SERVIR West Africa 2025 Training Program on Data Stewardship and the CoreTrustSeal Requirements: Session 1, R07 & R08 Requirements

Robert R. Downs, PhD

Center for International Earth Science Information Network (CIESIN)

Columbia Climate School, Columbia University

Coordinated by AFRIGIST

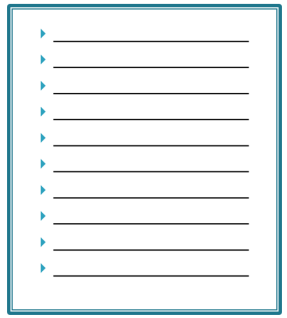
October 30, 2024, 13:00 – 14:30 GMT (9:00 - 10:30 a.m. EDT)

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

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

- Key topic:
  - R07 & R08 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 07 and Examples of Responses (20 minutes)
- Homework: Draft Self-Assessment Responses (5 Minutes)
- CoreTrustSeal Requirement 08 and Examples of Responses (20 minutes)
- Homework: Draft Self-Assessment Responses (5 Minutes)
- Introduction to CoreTrustSeal Requirements R09, R10, R11 (15 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, November 20, 2024, 12:00-13:30 GMT *
  - Topics: R09, R10, and R11 Requirements and Introduction to R12 and R13 Requirements

*** Please note that the start time is one hour earlier than usual.**

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

Questions about previous homework to compose draft responses in template

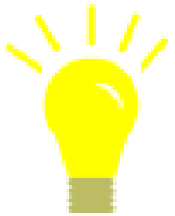
- Identifying resources that address a challenging requirement
  - Can we start listing and describing current resources that address the requirement?

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

Review of Homework in Template: R06 Expertise & Guidance

- Describe the type of training and expertise that supports repository services.
- Describe repository services that inform recruitment and community engagement.
- Describe the training and professional development of staff members.
- Describe the relevant expertise and affiliations of the organization’s staff.
- Describe the internal and external advisors that guide services offered.
- Describe how the repository communicate with experts for advice.



We have completed exploring the CoreTrustSeal requirements for Organizational Infrastructure, R01-R06.

We focus next on the CoreTrustSeal requirements for Digital Object Management, R07-R13 .

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:

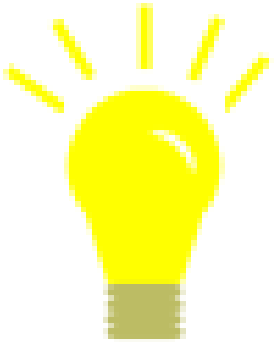
Digital Object Management

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 Requirement R07 Provenance and Authenticity



CoreTrustSeal Requirement R07 Provenance and Authenticity

- R07. The repository guarantees the authenticity of the digital objects and provides provenance information.

Compliance Level

Guidance

- The repository should provide evidence to show that it operates a data and metadata management system that maintains provenance information to ensure authenticity from deposit, and through curation and preservation to the point of access.
- Any intentional changes to data and metadata should be documented, including the rationale and originator of the change. Authenticity covers reliability and provenance, including the relationship between the deposited digital objects and those provided at the point of access.

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 R07 Provenance and Authenticity (cont'd)

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

- The repository approach to changing and versioning data and metadata. How the approach and records of changes are communicated to data depositors and users.
- The provenance information and audit trails recorded for data and metadata processing and versioning.
- How the repository compares the essential properties of different versions of the same file.
- Identification checks for depositors.

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

R07 Provenance and Authenticity Example: SAEON CoreTrustSeal Application (Excerpt 1of 2)

Compliance level: The repository is in the implementation phase - 3

✓ The aim of our measures in this respect are to ensure that data as deposited can always be retrieved, all modifications are recorded in the metadata, and that persistent identifiers (PIDs) always resolve in metadata landing pages that describe the provenance of the dataset being accessed. Some of these measures are currently manual, however we are working towards automating as much of the workflow as possible in the future. The following guiding principles are followed to ensure integrity and authenticity of data managed by SAEON:

✓ - The Submission Information Packages are archived as they were submitted and all changes and additions made during the creation of the Archival Information Package are recorded in the metadata, as are changes made to the Archival Information Package thereafter.

