



# Audit & Certification

## Data Seal of Approval

Hervé L'Hours. UK Data Archive & DSA Board

APARSEN Trust & Digital Preservation hosted by the Digital Repository of Ireland

Royal Irish Academy, Dublin 4-5 June 2013



## UK Data Archive

“Mucking about with your data since 1967”

- Board Member of the Data Seal of Approval
- APARSEN Test Audit Against ISO16363
- Secure Data Service, now part of UK Data Service
- ISO27001 Information Security
- CESSDA & EUDAT,

Trust Maturity Level	Key Guideline	Guideline Source
<b>1. OAIS Core Conformance</b>	<p>Support OAIS Information Model.</p> <p>Acknowledge OAIS Archive responsibilities.</p>	<p>OAIS Information Model: Section 2.2 of CCSDS 650.0-M-2 / ISO 14721:2012.</p> <p>OAIS Archive Responsibilities: Section 3.1 of CCSDS 650.0-M-2 / ISO 14721:2012.</p>
<b>2. Initial self-assessment, PLATTER/DRAMBORA</b>	Self-assessment through PLATTER and DRAMBORA.	<p>PLATTER Key Self-assessment questions.</p> <p>DRAMBORA Key Self-assessment questions.</p>
<b>3. Peer-reviewed self-assessment I, DSA</b>	Peer-reviewed self-assessment I, DSA.	<p>Data Seal of Approval Guidelines.</p> <p>Support: NESTOR criteria</p>
<b>4. Peer-reviewed self-assessment II, ISO 16363/DIN 31644</b>	<p>Conformance to the OAIS Detailed Functional Model.</p> <p>Self-audit with the ISO 16363.</p> <p>Alternatively, self-audit with DIN 31644.</p>	<p>OAIS Detailed Functional Model: Section 4.1 of CCSDS 650.0-M-2 / ISO 14721:2012.</p> <p>CCSDS 652.0-M-1 / ISO 16363:2012.</p> <p>DIN 31644</p>
<b>5. Certification and Optimization</b>	<p>External review and formal certification in conformance with the ISO 16363.</p> <p>Alternatively, with DIN 31644.</p>	<p>CCSDS 652.0-M-1 / ISO 16363:2012.</p> <p>DIN 31644.</p>



# Introduction: the Certification Framework

## Memorandum of Understanding for a Three-tiered Framework

- **Formal certification:** DSA + full external audit and certification based on ISO 16363\* or DIN 31644\*\*
- **Extended certification:** DSA + structured, externally reviewed and publicly available self-audit based on ISO 16363\* or DIN 31644\*\*
- **Basic certification:** Data Seal of Approval (DSA)

## Information and Documentation - Criteria for Trustworthy Digital Archives

<http://www.trusteddigitalrepository.eu>

\*ISO 16363 - Audit and Certification of Trustworthy Digital Repositories

\*\*DIN 31644 - Information and Documentation - Criteria for Trustworthy Digital Archives



# Foundations

## Trust

- **Trust is a process**
  - Towards greater quality
  - Towards better relationships
  - Towards more certainty



