

# Data Quality Assurance from the Perspective of CoreTrustSeal

re3data COREF / CoreTrustSeal Workshop on Data Quality  
Management at Repositories

2022-10-05

Hervé L'Hours

UK Data Service, UK Data Archive,

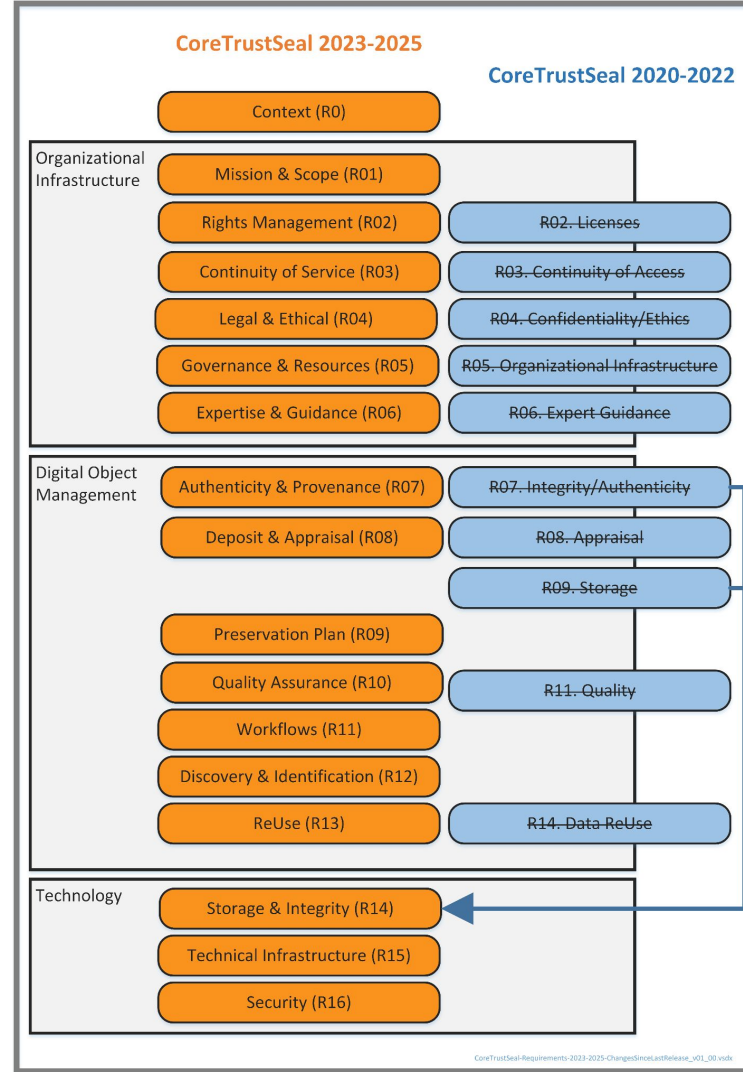
University of Essex

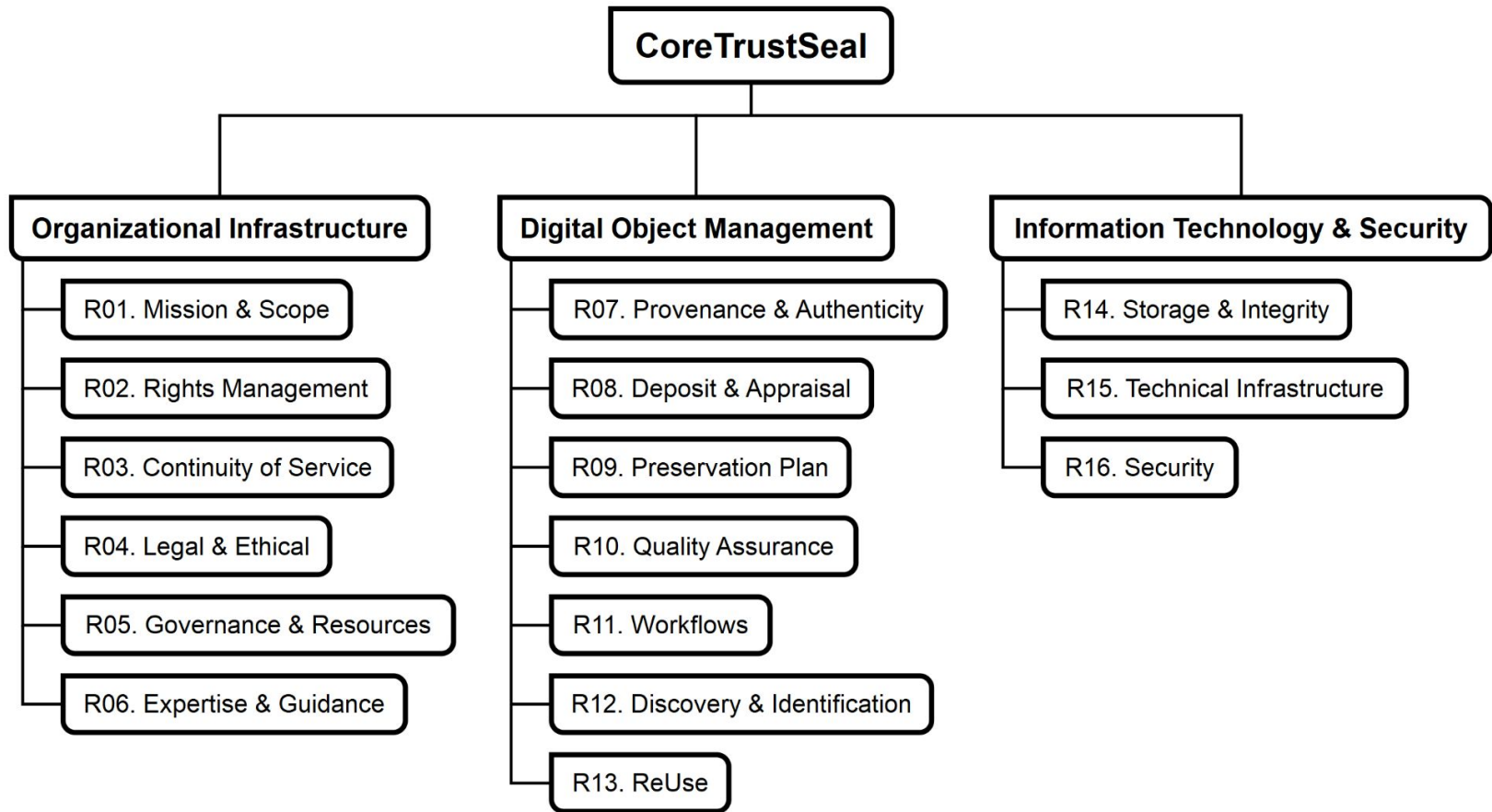
Vice-Chair CoreTrustSeal Board



# 2023-25

- From 2023-01
- Terminology & Structure
- Evidence Stable
- Clarification
- Deduplication