✓ - Data providers agree to the following in the Data Agreement: “The data producer grants SAEON the permission to copy, store and transform the uploaded data in order to publish data in standardised formats for access via services within license constraints, perform format migrations as part of the preservation process.” If the data in a Submission Information Package needs to be converted to a different format that the data producer is unable to provide, then a curator will make the necessary changes when they create the Archival Information Package.

✓ - SAEON metadata records are validated automatically against the relevant schema on creation and on update. Curators always verify that the data in the SAEON Open Data Platform

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

R07 Provenance and Authenticity Example: SAEON CoreTrustSeal Application (Excerpt 2of 2)

Submission Information Package matches what is described in the data provider's metadata record.

- The complete state of a metadata record is logged to an audit table for every change to a record - insert, update or delete. Similarly, metadata record tags - which may contain quality or other information associated with metadata records - are also logged to an audit table on insert, update or delete. Data versions are recorded using the related identifier fields in the DataCite metadata schema.

- Version control of the data is documented through the DataCite metadata schema which uses related identifiers to link versions that are derived from the Archival Information Package, or which provide major or minor versions of the original Archival Information Package. Changes in datasets are considered as major versions whereas additional details about the dataset are considered as minor versions. Related datasets and metadata are also linked through DataCite DOIs and the use of related identifiers.

- SAEON refers to the OAIS model for its workflow configuration and ingest process and makes use of international metadata and data standards including DataCite, ISO 19115, ISO/TS 19139 and CF Conventions for NetCDF data.

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

R07 Provenance and Authenticity Example: DataFirst CoreTrustSeal Application (Excerpt 1 of 4)

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

Compliance level: The guideline has been fully implemented in the repository - 4

✓ Our Digital Curation Reference Model aligns with the OAIS but uses our own terminology and the model has been adapted for our curation environment. The numbering in the model reflects stages of the digital data lifecycle at DataFirst https://www.datafirst.uct.ac.za/services/data-curation-process

✓ Managing data from our Depositor is depicted as Stage 2 in our Digital Curation Model. Stage 2 aligns with activities in the Ingest stage of the OAIS. The deposit dataset (SIP) is depicted as digital object A in our Curation Model. The ingest stage is important for dealing with ethical as well as practical issues around data deposits. Ethical concerns include ownership, consent, and privacy. Practical issues include locating data collection documents and managing secure data transfer, as well as agreeing with Depositors on timeframes for disseminating data. The ideal data transfer method is secure self-deposit by depositors using an application that captures basic metadata. We are working with software developers to upgrade our dissemination DataFirst software to accept online deposits of data and metadata. Currently however, we arrange methods with depositors depending on the level of security required. Depositors complete short Data Description form https://www.datafirst.uct.ac.za/files/20220914-df-data-desc-v6.xlsx to provide basic metadata on the Deposit Dataset and we sign a Memorandum of Agreement with each depositor which codifies data sharing arrangements. Information on depositing data with us and the MOA for depositors can be accessed on our website https://www.datafirst.uct.ac.za/services/deposit-data

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

R07 Provenance and Authenticity Example: DataFirst CoreTrustSeal Application (Excerpt 2 of 4)

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

Dataset Digital Objects

✓ Dataset entities include data files and all materials relevant to the collection and ongoing use of the data. Digital objects that are not data files are referred to as External Resources at DataFirst, in line with standard Nesstar DDI-compliant metadata terminology. They are generally documents, such as data collection instruments (administrative forms or questionnaires) and reference documents such as codebooks, as well as analytic documents and technical reports. However, these digital objects may be syntax files or other programmes, podcasts or videos. Like data files, external resources are prepared, versioned, and stored in accordance with our digital curation policy. The digital objects that are prepared for public access and data enclave access datasets are modelled on our site at https://www.datafirst.uct.ac.za/files/20210802-data-preparation-model.pdf and as Figure 5 in our Digital Curation and Preservation Document https://www.datafirst.uct.ac.za/files/df-digital-curation-preservation.pdf

✓ Data Preparation and Preservation is depicted in Stages 3-5 in our Digital Curation Reference Model. Preparing datasets for preservation and sharing is a component of Ingest in the OAIS model as well as part of Preservation and Planning in the OAIS system. The Deposit Dataset (SIP in the OAIS) is the integral copy and is never changed by DataFirst.