# Foundations

**Trust:** *“Reliance on and confidence in the truth, worth, reliability of a person or thing”*

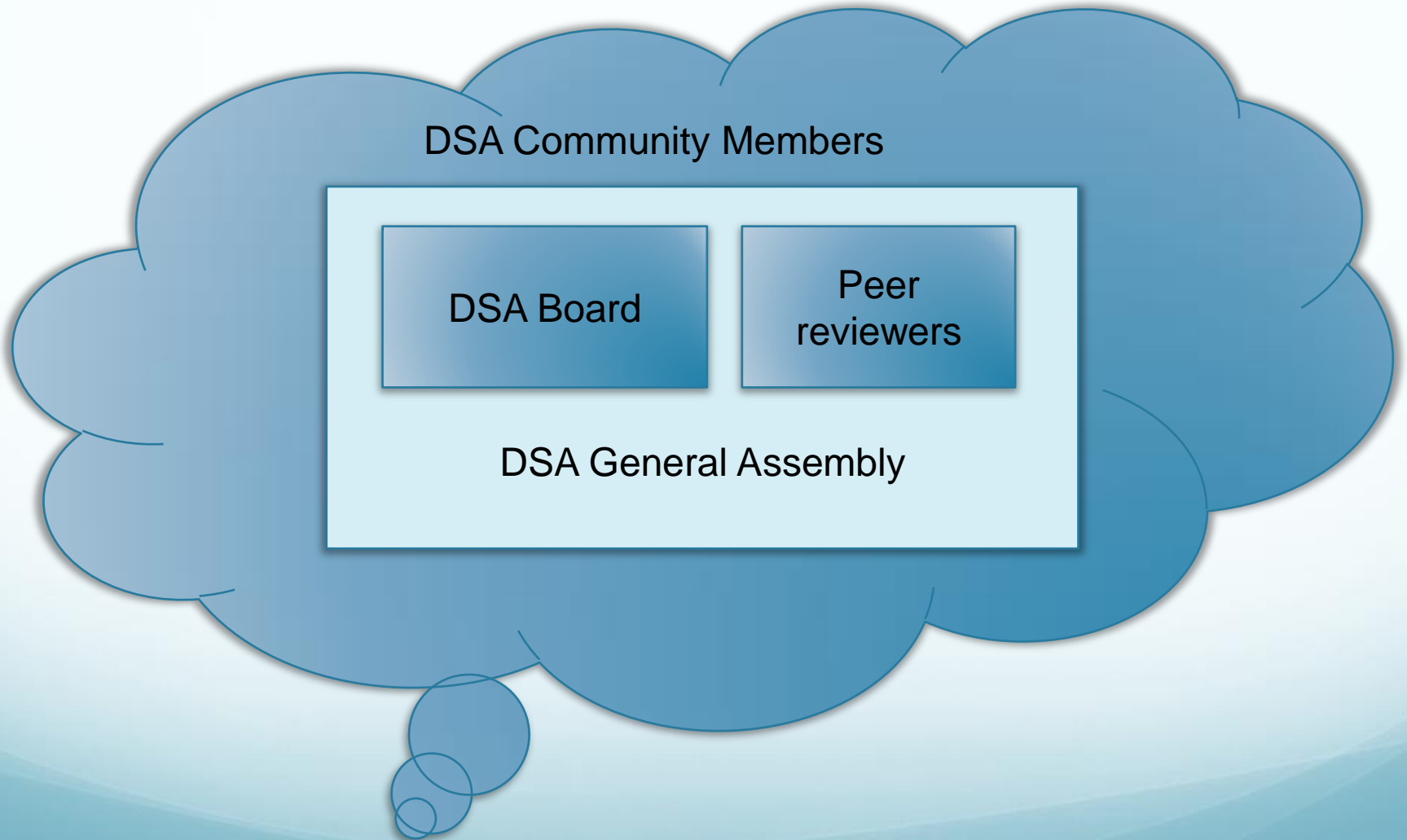
**Transparency:** *“Minimum degree of disclosure to which agreements, dealings, practices, and transactions are open to all for verification”*

**Community:** The Data Seal of Approval was established by a number of institutions committed to the long-term archiving of data. By assigning the seal, the DSA **community** seeks to guarantee the durability of the data concerned, but also to promote the goal of durable archiving in general.

**Trust is transitive**



# Governance





# Governance

- All recipients of a current DSA are automatically members of the **DSA Community**.
- Organisations that are part of the DSA Community may choose to become Members of the **DSA General Assembly** and designate one representative with voting rights.
- The General Assembly elects the **DSA Board** and provides advice to the Board when needed
- The **DSA Board** consists of six to eight representatives elected for two year terms with re-election possible.





# Governance

The current DSA Board

Different countries:

- NL, UK, FR, DE, USA

Different scientific fields

- Linguistics, Social sciences, Life sciences

Different functions:

- Archives, research centers, ICT centers

Different research-infrastructures:

- CLARIN, CESSDA, DARIAH ...



# Regulations

- The Data Seal of Approval Guidelines remain current for a period of two calendar years, the 'Seal Period'
- The DSA Board meets formally once a year, Board elections are held every two years
- The Board considers amendments to the DSA Guidelines and Regulations
- Any changes become effective at the end of the current Seal Period



# Regulations

- The Data Seal of Approval is awarded per Seal Period.
- The DSA logo displayed includes the two-year Seal dates.
- Each DSA Community member is contacted six months before the end of the current Seal period as a prompt for renewal.
- The organisation can then either choose to update its self-assessment for the new period and apply for the latest Seal, including a new version of the guidelines when applicable, or choose to display the outdated logo on its Web site.