# Data Quality Assurance?

vs. **(meta)**data Quality Assurance

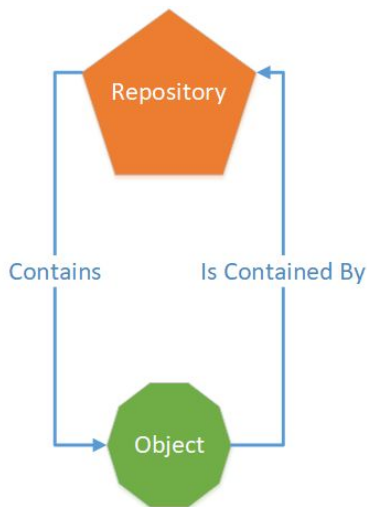
vs. **Digital Object** Quality Assurance

- Data
- Metadata (more structured)
- Documentation (less structured)

cf: Linked Data/RDF and the limitations of digital object boundaries and files.

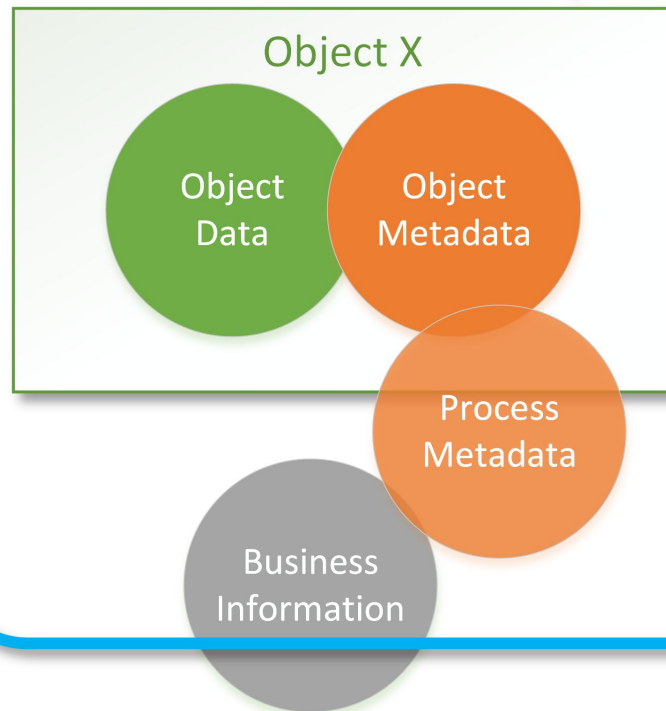
# CoreTrustSeal

Artefacts that support practice..  
...provide evidence...  
...to demonstrate compliance.