✓ The archival or preservation dataset (Archival Information Package (AIP) in the OAIS) will undergo changes in the data preparation stage which is part of the Ingest stage in the OAIS model. Preservation copies include all iterations of the dataset as we keep all previous copies as a change record and for reference purposes. For example, outdated versions may need to be consulted for tombstone pages https://support.datacite.org/docs/tombstone-pages on a removed record or citations to an earlier version of the data.

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

R07 Provenance and Authenticity Example: DataFirst CoreTrustSeal Application (Excerpt 3 of 4)

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

- ✓ The archival or preservation dataset may be preserved with additional resources such as administrative documents from the data collection stage or created during negotiations with Depositors (MOA, branding information, logos). These are not made publicly available and are therefore not included in the dissemination dataset.
- ✓ The dissemination dataset (Dissemination Information Package (DIP) in the OAIS) will be the latest and most usable version of the dataset and the one shared with researchers.
- ✓ The archival or preservation dataset (AIP) and dissemination dataset (DIP) may undergo changes and these changes are documented in dataset- and file-level metadata. Change metadata provides an audit trail from the SIP to other versions to ensure the authenticity of the data. Data quality checks are undertaken by our Data Analysts but we also crowd-source for quality input from data users who communicate with us through our online support site http://support.data1st.org/ Quality assessments include checks for accuracy and consistency.
- ✓ Consistency checks determine whether the data is coherent within the same dataset e.g., whether variables and values are consistently represented and whether separate data files within a dataset can be merged. We also check for data comparability across a dataset series which may be compromised by changes in data collection methods, for example, a change in the sample frame or survey question between survey rounds. Where possible, DataFirst strives to correct data errors. Error correction is carried out in consultation with Depositors. Where error correction is not possible, data errors and other quality issues are documented in the metadata we publish with each dataset.

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

R07 Provenance and Authenticity Example: DataFirst CoreTrustSeal Application (Excerpt 4 of 4)

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

Version Control/Change Procedures

✓ Some changes may need to be made to digital objects that make up the datasets for the data to be accurate and optimally usable. We use versioning or version control to manage multiple variations of datasets …

✓ Corrections and updates to any of the dataset components will be denoted by a new version number. In the case of a Dataset deposited without version information, we follow a policy of labelling the deposit dataset as version 1. Our versioning policy distinguishes between a new version and a new release. That is, changes we make to deposited data files or documents will be indicated by a new version number denoted by a minor version update and indicated by increments to the decimal place … . Re-releases may occur when errors are discovered in data entities, either by Depositors or by DataFirst. In these cases, the depositor recalls the data and reissues a corrected version.

✓ We version at file level and align the dataset version with the version of the most recently changed file. …

✓ All changes to all digital objects that make up datasets must be documented in the descriptive and structural metadata to ensure their authenticity. Version information is recorded in file labels and file-level metadata as well as in the metadata record for the whole dataset that we post online. This information can be found in the “Version Notes” field in the online metadata. Metadata. Metadata records are also versioned, in compliance with the DDI metadata schema …

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

R07 Provenance and Authenticity Example: SunScholarData CoreTrustSeal Application

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

Compliance level: The guideline has been fully implemented in the repository - 4

✓ MD5 integrity checks are performed when the file is uploaded to SUNScholarData and are displayed alongside file previews on the record. The computed value is stored along with the other metadata for further checks. The default storage platform that Figshare relies on for storage (Amazon AWS S3) performs regular, systematic data integrity checks as well.

✓ MD5 integrity checks are performed on all uploads and are displayed on file previews. DOIs are only issued when all the mandatory metadata elements, set by DataCite, have been filled out. Furthermore, version control of the metadata and the files is supported for all research outputs that are made publicly available.

