SERVIR West Africa 2025 Training Program on Data Stewardship and the CoreTrustSeal Requirements: Session 2, R09, R10 & R11 Requirements

Robert R. Downs, PhD

Center for International Earth Science Information Network (CIESIN)

Columbia Climate School, Columbia University

Coordinated by AFRIGIST

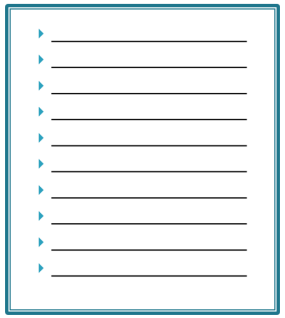
November 20, 2024, 13:00 – 14:30 GMT (8:00 - 9:30 a.m. EST)

Brief Review of Previous Webinars

- Data Stewardship Concepts, Principles, and Certification Instruments
  - CoreTrustSeal, ISO 16363, Nestor DIN 31644
- Preservation, Designated Community, Sustainability
- CoreTrustSeal Certification requirements
- Organizational aspects of Data Stewardship
- CoreTrustSeal Requirements 2020-2022: R01-R16
- Setting the stage for Planning to Meet the CoreTrustSeal Requirements
- Motivation and R0 Requirement, Background Information and Context
- Organizational Infrastructure Requirements: R01, R02, R03, R04, R05, R06
- Digital Object Management Requirements: R07, R08

SERVIR West Africa 2025 Training Program on Data Stewardship and CoreTrustSeal Requirements

- Key topic:
  - R09, R10, & R11 Requirements
- Instructional format
  - Lecture, questions, and discussion
- Session will be recorded
  - Recording will be posted on the SERVIR West Africa website



Outline for Current Session

- Progress Planning and Participant Introductions (5 minutes)
- Scheduling Training Sessions and CoreTrustSeal Application Preparation (5 minutes)
- Review of Homework: Draft Self-Assessment Responses in Template (10 minutes)
- CoreTrustSeal Requirement 09 and Examples of Responses (13 minutes)
- Homework: Draft Self-Assessment Responses (5 Minutes)
- CoreTrustSeal Requirement 10 and Examples of Responses (13 minutes)
- Homework: Draft Self-Assessment Responses (5 Minutes)
- CoreTrustSeal Requirement 11 and Examples of Responses (13 minutes)
- Homework: Draft Self-Assessment Responses (5 Minutes)
- Introduction to CoreTrustSeal Requirements R12 and R13 (11 Minutes)
- Preparing for Next Session (5)

Progress Planning and Introductions

- CoreTrustSeal application template created for each repository
  - CERSGIS, ICRISAT FMNR, AFRIGIST, AGRHYMET, CSE
- Each repository uses template to progressively revise their self-assessment
- Please put additional repository names and your email address in Zoom chat
- Brief introductions by those who have not previously introduced themselves

Scheduling Training Sessions and CoreTrustSeal Application Preparation

- Plan to schedule sessions twice monthly enabling time for progress
  - Repositories complete template sections between sessions ✓
  - Each session includes review and introduction to subsequent section ✓

- Next session planned for Wednesday, December 4, 2024, 13:00-14:30 GMT
  - Topics: R12 and R13 Requirements and Introduction to R14, R15, and R16 Requirements

Review repository self-assessments in templates:

Draft responses to

CoreTrustSeal Requirements

Review of Self-Assessments in CoreTrustSeal Application Templates

- R0 Background Information and Context
  - Re3data identifier, Repository Type, Overview, Designated Community, Level of Curation Performed
  - Cooperation and Outsourcing to third parties, partners, and host organizations
  - Applicants renewing – Summary of Significant Changes Since Last Application
- R01 Mission & Scope
  - The repository has an explicit mission to provide access to and preserve digital objects.
- R02 Rights Management
  - The repository maintains all applicable rights and monitors compliance.
- R03 Continuity of Service
  - The Repository has a plan to ensure ongoing access to and preservation of its data and metadata.
- R04 Legal & Ethical
  - The repository ensures to the extent possible that data and metadata are created, curated, preserved, accessed and used in compliance with legal and ethical norms.
- R05 Governance & Resources
  - The repository has adequate funding and sufficient numbers of staff managed through a clear system of governance to effectively carry out the mission.
- R06 Expertise & Guidance
  - The repository adopts mechanisms to secure ongoing expertise, guidance and feedback-either in-house, or external.

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Requirements 2023-2025 (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051012

Review of Self-Assessments in CoreTrustSeal Application Templates (Cont'd)

- R07 Provenance & Authenticity
  - The repository guarantees the authenticity of the digital objects and provides provenance information.
- R08 Deposit & Appraisal
  - The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for users.

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Requirements 2023-2025 (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051012

Questions about previous homework to compose draft responses in template

- Identifying resources that address a challenging requirement
  - Can we start listing and describing current resources that address the requirement?

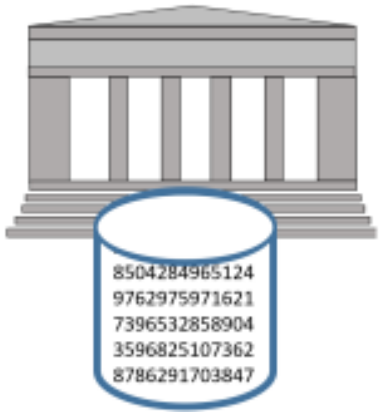
- Needed resources
  - What resources are needed to comply with a challenging requirement?