# Objectives

## Data Producers

- *Assurance* of reliable data Storage

## Funding Bodies

- *Confidence* that data is available for re-use

## Data Consumers

- *Enables* assessment of repositories



# Principles

The data are:

- available on the Internet
- accessible
  - while taking into account relevant legislation with regard to personal information and intellectual property of the data.
- usable (file formats)
- reliable
- citable (can be referred to)



# Stakeholders

## *Data Producer*

- responsible for the quality of the digital data

## *Data Repository*

- responsible for the quality of data storage & availability

## *Data Consumer*

- responsible for the quality of use of the digital data



# Responsibility: the DSA Focus

**The DSA focus is on the Repository as *enabler* of good Data Producer and Data Consumer practice**

A data repository is designated a *Trusted Digital Repository* (TDR) if:

- It enables Data Producers to adhere to Guidelines 1-3
- It meets guidelines 4-13
- It enables Data Consumers to adhere to guidelines 14-16

The Seal is displayed only on the repository web site



# Compliance

## Minimum level of compliance for each guideline

- Must be met to receive the Data Seal of Approval
- Compliance levels will be evaluated and will increase as:
  - Best practices emerge
  - Compliant tools become available
  - Implementation occurs





# Compliance Levels

Level	Compliance Level Definition	Requirements
0	Not Applicable	Provide an explanation
1	We have not considered this yet	Provide an explanation
2	We have a theoretical concept	Provide a URL for the initiation document.
3	We are in the implementation phase.	Provide a URL for the definition document.
4	This guideline has been fully implemented for the needs of our repository	Provide a URL for the definition document.



# Evidence

## Transparency

- Link to publicly available documentation
  - Or deadline for public release
- English or short summary in English

Reviewers Guide: “Topics for discussion and inclusion are suggested but they are neither exhaustive nor prescriptive”

## How do we know what is

- Appropriate?
- Sufficient?



# Peer Reviewers

## Guidance :

- Does the self-assessment response correspond to the guideline?
- Are links to supporting documentation available publically?
- Do you agree with the self-assessed compliance levels?
  - are they sufficient to award the DSA for this guideline?
- Have abbreviations been explained?

In responding to the self-assessment try to provide helpful comments rather than specific questions.



# The task at hand

#	Guideline	Abbreviated Guideline	Minimum Requirement	Compliance	Evidence	Notes
1	The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data and compliance with disciplinary and ethical norms.	1. Deposit Quality and Compliance	3-In Progress		Exists Written Public	



# Data Producer

#	Abbreviated Guideline	Minimum Requirement
1	1. Deposit Quality and Compliance	3-In Progress
2	2. Recommended Formats	3-In Progress
3	3. Requested Metadata	4-Implemented



# The Repository

#	Abbreviated Guideline	Minimum Requirement
4	4. Explicit Mission	4-Implemented
5	5. Legal/Contract Compliance	4-Implemented
6	6. Data Storage Processes	4-Implemented
7	7. Preservation Plan	3-In Progress
8	8. Explicit Data Lifecycle Workflows	3-In Progress
9	9: Responsibility for Access	4-Implemented
10	10. Discovery, Use and Reference	3-In Progress
11	11. Object/Metadata Integrity	3-In Progress
12	12. Object/Metadata Authenticity	3-In Progress
13	13. Technical Infrastructure Standards	3-In Progress



# Data Consumer

#	Abbreviated Guideline	Minimum Requirement
14	14. Compliance with Access Criteria	4-Implemented
15	15: Proper Use and Exchange of Information	4-Implemented
16	16. Usage Licence Compliance	4-Implemented



# The task at hand

#	Guideline	Abbreviated Guideline	Minimum Requirement	Compliance	Evidence	Notes
1	The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data and compliance with disciplinary and ethical norms.	1. Deposit Quality and Compliance	3-In Progress		Exists Written Public	





# LUNCH!



# A Bit of Context

A Trust-Based Trust Standard

A Work in Progress (Help!)

Scalable through the Network/Community

Challenges of Consistency (who watches the Watchmen?)

Legality & Liability (Access, Anonymisation, Ethics and all that)

(Research) Data Management, Information Technology,  
discipline-specific needs, “I’m Special”



# A Bit More Context

Show me the Money! /Cost/Risk of Failure

## **Sub-Certification Foundations (Achievable Goals):**

- Records Management
- Understanding business processes,
- Analysing risk
- Managing Change
- Policies, Procedures and Publicness



# Data Producers

Guidelines 1 to 3:

The level of guidance which the Repository gives to the Data Producer before and during submission to the Repository.

Responses concentrate efforts by the Repository in supporting compliance by the Data Producer.