The audit log covers actions performed through the website, api, or widgets systems. Generally, all actions which result in an update of an item are logged. In case of an emergency, Stellenbosch University can request (from Figshare) a dump of the audit log for actions in the organisation at any time via a support request.

✓ SUNScholarData supports version control of all publicly available data. Any privately stored data can also be altered or deleted at will.

✓ SUNScholarData supports versioning for both items and collections. Because projects are a continuous piece of work with a start date and finish date they are not candidates for versioning. There are a couple of changes that would trigger versioning. These rules are a bit different between items and collections.

✓ These actions can be done from SUNScholarData's website, the API or any submission method used by the user. Versions are listed and accessible in the drop-down menu under the item title, each is timestamped. …

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

R07 Provenance and Authenticity Example: WDCC CoreTrustSeal Application

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

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

✓ Quality assurance of the data and metadata is part of the publication process and documented in the metadata of the data collection (WDCC Quality Assurance). Deliberate changes to the metadata or to the data structure are documented in a separated table in the metadata database. Changes to data objects are not allowed after the publication. New versions of data objects are possible, but they are versioned and treated as independent objects and provided with mutual pointers and with the corresponding information about the changes. Old versions are kept in the archive. This also applies to the corresponding metadata entries. This is clearly communicated to the data producer during the editorial process of publication. WDCC's policy for ensuring the integrity of data objects and metadata is set out in the WDCC Preservation and Storage Policy.

✓ To ensure the integrity of the files, the WDCC verifies the checksums of the files. In most cases, the checksums are already calculated by the data producers. The checksums should be stored by the producer as metadata and can then be used by data users for verification after download. WDCC strongly recommends the use of the common community format NetCDF, which is a selfdescribing file format and allows rich file metadata. The NetCDF file format allows the implementation of the Climate Forecast Conventions (CF), which is a very widely used community metadata standard in climate science. The implementation of the CF conventions results in highly standardised and machine-readable file metadata. The WDCC accepts other non-proprietary file formats, like GRIB, e.g. when very large data volumes require specific compression.

✓ Both, NetCDF and GRIB file format provide the opportunity to store audit trail metadata and provenance information directly in the file header, as strongly recommended by WDCC. Provenance information is additionally included and maintained in the describing metadata. The WDCC always checks the correctness of the affiliation information when users request depositing data in the WDCC.

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

R07 Provenance and Authenticity Homework: Self-Assessment in Template

- How does the data and metadata management system maintain provenance?
- How does the repository ensure authenticity from deposit to the point of access?
- How does the repository document changes to data and metadata?
- Describe the repository's approach for changing and versioning data and metadata.
- How does the repository communicate data and metadata changes to depositors and users?
- Describe how provenance information and audit trails for data and metadata are recorded.
- How does the repository compare essential properties of different versions of the same file?
- How is the identification of depositors verified?

Questions?

CoreTrustSeal Requirement, R08 Deposit & Appraisal

CoreTrustSeal Requirement R08 Deposit & Appraisal

R08. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for users.

Self-Assessed Compliance Level:

The appraisal function during deposit is critical to evaluate whether digital objects meet all criteria for selection and to ensure appropriate management for their preservation. Appraisal ensures that deposited digital objects are relevant and are, or can become, understandable to the Designated Community.

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 R08 Deposit & Appraisal (Continued)

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

- Any documented deposit process that includes steps to ensure that data and metadata are sufficient for long-term preservation.
- A collection development policy or procedures to guide the selection of digital objects .
- Criteria for prioritisation and any different curation-levels or preservation levels defined during appraisal.
- The approach to digital objects that do not fall within the mission/collection profile.
- Procedures to determine that the metadata required to interpret and use the digital objects are provided.
- Any automated assessment of metadata adherence to relevant schemas.
- The repository approach if metadata provided is insufficient for long-term preservation .
- A list of preferred formats.
- Checks in place to ensure that depositors adhere to the preferred formats.
- The approach towards digital objects that are deposited in non-preferred formats.
- The transfer of custody and responsibility during the handover from the depositor to the repository.