- Progress on CoreTrustSeal application templates for R0:
  - We previously featured examples: CERSGIS, AGRHYMET, and ICRISAT FMNR

- More progress is needed to feature more examples
  - Suggestions have been offered in the templates

- Let’s all continue making progress on the CoreTrustSeal application templates

CoreTrustSeal Requirement R09 Preservation Plan



CoreTrustSeal Requirement R09 Preservation Plan

R09. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Self-Assessed Compliance Level

The repository, depositors, and Designated Community need to understand the level of responsibility undertaken for the long-term preservation of data and metadata. Procedures must be documented and their completion assured.

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Trustworthy Digital Repositories Requirements 2023-2025 Extended Guidance (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051096

CoreTrustSeal Requirement R09 Preservation Plan

The response statement and evidence should include references to the following items:

- The documented approach to preservation, including whether this involves format migration, emulation, etc. Ensuring bit level integrity is vital but not sufficient for preservation Ensuring bit level integrity is vital but not sufficient for preservation.
- File formats and metadata schemas for long term preservation.
- How the level of responsibility for the preservation of each item is defined.
- Plans related to future migrations or similar measures to address the threat of obsolescence.
- Actions relevant to preservation specified in documentation, including custody transfer, submission information criteria, and preservation information metadata.
- Measures to ensure these actions are taken.
- Any minimum stated retention and/or preservation periods.
- How often the digital objects are re-appraised and the possible outcomes of reappraisal.
- The repository approach to deleting/removing data and metadata from collection/holdings including the impact on persistent identifiers.

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Trustworthy Digital Repositories Requirements 2023-2025 Extended Guidance (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051096

CoreTrustSeal Requirement R09 Preservation Plan

The rights of the repository, including the right to preserve, are covered under Rights Management (R02). Bit level integrity is covered under Storage and Integrity (R14). Acceptable file formats at deposit should be covered under Deposit and Appraisal (R08). Measures to ensure that file formats, schemas and content are appropriate to the Designated Community should be covered under Reuse (R13).

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Trustworthy Digital Repositories Requirements 2023-2025 Extended Guidance (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051096

Preservation Plan Example: SAEON CoreTrustSeal Application (Excerpt)

SAEON has a Preservation Policy[1] that is available on the Wiki.

✓ We have three preservation levels, one of which is currently in use and two which are available for future use. The first is our online storage that ensures the data is immediately accessible. The second is nearline storage of data that is moved to magnetic tape, which could be made available within 24 hours and the third is offline storage. We currently do not need to make use of the latter two, but as the repository expands we may need to. The access policy and workflow for these latter preservation levels will be determined prior to their use, as will the migration policy within the two levels.

✓ Institutional data providers that allow SAEON to host and archive data on their behalf enter into Service Level Agreements (SLAs) that detail the conditions of hosting and specify both parties' responsibilities within the agreement, including granting SAEON all necessary rights to properly curate and archive the data, including format conversions.

✓ Individual data producers that submit data to SAEON are required to comply with the conditions of the SAEON data policy[2] that similarly specifies the responsibilities of both parties as well as granting SAEON the necessary rights to manage the data, including format migrations.

✓ The transfer of custody and responsibilities handover, as well as a description of what falls under the data users' responsibility and what the curators are responsible for, is described on both the standard operating procedures[3] and the data submission guidancen pages on the Wiki.

✓ Format conversions will be addressed as needed to avoid data obsolescence based on the accepted standards of our Designated Community.

Source: South African Environmental Observation Network (SAEON), "2026-09-04 - SAEON Open Data Platform - CoreTrustSeal Requirements 2020-2022", https://doi.org/10.34894/CZ1JFL, DataverseNL, V2

Preservation Plan Example: SunScholarData CoreTrustSeal Application (Excerpt 1 of 2)

SunScholarData is the institutional research data repository of Stellenbosch University in South Africa.

The repository is in the implementation phase - 3

- ✓ The most detailed approach to preservation is documented in the SUNScholarData Preservation Plan. This is a living document which maps out the plan for preserving Stellenbosch University's research data. It is worth pointing out that not all of the aspects covered in the plan have been implemented.
- ✓ Whilst there are several aspects which have actually been implemented there are some other aspects that can only be implemented in the future. The preservation plan lays the foundation for the treatment of the following core aspects of data preservation: (1) objectives, (2) scope, (3) file formats, (4) legal requirements, (5) regulatory requirements, (6) preservation strategy, (7) preservation planning, (8) monitoring and review. The first five of these core aspects have already been implemented whilst the remaining listed aspects have yet to be implemented.
- ✓ The preservation plan contains a broad outline of the preservation strategies that could be followed in the future. This relates to aspects such as: (1) data ingestion, (2) timely access, (3) archival storage, (4) integrity and security of archived content, (5) monitoring of community preservation activities, best practices and standards, (6) development of local preservation planning activities as well as, (7) the development of preservation planning actions in accordance with the Open Archival Information System (OAIS) Reference Model.
- ✓ Insofar as the more detailed aspects of preservation planning are concerned, the document adopts the tenets of the preservation planning functional entity as set out in the Open Archival Information System (OAIS) Reference Model (2012). Consequently, the document addresses the following aspects: (1) the monitoring of the designated community, (2) the monitoring of technology, (3) the development of preservation strategies and standards, (4) the development of packaging designs and migration plans as well as, (5) the administration of preservation planning.
- ✓ Lastly, the document provides some details about the measures that were devised in order to monitor and review the preservation plan.