# Data Producers: the Content

**“1. The *data producer* deposits the data in a data repository with sufficient information for others to assess the quality of the data and compliance with disciplinary and ethical norms.”**

- **Minimum:** We are in the implementation phase (3)
- **Outsourced:** No

Can users of the data assess the quality, value whether it is ‘of interest’:

- Scientific
- Scholarly
- Business



# Data Producers: the Content

**“1. The *data producer* deposits the data in a data repository with sufficient information for others to assess the quality of the data and compliance with disciplinary and ethical norms.”**

- Transparency
- Sector-specific/Designated Community quality criteria
- Adherence to disciplinary & ethical norms
- Assessment by experts and colleagues



# Data Producers: the Content

**“1. The *data producer* deposits the data in a data repository with sufficient information for others to assess the quality of the data and compliance with disciplinary and ethical norms.”**

Does the repository?

- Define the full package of information that should be deposited to facilitate assessment?
  - citations based on the data?
  - A methodology report?
  - official approval for data collection (to confirm adherence to legal or ethical requirements)?
- Promote data sharing and reuse?



# Data Producers: the Content

**“1. The *data producer* deposits the data in a data repository with sufficient information for others to assess the quality of the data and compliance with disciplinary and ethical norms.”**

Does the repository?

- Provide enough information in terms of:
  - Identification of the Data Producer and their organisation
  - Reputation of the depositor
  - References to related publications
  - Information regarding the methods and techniques used, including those for data collection.





# Data Producers: Data Formats

**“2. The *data producer* provides the data in formats recommended by the data repository.”**

- **Minimum:** We are in the implementation phase (3)
- **Outsourced:** No

Obsolete formats create a risk of unusable data

Preferred formats that a data repository can reasonably assure will remain readable and usable

Usually De-Facto Standards



# Data Producers: Data Formats

**“2. The *data producer* provides the data in formats recommended by the data repository.”**

**Does the Repository:**

- Publish a list of preferred formats?
- Complete quality control to ensure Data Producers adhere to the preferred formats?
- Use tools to check the compliance with official specifications of the formats?
- Have a standard approach to deposits in non-preferred formats?
- Request detailed information about file formats and creation tools/methods?



# Data Producers: Documentation

**“3. The data producer provides the research data together with the metadata requested by the data repository..”**

- **Minimum:** Fully implemented (4)
- **Outsourced:** No

The data repository specifies the level of producer-created metadata required and provides the tools for its effective capture

- *Descriptive metadata*
- *Structural metadata*
- *Administrative metadata data.*



# Data Producers: Documentation

**“3. The data producer provides the research data together with the metadata requested by the data repository..”**

Does the repository:

- Offer deposit forms and/or other user-friendly ways to submit metadata?
- Have quality control checks to validate the metadata provided?
- Are there tools to create metadata at the file level?
- Use established metadata standards, registries or conventions?
  - Show the level of adherence to those standards
- Ensure the metadata provided are relevant for the data consumers?

What is the repository's approach if the metadata provided are insufficient for long term preservation?



# Data Repositories: Organisation and processes

“Organisations that play a role in digital archiving and are establishing a *Trusted Digital Repository* minimally possess a sound, long-term financial, organisational and legal basis”



# Data Repositories: Organisation and processes

**“4. The data repository has an explicit mission in the area of digital archiving and promulgates it.”**

- **Minimum:** Fully implemented (4)
- **Outsourced:** Yes



# Data Repositories: Organisation and processes

**“4. The data repository has an explicit mission in the area of digital archiving and promulgates it.”**

Does the Repository

- Have a Mission Statement?
- Describe how the Mission Statement is implemented?
- Carry out related promotional activities?
- Have a succession plan in place for its digital assets?

Or if applicable

- To which TDR have you outsourced?



# Data Repositories: Organisation and processes

**“5. The data repository uses due diligence to ensure compliance with legal regulations and contracts. ”**

- **Minimum:** Fully implemented (4)
- **Outsourced:** No

This guideline relates to the legal regulations which impact on the repository.





# Data Repositories: Organisation and processes

**“5. The data repository uses due diligence to ensure compliance with legal regulations and contracts. ”**

**Does the Repository:**

- Exist as a legal entity? Please describe its legal/organisational status.
- Use model contract(s) with Data Producers?
- Use model contract(s) with Data Consumers?
- Publish conditions of use?
- Have procedures for breaches of conditions?
- Ensure knowledge of and compliance with national and international laws? How?
- Have trained staff and procedures for data with disclosure risk including:
  - Review (including anonymisation and/or provision of secure access)storage
  - Secure access



# Data Repositories: Organisation and processes

**“6. The data repository applies documented processes and procedures for managing data storage.”**

- **Minimum:** Fully implemented (4)
- **Outsourced:** Yes

This guideline relates to the ability of the repository to manage archival storage.