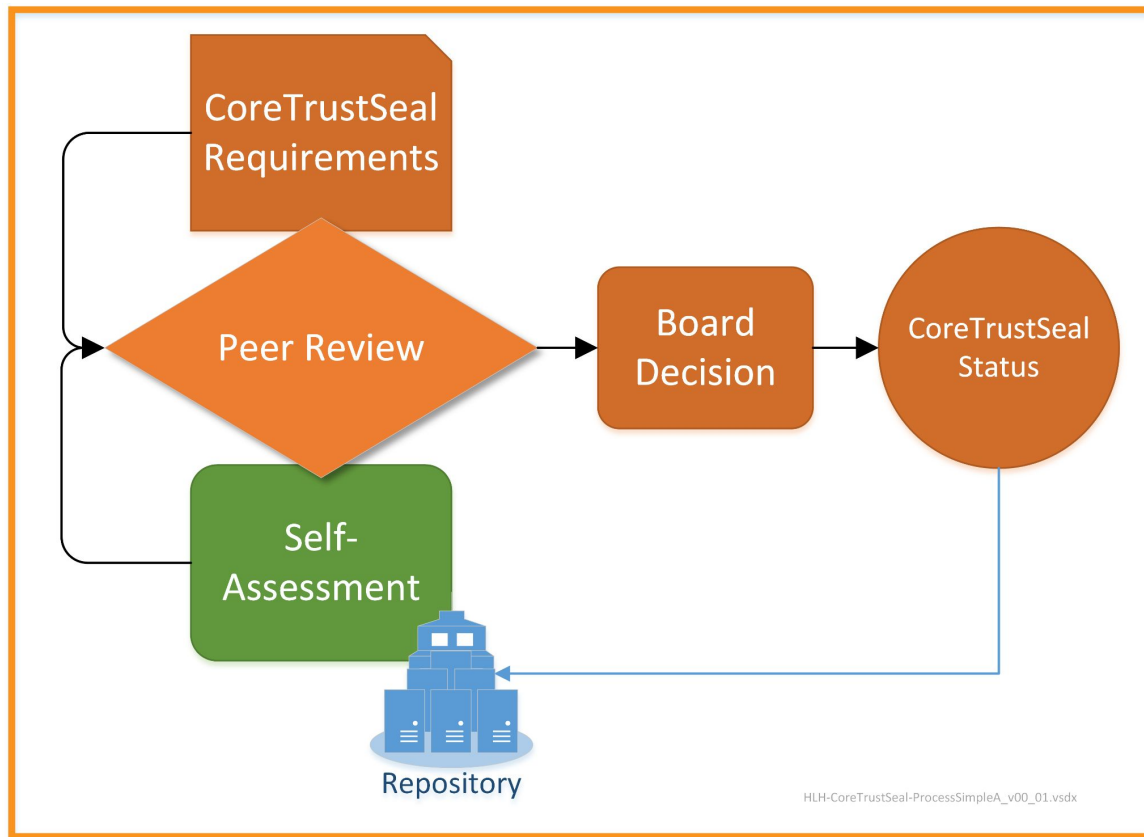


Repository ObjectMetadata A\_00\_01.vsd

Repository...



# CoreTrustSeal Review



# CoreTrustSeal Perspective

Quality in terms of:

## Repository Characteristics

- Organisational
- Technical-Security
- Digital Object Management
  - Collection Scope
  - *Quality Indicators*
  - *Quality Criteria*
  - *Assessment of (meta)data*
  - *Curation*
  - *Review* (appraisal & re-appraisal)
  - Feedback mechanism (data reUsers)

Digital Object Characteristics (implied)



# CoreTrustSeal Perspective

Quality in terms of:

Repository Focus

- Generic
- Discipline or Domain

CoreTrustSeal will seek appropriate evidence,  
but remains domain-agnostic.

Repository as **enabler** and **communicator** of quality vs  
**guarantor** of quality.





# Quality Assurance (R10)

The repository addresses:

- **technical quality** and **standards compliance**
- ensures that sufficient **information** is **available** for **end users** to make **quality-related evaluations**.