Source: SUNScholarData, 2024, - CoreTrustSeal Requirements 2020-2022", https://doi.org/10.34894/D0FFK8, DataverseNL, V2

Preservation Plan Example: SunScholarData CoreTrustSeal Application (Excerpt 2 of 2)

SunScholarData is the institutional research data repository of Stellenbosch University in South Africa.

- The Library and Information Service has adopted a holistic approach to bit preservation. This approach encompasses bit preservation in its entirety as well as certain aspects of functional preservation. Although SUNScholarData does implement some aspect of functional preservation these are not yet sufficient to classify the preservation approach as true digital preservation. The details pertaining to the manner in which preservation is addressed by the Library and Information Service is set out in the SUNScholarData Preservation Approach.
- The Library and Information Service has assessed SUNScholarData's preservation activities against the Levels of Digital Preservation (LoP). The levels that are currently addressed range from levels 1-4. The LoP assessment tool was utilised in order to assess five functional areas namely: storage, SUNScholarData integrity, control, metadata and content.
- The results enabled the LIS to understand which preservation activities are currently being addressed in a holistic manner. The LoP assessment revealed a number of gaps in SUNScholarData's approach to preservation. A number of steps will need to be taken in order to address the existing gaps. Plans to address this issue in the future have been set out in detail in the SUNScholarData Digital Preservation Roadmap.

SUNScholarData Preservation Roadmap:

- A high-level approach for addressing the digital preservation of research data at Stellenbosch University is set out in the SUNScholarData Preservation Roadmap. It provides direction for the future program development and implementation of the University's digital preservation initiatives. This information shed light on the manner in which digital preservation will be implemented. Emphasis is placed on three phases: information gathering, development and implementation. The preservation roadmap should be read together with the SUNScholarData Preservation Plan.

Source: SUNScholarData, 2024, - CoreTrustSeal Requirements 2020-2022", https://doi.org/10.34894/D0FFK8, DataverseNL, V2

Preservation Plan Example: DataFirst CoreTrustSeal Application

DataFirst, the University of Capetown domain data repository, shares research data on African countries.

The guideline has been fully implemented in the repository - 4

Our repository's digital preservation plan is our Digital Curation and Preservation Document …

The plan documents the principles, policies and strategies followed for data and metadata management at DataFirst. Recording our digital preservation guarantees that DataFirst's digital preservation practices are transparent to staff to ensure ongoing adherence to the plans to maintain our standing as a Trusted Digital Repository. Preservation documentation also enables data Depositors and data users to understand the services we provide. Depositors can also be clear on what we offer in terms of curating and disseminating their datasets.

The Digital Curation and Preservation Document lays out standards adopted by the repository for accepting data deposits (SIPs) archiving (AIPs), preparing, and disseminating data (DIPs). It also covers how we ensure useful data documentation to support analysis of the data by our user community. The preservation plan must clearly articulate responsibilities for preserving digital objects. Our preservation plans are a guide for repository staff and provide an overview of our work for stakeholders.

The Digital Curation and Preservation Document includes details of the Memorandum of Agreement we sign with each Depositor. As documented in the Digital Curation Document, the MOA we sign with Depositors clearly spells out the rights and responsibilities of the Depositor and DataFirst's repository team, as parties to the Agreement. The MOA covers actions of both parties necessary to meet the responsibilities documented in the agreement. Under the MOA, Depositors agree on data transfer date and access conditions, including the open copyright license to be applied to their data. …

Source: DataFirst, 2024, "2026-12-06 - DataFirst - CoreTrustSeal Requirements 2020-2022", https://doi.org/10.34894/9EWIVW, DataverseNL, V2

R09 Preservation Plan Example: WDCC CoreTrustSeal Application

The German Climate Computing Center (DKRZ), Hamburg, Germany, operates the World Data Center for Climate (WDCC).

Implemented: the requirement has been fully implemented by the repository - 1

Since 2003, WDCC has been providing long-term preservation of data and metadata for the climate science community, including an optional DataCite DOI data publication. Over the years, WDCC has continuously refined its preservation strategy, aimed at maintaining and enhancing the FAIRness of WDCC data publications. This strategy involves regularly updating metadata, adopting advancements in data preservation practices, and maintaining the integrity of archived data through continuous backup and monitoring. Detailed descriptions of this established strategy can be found in WDCC Preservation and Storage Policy, WDCC Preservation Plan and Workflow, and WDCC Risk Management.

The WDCC employs a standardized preservation approach for all digital assets, ensuring that all data are archived with the same preservation level. It primarily preserves non-proprietary file formats recognized in the climate science community and strongly recommends using NetCDF or WMO GRIB formats. NetCDF, a file format and a set of open software libraries designed for sharing scientific data in a backward-compatible manner, is the format used for the vast majority of files archived in WDCC. If WDCC accepts non-recommended file formats, it guarantees that at least one version of the data is archived in an open-source format. Once data files are archived in the WDCC, they remain unaltered, ensuring the preservation and integrity of the data. As explained in R08, WDCC's commitment to backward-compatible file formats makes format migration obsolete.