# Data Repositories: Organisation and processes

**“6. The data repository applies documented processes and procedures for managing data storage.”**

**Does the repository:**

- Have a preservation policy?
- Have a strategy for backup / multiple copies? please describe.
- Are data recovery provisions in place? What are they?
- Are risk management techniques used to inform the strategy?
- Are there checks on the consistency of the Archival Storage ?
- What levels of security are acceptable for the repository?
- How is deterioration of storage media handled and monitored?

**Or if applicable**

- To which TDR have you outsourced?



# Data Repositories: Organisation and processes

**“7. The data repository has a plan for long-term preservation of its digital assets.”**

- **Minimum:** We are in the implementation phase (3)
- **Outsourced:** Yes

This guideline relates to the provision of continued access to data.



# Data Repositories: Organisation and processes

**“7. The data repository has a plan for long-term preservation of its digital assets.”**

- Are there provisions in place to take into account the future obsolescence of file formats? Please describe.
- Are there provisions in place to ensure long-term data usability? Please describe.

Or if applicable

- To which TDR have you outsourced?



# Data Repositories: Organisation and processes

**“8. Archiving takes place according to explicit workflows across the data life cycle.”**

- **Minimum:** We are in the implementation phase (3)
- **Outsourced:** Yes

This guideline relates to the levels of procedural documentation for the repository.



# Data Repositories: Organisation and processes

**“8. Archiving takes place according to explicit workflows across the data life cycle.”**

Does the repository:

- have procedural documentation for archiving data? If so, provide references to:
  - Workflows
  - Decision-making process for archival data transformations
  - Skills of employees
  - Types of data within the repository
  - Selection process
  - Approach towards data that do not fall within the mission
  - Guarding privacy of subjects, etc.
  - Clarity to data producers about handling of the data

Or if applicable

- To which TDR have you outsourced?



# Data Repositories: Organisation and processes

**“9. The data repository assumes responsibility from the data producers for access to and availability of the digital objects.”**

- **Minimum:** Fully implemented (4)
- **Outsourced:** No

This guideline relates to the levels of responsibility which the repository takes for its data.





# Data Repositories: Organisation and processes

**“9. The data repository assumes responsibility from the data producers for access to and availability of the digital objects.”**

**Does the repository:**

- have licences / contractual agreements with data producers? Please describe.
- Does the repository enforce licences with the data producer? How?
- Does the repository have a crisis management plan? Please describe.



# Data Repositories: Organisation and processes

**“10. The data repository enables the users to utilize the data and refer to them.”**

- **Minimum:** We have a theoretical concept (2)
- **Outsourced:** No

This guideline relates to the formats in which the repository provides its data and its identifiers.



# Data Repositories: Organisation and processes

**“10. The data repository enables the users to utilize the data and refer to them.”**

- Are data provided in formats used by the designated community? In what forms?
- Does the repository offer search facilities?
  - Is OAI harvesting permissible?
  - Is deep searching possible?
- Does the repository offer persistent identifiers?



# Data Repositories: Organisation and processes

**“11. The data repository ensures the integrity of the digital objects and the metadata.”**

- **Minimum:** We are in the implementation phase (3)
- **Outsourced:** No

This guideline relates to the information contained in the digital objects and metadata and whether:

- it is complete
- all changes are logged
- intermediate versions are present in the archive .



# Data Repositories: Organisation and processes

**“11. The data repository ensures the integrity of the digital objects and the metadata.”**

- Does the repository utilise checksums? What type? How are they monitored?
- How is the availability of data monitored?
- Does the repository deal with multiple versions of the data? If so, how? Please describe the versioning strategy.



# Data Repositories: Organisation and processes

**“12. The data repository ensures the authenticity of the digital objects and the metadata.”**

- **Minimum:** We are in the implementation phase (3)
- **Outsourced:** No

This guidelines relates to the relationship between the original data and that disseminated :

- the degree of reliability of the original
- the provenance of the data
- Maintenance of existing relationships/links for data and metadata



# Data Repositories: Organisation and processes

**“12. The data repository ensures the authenticity of the digital objects and the metadata.”**

. Does the repository:

- have a strategy for data changes? Are data producers made aware of this strategy?
- Does the repository maintain provenance data and related audit trails?
- Does the repository maintain links to metadata and to other datasets, and if so, how?
- Does the repository compare the essential properties of different versions of the same file? How?
- Does the repository check the identities of depositors?