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 R08 Deposit & Appraisal (Continued)

This Requirement covers the selection criteria applied at the point of deposit. Data Quality (R11) should be used to address steps taken by the repository during the curation process.

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

R08 Provenance and Authenticity Example: SAEON CoreTrustSeal Application (Excerpt 1of 2)

✓ SAEON archives any publicly funded data that is captured in the domain of earth and environmental observation in South Africa (the collection policy is described in the Preservation Policy[1]. If data is submitted that does not fall within the collection policy it will be sent back to the data provider and the curators will suggest an alternative repository if possible. SAEON also provides decision and policy support to government institutions, particularly with regards to earth and environmental observation, and we also work on obtaining any datasets that can potentially supplement decision support and policy making in these areas.

✓ The SAEON ODP mostly accepts publication ready data and the onus is on the data provider to ensure the quality of the data that they provide. Other than assuring that the data and metadata are standard compliant and in a preferred format, SAEON does not perform full quality assurance on the content of datasets. Quality assurance on SAEON produced datasets falls under the responsibility of the staff member submitting the data to the ODP. If there is feedback from data users that indicates an issue with the data quality the data provider will be informed and it is their responsibility to rectify the issue and resubmit the Submission Information Package.

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

R08 Provenance and Authenticity Example: SAEON CoreTrustSeal Application (Excerpt 2of 2)

- ✓ SAEON makes use of metadata standards accompanying the original submission (such as ISO 19115), which we map to an extended DataCite schema for harmonisation. For both metadata sets (original and mapped) automated schema validation is performed and all mandatory fields need to be provided. At the very least this includes enough metadata for citation and includes the data provider's contact details. If the mandatory metadata fields are not filled in or if the data producer has not provided enough information to make the data reusable, for example, then the curator working on the submission will request additional information from them before going ahead with the submission.
- ✓ SAEON has developed guidance documents on the preferred formats for the main data types that are archived in the repository. These are compliant with best practice and international standards and are reviewed as the standards are updated.
- ✓ There is a list of preferred formats published on the data submission guidance page[2] on the Wiki. The curator creating the Archival Information Package will check the format of the data uploaded, as well as confirming that it corresponds with what is described in the metadata. If the data are not in a format that is accepted by the relevant community then the data producer will be informed and either they, or the curator, will migrate it to the preferred format. If a format becomes outdated then the curators will migrate the data to the new preferred format, but this process is currently manual.
- ✓ We discourage the removal of published datasets, but if a data provider needs to remove a dataset then the metadata record will be updated to indicate that that has happened.

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

R08 Provenance and Authenticity Example: DataFirst CoreTrustSeal Application (Excerpt)

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

Documentation

✓ We encourage depositors to provide all documents used in data collection and reporting on the data. We use this, often in consultation with depositors, to create useful metadata. Our repository has adopted the Data Documentation Initiative (DDI) metadata schema for describing socioeconomic datasets. Records created according to the DDI schema are expressed in XML (eXtensible Markup Language) which allows the markup of the content of metadata records so that they are standardised and thus machine-readable and interoperable. We use the Nesstar Publisher free data markup software for the creation of our xml-compliant metadata http://www.ihsn.org/software/ddi-metadata-editor

✓ In Publisher we can create dataset-level metadata using templates customised for our Dataset Collection and harvest variable-level metadata from data files for inclusion in the final metadata record. Variable-level metadata includes variable and value names and labels and summary statistics.

✓ Using the Nesstar Publisher application we also create metadata records for documents and other resources to be shared with data files. In Nesstar these elements are referred to as External Resources. Publisher uses the simpler Dublin Core metadata schema for creating metadata records for External Resources which are exported in the Resource Description Format (rdf) standard for web data. The dataset and variable-level metadata record (in xml) and the external resource-level metadata record (in rdf) are web-compliant which enables them to be posted online with the data and document files. The Nesstar Publisher markup software is preferred by us as it is freeware and so can be shared with Data Managers in under-resourced institutions. It is bundled with the freeware data dissemination software application we use for disseminating data from DataFirst our open data site.