Source: World Data Center for Climate (WDCC) 2024 - CoreTrustSeal Requirements 2023-2025", https://doi.org/10.34894/LWLCTT, DataverseNL, V1

R09 Preservation Plan Homework: Self-Assessment in Template

- Describe the documented approach to preservation (format migration, emulation, etc.).
- What file formats and metadata schemas are utilized for long term preservation?
- How is the level of responsibility for the preservation of each item defined?
- Describe plans for future migrations or similar measures to address the threat of obsolescence.
- What documentation specifies actions relevant to preservation, including custody transfer, submission information criteria, and preservation information metadata?
- What are the measures to ensure that preservation actions are taken?
- What are the minimum stated retention and/or preservation periods?
- How often are the digital objects re-appraised and what are the possible outcomes of reappraisal?
- What is the approach to deleting/removing data and metadata from collection/holdings and the impact on persistent identifiers?

CoreTrustSeal Requirement R10 Quality Assurance

CoreTrustSeal Requirement R10 Quality Assurance

R10. The repository addresses technical quality and standards compliance, and ensures that sufficient information is available for end users to make quality-related evaluations.

Self-Assessed Compliance Level:

Different repositories undertake different levels of curation on data, metadata and documentation depending on the needs and expectations of their depositors and Designated Community. Quality assurance by the repository ensures that digital objects comply with a range of standard criteria including acceptable formats, metadata schema, metadata content and links to other digital objects. This relates to 'technical quality' rather than the 'scientific quality' of the original digital objects creation or collection prior to deposit, though the repository must ensure there is sufficient information about the digital objects for the Designated Community to assess their fitness for use. 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.

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Trustworthy Digital Repositories Requirements 2023-2025 Extended Guidance (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051096

CoreTrustSeal Requirement R10 Quality Assurance

The response statement and evidence should include references to the following items:

- 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.

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Trustworthy Digital Repositories Requirements 2023-2025 Extended Guidance (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051096

CoreTrustSeal Requirement R10 Quality Assurance

This Requirement refers to data and metadata quality standards and assurance during curation. Selection criteria are covered during Deposit and Appraisal (R08). Measures to ensure that digital objects remain fit for purpose over time are covered under Preservation Plan (R09).

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Trustworthy Digital Repositories Requirements 2023-2025 Extended Guidance (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051096

R10 Quality Assurance Example: WDCC CoreTrustSeal Application (Excerpt)

The German Climate Computing Center (DKRZ), Hamburg, Germany, operates the World Data Center for Climate (WDCC).

Implemented: the requirement has been fully implemented by the repository - 1

Prior to any data publication in the WDCC, the data depositor has to contact the WDCC curator team and provide a short description of the planned preservation. When the request for preservation is accepted, a WDCC curator takes over the responsibility for this request and will guide the data depositor through the entire preservation workflow (see WDCC Preservation Plan and Workflow). Metadata are provided by the data depositor. Additional metadata like information about the methodology used to create the data and references, like scientific publications using the data, are optional, but recommended by WDCC. Before metadata are ingested into the WDCC relational database (Oracle), the data curator checks all metadata against the WDCC metadata scheme and regarding their overall understandability. For all DataCite DOI data publications, these optional metadata are checked by the data curators. After the metadata ingestion, all collected metadata become accessible in the WDCC metadata scheme on the data publication landing page. These include links to previous/subsequent data publication versions in the WDCC, describing scientific publications or other references.

WDCC strongly recommends the usage of NetCDF (selfdescribing file format) and encourages data depositors to standardize the NetCDF files according to Climate Forecast (CF) metadata conventions, which is the most established community standard in climate science. The CF Conventions provides a comprehensive controlled vocabulary for variables used in Climate and Earth System Sciences. The adherence to the CF conventions ensures the understandability of data by the provision of rich and standardized file metadata. …

Source: World Data Center for Climate (WDCC) 2024 - CoreTrustSeal Requirements 2023-2025", https://doi.org/10.34894/LWLCTT, DataverseNL, V1

R10 Quality Assurance Example: Data@UNIMI CoreTrustSeal Application (Excerpt)

Data@UNIMI is the open research data repository for the University of Milan, Italy.

Compliance level: Implemented: the requirement has been fully implemented by the repository - 1

At the University of Milan, we value research data quality as a major asset (see, for instance, our presentation about “Institutional metadata & data quality check” in [27] GenOAWeek2022). Specifically, Data@UNIMI follows the Dublin Core standard schema for metadata (see [48] Dublin Core), whilst strictly complying with the FAIR principles, as reported in the RDM policy of the University (see [1] RDM policy). As so, the quality check of data, of metadata and of the FAIRness of the datasets deposited in Data@UNIMI is operated through an automatic tool (see [39] Data@UNIMI quality tool) and with a final double review, carried out by human resources. Human curation is performed by data stewards and central staff, from now on called the curators. The responsibility of the quality check is ultimately left to human skill, whilst the automatic check is just to support.

The curation levels are essentially two:

- The data steward disciplinary curation activity, which is more discipline-oriented and can better apply all the specific measures that the discipline will support and recommend.