# Data Repositories: Technical Infrastructure

**“13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.”**

- **Minimum:** We are in the implementation phase (3)
- **Outsourced:** Yes

The technical infrastructure constitutes the foundation of a Trusted Digital Repository. The OAIS reference model, an ISO standard, is the de facto standard for using digital archiving terminology and defining the functions that a data repository fulfils





# Data Repositories: Technical Infrastructure

**“13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.”**

This guideline refers to the level of conformance with accepted standards.

- What standards does the repository use for reference?
- How are the standards implemented, Please note any significant deviations from the standard with explanations.
- Does the repository have a plan for infrastructure development? Please describe.

Or if applicable

- To which TDR have you outsourced?



# Data Consumers

The data consumer uses the digital research data in compliance with guidelines 14-16

The quality of the use of research data is determined by the degree to which the data can be used without limitation for scientific and scholarly research by the various target groups, while complying with certain applicable codes of conduct.

The open and free use of research data takes place within the relevant legal frameworks and the policy guidelines as determined by the relevant national authorities.

The data consumer is bound by relevant national legislation. The data repository may have separate access regulations, which include restrictions imposed by the laws of the country in which the data repository is located. Access regulations should be based on relevant international access standards (e.g., Creative Commons) as much as possible.

Most nations have legal frameworks relating to the ethical use and re-use of data. These frameworks range from the statutory — which protect the privacy of individuals — to formal codes of conduct which inform ethical issues. Repositories must be aware of these local legal frameworks and ensure that they are taken into account when providing data for re-use.



# Data Consumers

**“14. The data consumer must comply with access regulations set by the data repository.”**

- **Minimum:** Fully implemented (4)
- **Outsourced:** No



# Data Consumers

## **“14. The data consumer must comply with access regulations set by the data repository.”**

This guideline refers to the responsibility of the repository to create legal access agreements which relate to relevant national (and international) legislation and the levels to which the repository informs the data consumer about the access conditions of the repository.

- Does the repository use End User Licence(s) with data consumers?
- Are there any particular special requirements which the repository's holdings require?
- Are contracts provided to grant access to restricted-use (confidential) data?
- Does the repository make use of special licences, e.g., Creative Commons?
- Are there measures in place if the conditions are not complied with?



# Data Consumers

**“15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.”**

- **Minimum:** Fully implemented (4)
- **Outsourced:** No

This guideline refers to the responsibility of the repository to inform data users about any relevant codes of conduct.



# Data Consumers

**“15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.”**

- Does the repository show awareness of and apply appropriate codes of conduct?
  - Including those designed for protection of human subjects?
- What are the terms of use to which data consumers agree?
- Are institutional bodies involved?
- Are there measures in place to address breaches of a code
- Does the repository provide guidance in the responsible use of confidential data?



# Data Consumers

**“16. The data consumer respects the applicable licences of the data repository regarding the use of the data.”**

- **Minimum:** Fully implemented (4)
- **Outsourced:** No

This guideline refers to the responsibility of the repository to inform data users regarding the applicable licences.



# Data Consumers

**“16. The data consumer respects the applicable licences of the data repository regarding the use of the data.”**

- Are there relevant licences in place?
- Are there measures in place to address licence breaches





# A Work in Progress

These Guidelines and their implementation are a work in progress which will evolve as further DSA assessments are performed, we welcome your professional insight into this evolution either as a member of the DSA community or by directing your comments to [info@datasealofapproval.org](mailto:info@datasealofapproval.org).



## And another thing...

Experts with expertise vs Evidence. What can be audited?

