access to data.

# *III. Continuity of access*

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

The level of responsibility for data should be indicated in the evidence.

This information helps the reviewer to judge whether the organisation is sustainable in terms of its finances and processes; in particular the continuity of its collections and responsibilities in the case of cessation of funding.

# 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.*

- All organisations responsible for data have an ethical duty to manage them to the level as expected by the scientific practice of its designated community. For repositories holding data about individuals, businesses, or other organisations, there are furthermore obligations and obligations that the rights of the data subjects will be protected. These will be both of a legal and ethical nature.
- Disclosure of these data could also present a risk of personal harm, a breach of commercial confidentiality, or the release of critical information (e.g., the location of protected species or an archaeological site).
- Reviewers expect to see evidence that the applicant understands their legal environment and the relevant ethical practices, and has documented procedures.
- Minimum compliance level should be a 4 if the repository is currently providing access to personal data.

# 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.*

- The description of this requirement should contain evidence describing the organisation's governance/management decision making processes, and the entities involved. Staff should have appropriate training in data management to ensure consistent quality standards.
- In what degree is funding structural or project-based? Can this be expressed in FTE numbers?
- How often does periodic renewal occur?

# 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).*

- The reviewer should be looking for evidence that the repository is linked to a wider network of expertise in order to demonstrate access to advice and guidance for both its day-to-day activities and the monitoring of potential new challenges on the horizon (science and technology watch).
- Part of this information may already have been given under “R0. Brief description of the designated community” and ‘Other relevant information’, if so, then please refer to it.

# VII. Data integrity and authenticity

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

- A clear and complete context section is important for all requirements but this is particularly the case for this long requirement 7. The organisation of the curation and the types of data will help guide the reviewer expectation. The reviewer would benefit from a clear overview of the processes and tools used to curate the data including the level of manual and automated practice, and the how the processes, tools and practices are documented. Most useful would be when the applicant responds to each bullet point separately and to address integrity and authenticity independently as defined in the requirement.
- Audit trails (written evidence on which actions have been performed on the data) should be elaborated on in the evidence

# VIII. Appraisal

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

- The applicant should be able to demonstrate that procedures are in place to ensure that only data appropriate to the collection policy are accepted and that they have all the necessary information and procedures and skills to ensure long term preservation and use relevant for the designated community.



# *IX. Documented storage procedures*

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

- The reviewer will be looking to understand each of the storage locations which support curation processes, how data are appropriately managed in each environment and that processes are in place to monitor and manage change to storage documentation.
- Can the repository recover from short-term disasters?
- Are procedures documented and standardised in such a way that different data managers, while performing the same tasks separately, will arrive at substantially the same outcome?

# X. Preservation plan

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

- The reviewer will be looking for clear managed documentation to ensure (1) a managed approach to long term preservation (2) continued access for data types despite format changes and (3) with sufficient documentation to support usability by the designated community.
- 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.*

The applicant should make clear in his statements that he understands the quality levels which can reasonably be expected from depositors. This should describe the quality assurance and improvement it will undertake during curation and the quality expectations of users, which may involve documentation of areas where quality thresholds have not been met.

## XII. Workflows

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

- The reviewer is looking for evidence that the applicant takes a consistent, rigorous, documented approach to managing its activities throughout their processes and that changes to those processes are appropriately evaluated, documented, managed and implemented

## *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.*

This should contain evidence that the curation of data and metadata is designed to support resource discovery of clearly defined and identified digital objects. It should be clear to the users of this data how it must be cited to provide appropriate academic credit and linkage between related research.

## *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.*

The applicant should understand the needs of the designated community in terms of their research practises and technical environment and used standards. Changes in technology are important, but appropriate high quality metadata should also play an essential role and should be mentioned in the evidence provided. The latter information is critical to design curation processes which result in digital objects that meet the needs of the end user as well as generic or disciplinary standards.

# *XV. Technical infrastructure*

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

The workflows and human actors providing repository services must be supported by a technological infrastructure. If possible this should be demonstrated by using a reference model.

The reviewer is looking for evidence that the applicant understands the wider ecosystem of standards, tools and technologies available for (research) data management and curation and has selected options which align with local requirements.

# XVI. Security

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

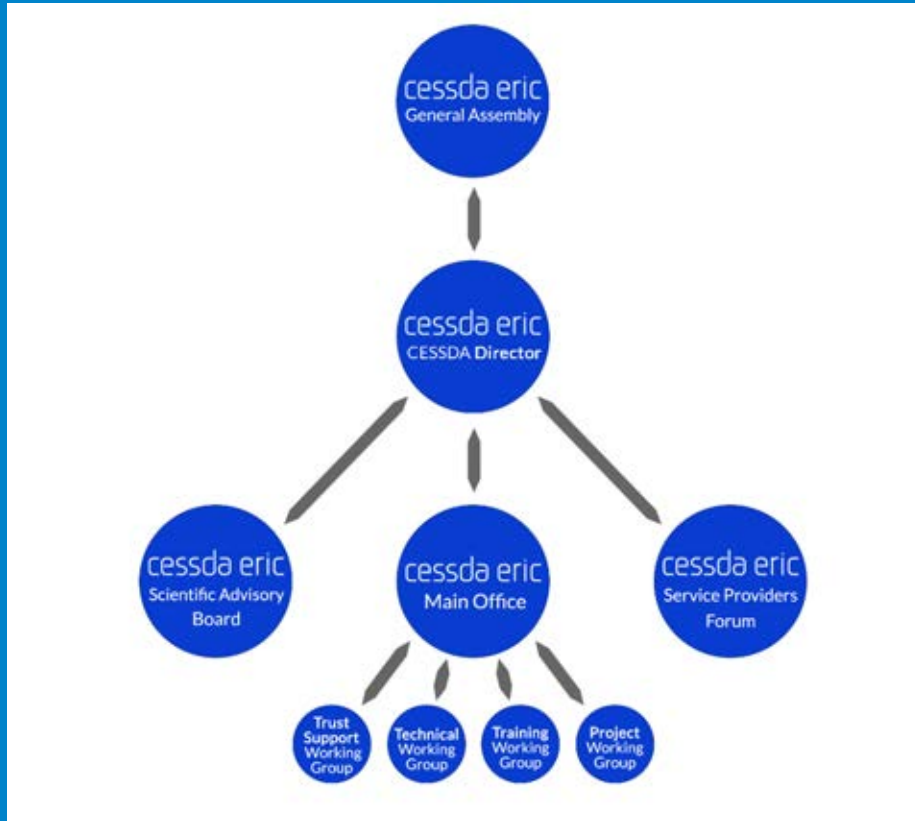
The applicant should understand the technical risks applicable to its particular service data user and physical environment and that it has mechanisms in place to respond to incidents.

Evidence must focus on technical infrastructure rather than on managerial and procedural aspects of business continuity.

In what way is the technical infrastructure determined by the repository or by their host /outsource institution?



# CESSDA ERIC Governance



## Bodies of CESSDA ERIC

- » General Assembly
- » Service Providers' Forum
- » Scientific Advisory Board
- » Working Groups

# CESSDA Trust Group

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



# What is CESSDA?

- » Cross-European resource discovery
- » Improved quality of data and metadata
- » A wider selection of comparable data
- » Certification of data archiving organisations
- » Professional training for data archivists and scientific community
- » Improved mechanisms for data dissemination and analysis
- » Strong involvement of organisations outside Europe

*The mission of CESSDA is to provide a full scale sustainable research infrastructure that enables the research community to conduct high-quality research in the social sciences which can contribute to effective solutions to the major challenges facing society today.*

# MEMBERS

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