- The central staff curation activity, which is the very last step of curation before publication. …

Source: 2027-10-22 - Data@UNIMI - CoreTrustSeal Requirements 2023-2025, https://doi.org/10.34894/F1XDX5, DataverseNL, V1

R10 Quality Assurance Example: PURR CoreTrustSeal Application (Excerpt)

Compliance level: Implemented: the requirement has been fully implemented by the repository - 1

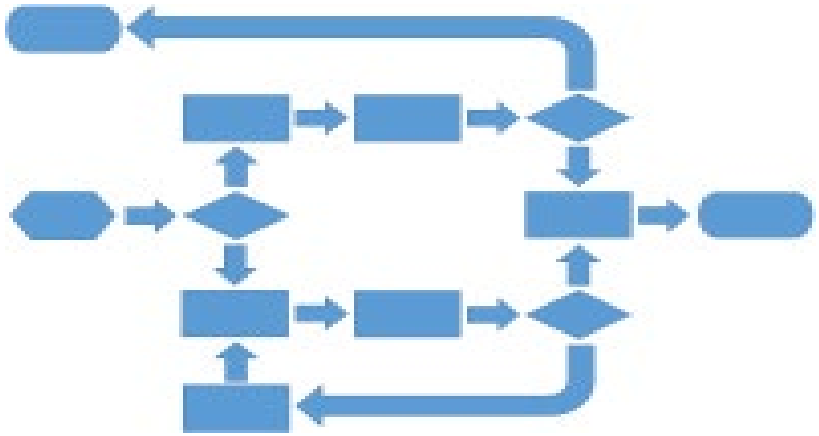
✓ PURR assures its Designated Community that the digital objects it holds are effectively preserved into the future. PURR's policies aim to make preserved information available and usable. PURR's preservation activities listed in the PURR Preservation Strategic Plan [9] address the preservation of the content of datasets published on PURR which includes the submitter-supplied descriptive metadata and files. It also addresses risks to long term access including media degradation, format and software obsolescence, human error, and natural disaster. …
✓ Communication with the producer/submitter of data during the publication process is important for making certain that dataset publication proceeds as expected and without error. … The producer/submitter is informed if errors occur in data or metadata during the publication workflow. Once a draft of publication is ready the submitter submits the dataset for review. … The review starts within two business days of dataset submission. The reviewer reviews the dataset for completeness of data and metadata and provides comments for each component of the dataset that it requires. … If a requirement is satisfied a green check is placed next to the component. ✓ Building these notifications into the publication workflow helps to make sure that lapses in communication have not affected their submission or caused a loss of the data publication or any of its contents ...

Source: PURR, 2024, "2027-08-08 - PURR - CoreTrustSeal Requirements 2023-2025", https://doi.org/10.34894/0QSAJN, DataverseNL, V1

R10 Quality Assurance Homework: Self-Assessment in Template

- What is the approach to data and metadata quality taken by the repository, including variations for different curation-levels?
- What standards, external or internally developed, must the data, metadata and documentation comply with to be acceptable for preservation and access?
- What quality control checks are in place to ensure the completeness and understandability of data and metadata?
- Describe 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.
- What is 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?
- Describe the links that are provided to other digital objects' data and metadata e.g. related digital objects, publications, or the use of controlled vocabularies and ontologies.

CoreTrustSeal Requirement R11 Workflows



CoreTrustSeal Requirement R11 Workflows

R11. Digital object management takes place according to defined workflows from deposit to access.

Self-Assessed Compliance Level:

For Quality Assurance (R10) to be achieved, it is necessary to avoid ad hoc actions and to deliver consistency of practice for all digital objects and across repository functions. This requires that workflows be defined, documented, and change-managed. Workflows may be specified in a mixture of standard operating procedures, business process descriptions and diagrams that guide normal practice and provide mechanisms for handling exceptions.

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Trustworthy Digital Repositories Requirements 2023-2025 Extended Guidance (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051096

CoreTrustSeal Requirement R11 Workflows

The response statement and evidence should include references to the following items:

- Workflows/business process descriptions covering the curation levels performed.
- How workflows are adjusted for different types of data and metadata.
- Decision handling within the workflows.
- Change management of workflows.
- Ability to track workflow execution, with mechanisms to handle exceptions.

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Trustworthy Digital Repositories Requirements 2023-2025 Extended Guidance (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051096

CoreTrustSeal Requirement R11 Workflows

This Requirement confirms that all workflows are documented. It should be noted if there are different workflows for different levels of security mentioned in the Legal and Ethical (R04) response statement. Workflows may include qualitative and quantitative checking of outputs, but any detail on checks and compliance should be addressed under Quality Assurance (R10).

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Trustworthy Digital Repositories Requirements 2023-2025 Extended Guidance (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051096

Workflows Example: SAEON CoreTrustSeal Application (Excerpt)

The repository is in the implementation phase - 3

✓ The operational framework of the SAEON ODP is based on the SAEON Data Policy[1], the SAEON Preservation Policy[2] and the SAEON ODP Standard Operating Procedures[3]. The data workflows that have been developed are compliant with these policies and procedures and make use of the Open Archival Information System (OAIS) reference model as a best practice standard to work towards.

✓ The data submissions guidance page on the Wiki walks data providers through the process of submitting data and is linked to from the SAEON website, which the SAEON Data Portal and uLwazi website directs users to. Institutional data agreements and data management plans are the first steps for data providers unless they are submitting once off datasets. This also enables for data infrastructure provisioning and resource planning to accommodate very large data submissions.

✓ The steps for making data available in the SAEON ODP are currently executed by members of the data curation team because user registration has not been fully implemented. In future, the system will allow data providers to automatically execute the ingest steps prior to curation. In the current iteration of the ODP the curation and publication of the data follows this workflow:

✓ On verification of a SAEON Data Policy compliant data submission, a Submission Information Package consisting of data and metadata submitted by a data provider is created by the data curation team and uploaded to the SAEON ODP file repository. …

Source: South African Environmental Observation Network (SAEON), "2026-09-04 - SAEON Open Data Platform - CoreTrustSeal Requirements 2020-2022", https://doi.org/10.34894/CZ1JFL, DataverseNL, V2

Workflows Example: SunScholarData CoreTrustSeal Application

SunScholarData is the institutional research data repository of Stellenbosch University in South Africa.