Different repositories, different levels of curation on data and metadata and documentation- depending on the expectation of their depositors and their **Designated Community** of users.



# Quality Assurance (R10)

**Technical quality and standards compliance.**

Comply with a range of standard criteria including acceptable formats, metadata schema, metadata content and links to other digital objects.

# Quality Assurance (R10)

**Sufficient information about the digital objects for the Designated Community to assess their fitness for use.**

Not (necessarily) an evaluation of “scientific quality” of the original objects prior to deposit.

Data, or associated metadata, may have quality issues relevant to their research value, but this does not preclude their use if a user can make a well-informed decision on their suitability through provided documentation.

Quality assessment becomes increasingly relevant when the Designated Community is multidisciplinary, where users may not have the personal experience to make an evaluation of quality from the data alone.



# Quality Assurance (R10)

- “The approach to data and metadata quality taken by the repository including variations for different curation-levels.
- The standards that data, metadata and documentation must comply with to be acceptable for preservation and access. Whether these are general external standards, internally developed standards or specific to a community of practice.
- The quality control checks in place ensure the completeness and understandability of data and metadata.
- The approach to resolving issues e.g. whether the digital objects are returned to the depositor for rectification, fixed by the repository, noted by quality flags, and/or included in the accompanying metadata.
- The approach to managing changes to expected standards (e.g. new or updated data formats of metadata schemas) in response to changes in the technical environment or to changes in the needs of the Designated Community.
- Any links provided to other digital objects’ data and metadata e.g. related digital objects, publications, or the use of controlled vocabularies and ontologies.”



# Quality Definition

- *Formats* (but cf: Linked open data & RDF)
- *Collection policies*
- *Deposit guidelines*
- *Quality checklists*
- **Repository Mission/Scope (R01)**
- **Designated Community (R0)**
- **Workflows (R11)**

## **Formats:**

Acceptable and preferred for:

- **Deposit** (R08) so they can be used (directly, or converted) for
- **ReUse** (R13) and
- as the basis for active long term **Preservation** (R09)



# Quality Development

- *Templates*
  - **Workflows (R11)**
- *Individual support*
  - < specific focus in CoreTrustSeal
- *Training*
  - < specific focus in CoreTrustSeal

Points of Contact:

Depositor: **Deposit & Appraisal (R08)**  
End User: **Reuse (R13)**



# Quality Control

- *Technical routines* **Workflows (R11)**
- *Assessment and review* **Quality Assurance (R11)**
- *Quality monitoring* **Preservation Plan (R09)**
  - Changes in expectations of technical & scientific quality
  - Changes in the Designated Community
  - Technical Watch, Community Watch, ReAppraisal

# Quality Improvement

*Curation measures to improve data quality?*

## Workflows (R11)

- *standard data curation processes*
- *metadata enrichment*
- *data use specifications*

Curation vs Preservation:  
Immediate vs long term perspective?





# Quality Evaluation

## Quality Evaluation Procedures

- **Quality Assurance (R11)**
- **Workflows (R11)**
- *quality ratings*
- *user surveys*
- *user reports*

Inconsistent across repositories,  
lack of agreed ideal practice

# Quality Documentation

## *Document Quality Measures and Results*

- *quality reports*
- *quality flags*
- *related documentation*
- Variability across repositories
- Quality as part of **Provenance and Authenticity (R07)**?

Lack of standardised model for data or metadata quality assessment in general or per discipline.

# CoreTrustSeal Perspective-Outcomes?

Quality in terms of:

## Repository Characteristics

- Organisational
- Technical-Security
- Digital Object Management
  - Collection Scope
  - Quality Indicators
  - Quality Criteria
  - Assessment of (meta)data
  - Curation
  - Review (appraisal & re-appraisal)
  - Feedback mechanism (data reUsers)

Digital Object Characteristics (implied)



# CoreTrustSeal Perspective-Outcomes?

- Absolute vs relative assessments of quality
  - Repository Mission & Scope
  - Designated Community
- “Scientific” vs. “Technical” Quality
- Enabler and communicator of quality vs guarantor of quality.
- Curation vs Preservation: “Now” vs long term quality perspective
- Lack of agreed ideal practice for quality assessment or outcomes
  - in general or per discipline.
- Machine actionable vs human mediated evaluation
- Quality Characteristics @Repository (re3data)
- Quality Characteristics @Object (DataCite)



# Thank You

Email: [info@coretrustseal.org](mailto:info@coretrustseal.org)

Website: <https://www.coretrustseal.org/>