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

R08 Provenance and Authenticity Example: Arca Dados CoreTrustSeal Application (Excerpt 1 of 2)

Arca Dados is the official digital research data repository of Fundação Oswaldo Cruz (Fiocruz) in Rio de Janeiro, Brazil.

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

✓ The Arca Dados repository accepts data and metadata based on the process and deposit criteria described below.

The deposit is carried out using a procedure established in the operational plan [6] and with a workflow specified in BPMN [33] (Business Process Model and Notation). In summary, in this process the researcher requests the deposit by fulfilling the Deposit Form [65], then the Arca Dados staff evaluates the deposit demand, creates the dataset, and instructs the researcher to submit the files directly into

✓ the Dataverse platform (following the criteria detailed below). After that, Arca Dados staff instructs the producer to document and describe the dataset (metadata and data) according to data and metadata models for Dataset from the Dataverse platform [5] (see item 3.1 Metadata references). In this way, it allows metadata to be mapped and exported to standard metadata schemes, such as DDI, DataCite and Dublin Core. Finally, Arca Dados team checks and publishes the dataset.

Source: Arca Dados, 2024, "2027-09-12 - Arca Dados - CoreTrustSeal Requirements 2023-2025", https://doi.org/10.34894/BSELBC, DataverseNL, V1

R08 Provenance and Authenticity Example: Arca Dados CoreTrustSeal Application (Excerpt 2 of 2)

The criteria for deposits and evaluation of deposited data and metadata are established according to the following documents: Arca Dados Operational Plan - Content Guidelines [6] (see item 6): establishes criteria for the datasets and related documents allowed for Arca Dados (institutional origin of data producers, secrecy, intellectual property and potential to generate products and services, anonymization of data, institutional interest, long-term interest, completeness, assignment of rights, co-authors of institutions); Arca Dados Operational Plan - Content Organization [6] (see item 5): establishes organisation of repository datasets into communities (unit, project, thematic collections, publication data, institutions) and approval rules and responsibilities for communities (called Dataverses); Arca Dados Operational Plan - File Formats [6] (see item 6.1): Presents data types and exemplifies preferable formats, pointing to Fiocruz Digital Preservation Programme [7] as for file formats established there; Fiocruz Digital Preservation Programme - File Formats [7] (see item 5.2): Recommends open and non-proprietary formats, recommends formats for text, audio, video and database files, for digitised data, lists formats for digitised objects established in the Fiocruz Digitization Manual [64]; Arca Dados Operational Plan - File Formats [6] (see item 6.2): establishes the file size limit (120MB per file); Arca Dados Operational Plan - Metadata [6]: Establishes Dataverse as infrastructure for the repository, with the production of metadata compatible with the DataCite, Dublin Core and DDI schemes; …

Source: Arca Dados, 2024, "2027-09-12 - Arca Dados - CoreTrustSeal Requirements 2023-2025", https://doi.org/10.34894/BSELBC, DataverseNL, V1

R08 Deposit & Appraisal Homework: Self-Assessment in Template

- How are the steps of the deposit process documented to ensure that data and metadata are sufficient for long-term preservation?
- How does the repository address metadata that are insufficient for long-term preservation?
- Describe the policy or procedures that guide the selection of digital objects .
- Describe the appraisal criteria for different levels of curation and preservation.
- How are digital objects handled if they are not within the scope of the repository?
- How does the repository determine if metadata can be interpreted to enable use of digital objects?
- How does the repository manually or automatically assess metadata adherence to relevant schemas?
- Describe the repository's preferred formats and the process for verifying compliance by depositors.
- How does the repository treat digital objects that are deposited in non-preferred formats?
- Describe the transfer of custody and responsibility during the handover from the depositor.

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

Preparing for Next Session

Introduction to CoreTrustSeal Requirements R09, R10, and R11

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

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

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

Preparing for Next Session: Activities in Your Repository

- Update template assigned to each repository for self-assessment
- Review requirements for R07 & R08 within the context of your repository
- Discuss among managers and staff within your repository
- Draft responses to R07 & R08 for your repository within 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