The guideline has been fully implemented in the repository – 4

Archiving takes place in terms of a data curation workflow that was derived directly from the SUNScholarData Regulation – more specifically the provisions of sections 5 (submissions), 6 (selection and appraisal), 7 (processing and treatment), 8 (ingestion and storage) and 9 (metadata assignment) of the Regulation. The detailed aspects relating to the submissions, selection, appraisal, processing, treatment and metadata assignment have already been addressed under R8 and R11.

The data curation workflow has been reduced to a diagram. In addition to this, the granular steps of the workflow have been reduced to two documents – a summary of the steps and detailed step-by-step guide. Furthermore, the details relating to the metadata management aspects of the workflow have also been reduced to two documents – a summary of the steps and a detailed step-by-step guide.

Source: SUNScholarData, 2024, - CoreTrustSeal Requirements 2020-2022", https://doi.org/10.34894/D0FFK8, DataverseNL, V2

Workflows Example: DataFirst CoreTrustSeal Application (Excerpt 1 of 2)

DataFirst, the University of Capetown domain data repository, shares research data on African countries.

The guideline has been fully implemented in the repository - 4

Data curation workflows at DataFirst are clearly documented in our Digital Curation Document … and our Digital Curation Reference Model based on the OAIS framework. Both are published on our website … .

✓ Our curation processes cover the entire life cycle of datasets, from ingest of the SIP to dissemination of the DIP and data user support. Digital curation processes and procedures are pre-defined for different dataset types and access levels. Documenting business processes supports change management of these processes. Handling of data at each stage of our workflow is governed by the principles we adhere to and our digital curation policies. Tasks at each stage are also informed by international data curation standards, as depicted in Figure 4 in our Digital Curation Document …

Digital Curation Workflows as Depicted in our Digital Curation Reference Model

✓ As depicted in our model, Digital Preservation activities at DataFirst begin with Stage 2 which involves accepting the deposit dataset (SIP - which will include data collection documents) (Ingest) The Deposit Dataset (SIP) is depicted as A in our model. The workflow includes liaison with Depositors to agree on the contents of a Memorandum of Agreement with DataFirst. Information for Depositors and our MOA are available on our website … .

✓ Stage 2 workflows are underpinned by open data principles … and FAIR data principles. They are also governed by our Collections Policy which specifies the selection criteria for the data we accept. We collect data on African countries or multi-country data which has an African component. We do not accept data we cannot share in some manner, or data where ownership or permissions to share are not clear. Our Collections Policy is covered in Section 2.8 of our curation document … .

Source: DataFirst, 2024, "2026-12-06 - DataFirst - CoreTrustSeal Requirements 2020-2022", https://doi.org/10.34894/9EWIVW, DataverseNL, V2

Workflows Example: DataFirst CoreTrustSeal Application (Excerpt 2 of 2)

DataFirst, the University of Capetown domain data repository, shares research data on African countries.

Stage 3 involves digital preservation activities in which we make changes to the deposit dataset to produce a copy suitable for dissemination (DIP). These activities contribute value to datasets, by protecting the integrity of their content and by increasing their usability. At this stage we run data accuracy and consistency checks on the digital objects that make up the datasets. We also undertake disclosure control to ensure we do not disseminate personally identifiable data that could compromise respondents' privacy. Our disclosure control workflows are documented and diagrammatically represented online … and as Figure 7 in our digital curation document …

Stages 4 and 5 are where we create metadata for the digital objects that make up the dissemination dataset (DIP) Our descriptive metadata includes metadata elements that allow for identification of datasets, for example, title, author, and abstract. In addition, administrative metadata includes management and use information, such as version history and access permissions. …

Stage 6 shows the storage of the original deposit which remains archived unchanged as the deposit copy (SIP). At this stage it is stored alongside a copy of the dataset prepared for dissemination (DIP). Together these form archival copies (AIP), shown as B in our model. The AIP may be several copies as changes to digital objects produce new versions. All previous iterations of the dataset are archived as a change record and for reference purposes. For example, outdated versions may need to be consulted for tombstone pages for citations to an earlier version of the dataset. The AIP is therefore not a static record but is subject to alterations so that digital objects are accessible and optimally usable. …

Source: DataFirst, 2024, "2026-12-06 - DataFirst - CoreTrustSeal Requirements 2020-2022", https://doi.org/10.34894/9EWIVW, DataverseNL, V2

R11 Workflows Example: WDCC CoreTrustSeal Application

The German Climate Computing Center (DKRZ), Hamburg, Germany, operates the World Data Center for Climate (WDCC).

Implemented: the requirement has been fully implemented by the repository – 1

- ✓ The WDCC is a thematically oriented archive. As the focus is on output from numerical Climate and Earth System Models, the range of data and metadata is restricted (see also R04). This narrow focus allows for the use of a single standardized preservation workflow.
- ✓ The WDCC adheres to the OAIS reference model throughout the whole workflow, from the submission to the publication of the data. This workflow is described in the WDCC Preservation Plan and Workflow (see Figure 1).
- ✓ When the data producer requests archiving of his/her data in the WDCC, the standardized workflow is initiated (see also R10). The practical implications of the workflow for data providers are detailed in the WDCC User Guide while the more general consequences are described in the WDCC Preservation and Storage Policy. The contact between data provider and WDCC curator, as well as the internal communication within the curator team are moderated and documented in the WDCC ticket system (Request Tracker).
- ✓ All WDCC workflows and documentation are updated on a regular basis to incorporate recent changes in data and metadata best practices.