Not crushing people with bureaucracy

The initial effort vs the long haul

Academic Theory vs Best Practice in the real world

The Distributed Repository, Outsourcing & the Cloud



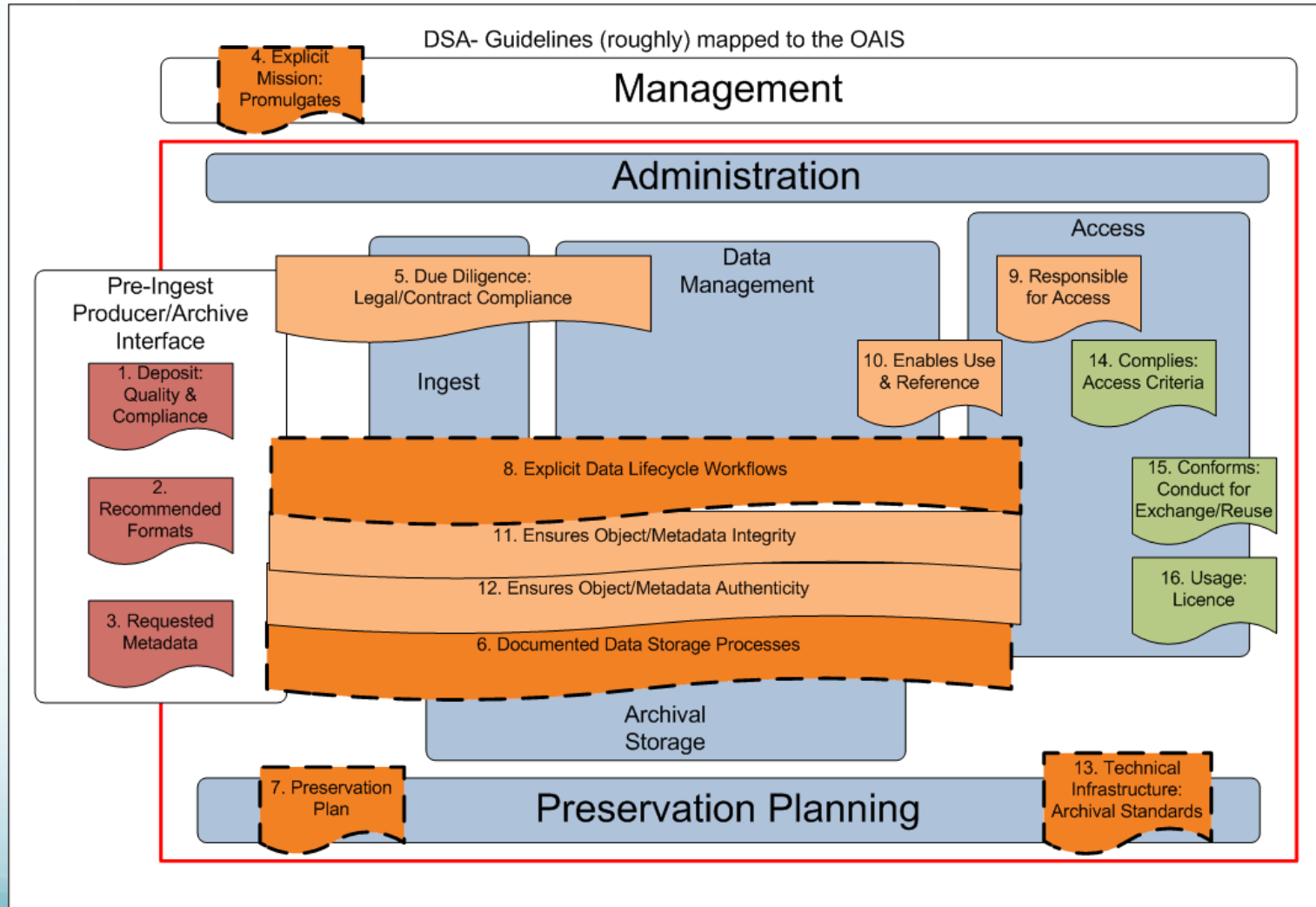
## Outsourcing

In the original version of the DSA outsourcing to third parties was permitted for Guidelines 4, 6, 7, 8 and 13 as long as the outsource partner had a DSA or better level of trust certification.

To take account of the increasingly distributed and service-based nature of modern repositories, the DSA Board expanded the possibility of outsourcing to all Guidelines. This decision will be monitored over time and may be amended in future in cooperation with the DSA community. Applicant information relevant to outsourcing is requested in the 'Repository Context' section and must form part of the evidence for each applicable Guideline.