Source: World Data Center for Climate (WDCC) 2024 - CoreTrustSeal Requirements 2023-2025", https://doi.org/10.34894/LWLCTT, DataverseNL, V1

R11 Workflows Homework: Self-Assessment in Template

- Describe the workflows/business processes covering the curation levels performed.
- How are workflows adjusted for different types of data and metadata?
- What decision handling occurs within the workflows?
- How is the change management of workflows facilitated?
- How is workflow execution tracked, with mechanisms to handle exceptions?

Questions?

Introduction to CoreTrustSeal Requirements R12 and R13

R12 Discovery and Identification

R12. The repository enables users to discover the digital objects and refer to them in a persistent way through proper citation.

Self-Assessed Compliance Level:

Effective data and metadata sharing discovery is key to resource discovery. Once discovered, digital objects should be referenceable through full citations, including persistent identifiers (PIDs) to help ensure that they can be accessed into the future.

R12 Discovery and Identification

The response statement and evidence should include references to the following items:

- The search facilities offered by the repository.
- The standards that a searchable metadata catalogue complies with.
- The approach to ensuring that identifiers are unique and persistent.
- Machine harvesting of the metadata.
- Repository, or repository data and metadata, inclusion in disciplinary or generic registries of resources.
- Recommended data citations.

R12 Discovery and Identification

Applicants should describe their use of a third party persistent identifier system, or document their own approach to ensuring that identifiers remain globally unique and persistent. The use of a third party to support PID creation and resolution is not sufficient; applicants should describe how they ensure that identifiers continue to resolve to the correct data or metadata over time, including the version rules that guide when a new identifier is created for a digital object.

Applicants that do not have a persistent identifier solution cannot achieve "Implemented: the requirement has been fully implemented by the repository" for this requirement.

R13 Reuse

R13. The repository enables reuse of the digital objects over time, ensuring that appropriate information is available to support understanding and use.

Self-Assessed Compliance Level

Repositories must ensure that data and metadata continue to be understood and used effectively into the future despite changes in technology and the Designated Community's knowledge base. This Requirement evaluates the measures taken to ensure that data and metadata are reusable.

R13 Reuse

The response statement and evidence should include references to the following items:

- The ways in which the repository engages with their Designated Community of users to identify their needs.
- The data formats, metadata schemas, controlled vocabularies and ontologies used to support reuse, and how these meet the community needs.
- The metadata and documentation provided at the point of access to support understandability and reuse appropriate to the Designated Community. This may include information specific to data type, e.g. manuals, calibration records, photos, protocols.
- Measures to ensure that data and metadata remain understandable.
- Management of changes to data, metadata, documentation or other information that supports reuse.

R13 Reuse

Responses to this Requirement should focus on engagement with the Designated Community, identification of their needs and specifying how their needs are met.

Preparing for Next Session

Preparing for Next Session: Activities in Your Repository

- Update template assigned to each repository for self-assessment
- Review requirements R09, R10 and R11 for your repository
- Discuss among managers and staff within your repository
- Draft responses to R09, R10 and R11 for your repository in your template
- Review draft responses to other requirements for consistency

Review of the CoreTrustSeal Requirements

CoreTrustSeal Requirements 2023-2025: Context

R0 Context

- General information about the repository
- Re3data identifier
- Repository Type. Select all relevant types from:
  - Generalist or Specialist; Specialist describes scope
- Overview – Brief Description of Repository
- Brief Description of the Designated Community
- Level of Curation Performed. Select all relevant types from:
  - Content distributed as deposited Basic curation – e.g., brief checking, addition of basic metadata or documentation. Enhanced curation – e.g., conversion to new formats during ingest, enhancement of documentation or metadata; Data -level curation – as in C above, but with additional editing of deposited data for accuracy
- Cooperation and Outsourcing to third parties, partners, and host organizations.
- Applicants renewing – Summary of Significant Changes Since Last Application

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Requirements 2023-2025 (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051012

CoreTrustSeal Requirements 2023-2025: Organizational Infrastructure

- R01 Mission & Scope
  - The repository has an explicit mission to provide access to and preserve digital objects.
- R02 Rights Management
  - The repository maintains all applicable rights and monitors compliance.
- R03 Continuity of Service
  - The Repository has a plan to ensure ongoing access to and preservation of its data and metadata.
- R04 Legal and Ethical
  - The repository ensures to the extent possible that data and metadata are created, curated, preserved, accessed and used in compliance with legal and ethical norms.
- R05 Governance & Resources
  - The repository has adequate funding and sufficient numbers of staff managed through a clear system of governance to effectively carry out the mission.
- R06 Expertise & Guidance
  - The repository adopts mechanisms to secure ongoing expertise, guidance and feedback-either in-house, or external.

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Requirements 2023-2025 (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051012

CoreTrustSeal Requirements 2023-2025: Digital Object Management

- R07 Provenance & Authenticity
  - The repository guarantees the authenticity of the digital objects and provides provenance information.
- R08 Deposit & Appraisal
  - The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for users.
- R09 Preservation Plan
  - The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.
- R10 Quality of Service
  - The repository addresses technical quality and standards compliance, and ensures that sufficient information is available for end users to make quality-related evaluations.
- R11 Workflows
  - Digital object management takes place according to defined workflows from deposit to access.
- R12 Discovery & Identification
  - The repository enables users to discover the digital objects and refer to them in a persistent way through proper citation.
- R13 Reuse
  - The repository enables reuse of the digital objects over time, ensuring that appropriate information is available to support understanding and use.

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Requirements 2023-2025 (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051012

CoreTrustSeal Requirements 2023-2025: Information Technology & Security

- R14 Storage & Integrity
  - The repository applies documented processes to ensure data and metadata storage and integrity.
- R15 Technical Infrastructure
  - The repository is managed on well-supported operating systems and other core infrastructural software and hardware appropriate to the services it provides to its Designated Community.
- R16 Security
  - The repository protects the facility and its data, metadata, products, services, and users.

Source: CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Requirements 2023-2025 (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051012

Online Guidance and Related Resources for Improving Data Stewardship and Attaining CoreTrustSeal Certification 2023-2025

- CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Requirements 2023-2025 (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051012
- CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Trustworthy Digital Repositories Requirements 2023-2025 Extended Guidance (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051096
- CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Trustworthy Data Repositories Requirements: Glossary 2023-2025 (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051125
- CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Revision Working Group Change Log and Associated Materials (v01.00). Zenodo. https://doi.org/10.5281/zenodo.7051237
- Apply for CoreTrustSeal Certification: https://www.coretrustseal.org/apply/
- CoreTrustSeal Requirements (& webinars): https://www.coretrustseal.org/why-certification/requirements/
- CoreTrustSeal Administrative Fee (& waiver): https://www.coretrustseal.org/apply/administrative-fee/
- CoreTrustSeal: https://www.coretrustseal.org/

Online Resources for Improving Data Stewardship and Attaining CoreTrustSeal Certification (Continued)

- Consultative Committee for Space Data Systems. 2012. Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-M-2. Also published as ISO 14721:2012. https://public.ccsds.org/Pubs/650x0m2.pdf
- Consultative Committee for Space Data Systems (CCSDS). 2011. Audit and Certification of Trustworthy Digital Repositories. CCSDS 652.0-M-1. Also published as ISO 16363:2012. https://public.ccsds.org/pubs/652x0m1.pdf
- Group on Earth Observations (GEO). 2016. Global Earth Observations System of Systems (GEOSS) Data Management Principles. http://earthobservations.org/open_eo_data.php#
- Group on Earth Observations (GEO). 2016. Global Earth Observations System of Systems (GEOSS) Data Sharing Principles. http://earthobservations.org/open_eo_data.php#

Online Resources for Improving Data Stewardship and Attaining CoreTrustSeal Certification (Continued)

- Carroll, S.R., et al. 2020. The CARE Principles for Indigenous Data Governance. Data Science Journal, 19(1). https://doi.org/10.5334/dsj-2020-043
- Lin, D., et al. 2020. The TRUST Principles for digital repositories. Scientific Data 7, 144. https://doi.org/10.1038/s41597-020-0486-7
- Wilkinson, M, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Online Resources for Improving Data Stewardship and Attaining CoreTrustSeal Certification (Continued)

- International Science Council (ISC) CODATA international discussion list
  - http://lists.codata.org/mailman/listinfo/codata-international_lists.codata.org
- Research Data Alliance (RDA) https://www.rd-alliance.org/
  - https://www.rd-alliance.org/groups/rdawds-certification-digital-repositories-ig.html
- Group on Earth Observations (GEO)
  - https://earthobservations.org/
- World Data System (WDS)
  - https://www.worlddatasystem.org/

Online Resources for Improving Data Stewardship and Attaining CoreTrustSeal Certification (Continued)

- Downs, 2021. Improving Opportunities for New Value of Open Data: Assessing and Certifying Research Data Repositories. Data Science Journal. https://doi.org/10.5334/dsj-2021-001
- Downs RR. 2019. International Standards for Trustworthy Data Repositories. National Institutes of Health (NIH) Trustworthy Data Repositories Workshop. Rockville, MD. https://doi.org/10.7916/d8-h7xf-ha70
- Downs RR. 2021. Adopting the TRUST Principles for Digital Repositories with the GEOSS Data Management Principles and the GEOSS Data Sharing Principles. ESIP 2021 Winter Meeting (Virtual). https://doi.org/10.6084/m9.figshare.13607747.v1
- Downs RR, Chen RS. 2016. A Portfolio Approach to a Sustainable Business Model for Scientific Data Stewardship. SciDataCon 2016, Denver, CO. https://doi.org/10.7916/d8-fae5-cz67
- Downs RR, Chen RS. 2015. Bridging Disciplines: Assessing the Interdisciplinary Impact of Open Data. 41st IASSIST Annual Conference, Minneapolis, MN. https://doi.org/10.7916/D8J38SDZ
- Downs RR, Lehnert KA. 2020, Opportunities for Sharing Resources Among Research Data Repositories. Research Data Alliance 16th Plenary (RDA P16). https://doi.org/10.7916/d8-kv81-sv37
- Downs RR, Lehnert KA, Zhang L, Donaldson DR, Silva JR. 2020. Selecting Research Data Repository Platforms for Open Science. RDA P16. https://doi.org/10.7916/d8-051c-y379

Thank you!

rdowns@ciesin.columbia.edu