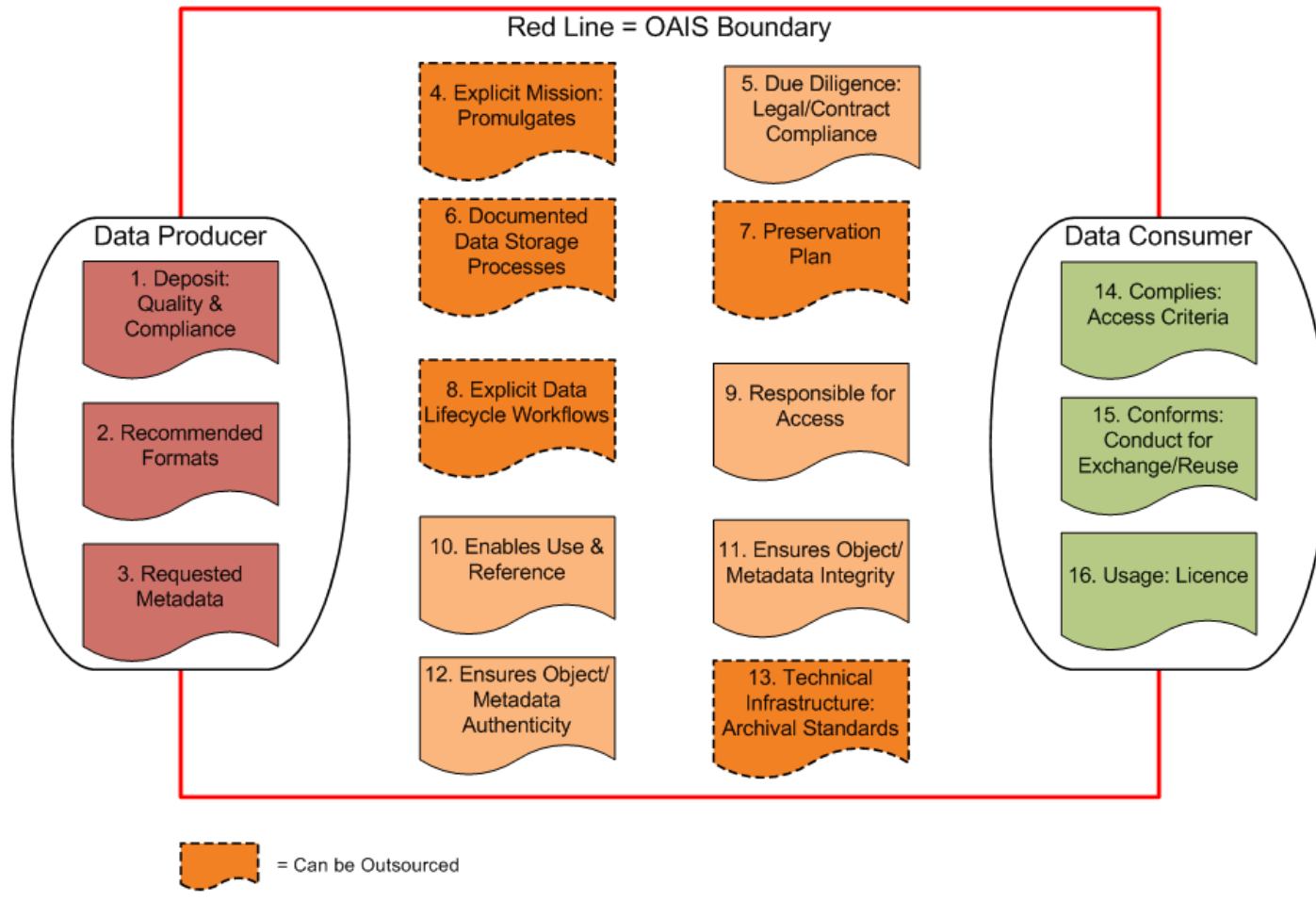
# The 16 Guidelines & the OAIS





# The 16 Guidelines & the OAIS

DSA- Guidelines (roughly) mapped to the OAIS Boundary





# The 16 Guidelines

## Producer

- 1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data and compliance with disciplinary and ethical norms.*
- 2. The data producer provides the data in formats recommended by the data repository.*
- 3. The data producer provides the data together with the metadata requested by the data repository.*



# The 16 Guidelines

## *Repository*

- 4. The data repository has an explicit mission in the area of digital archiving and promulgates it.*
- 5. The data repository uses due diligence to ensure compliance with legal regulations and contracts.*
- 6. The data repository applies documented processes and procedures for managing data storage.*
- 7. The data repository has a plan for long-term preservation of its digital assets.*
- 8. Archiving takes place according to explicit workflows across the data life cycle.*



# The 16 Guidelines

## *Repository*

- 9. The data repository assumes responsibility from the data producers for access to and availability of the digital objects.*
- 10. The data repository enables the users to utilize the data and refer to them.*
- 11. The data repository ensures the integrity of the digital objects and the metadata.*
- 12. The data repository ensures the authenticity of the digital objects and the metadata.*
- 13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.*





# The 16 Guidelines

## *Consumer*

*14. The data consumer must comply with access regulations set by the data repository.*

*15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.*

*16. The data consumer respects the applicable licences of the data repository regarding the use of the data.*