

SERVIR West Africa Training Program on Data Stewardship and the CoreTrustSeal Requirements: Session 2

Instructor: Robert R. Downs, PhD

Center for International Earth Science Information Network (CIESIN)

The Earth Institute, Columbia University

Coordinated by AFRIGIST

May 12, 2021, 10:00 a.m. EDT - 11:30 a.m. EDT (14:00 - 15:30 UTC)



Opening and Introduction to SERVIR West Africa Training Program on Data Stewardship and CoreTrustSeal Requirements

Schedule of Training Program Sessions

- Dates: May 5 (Session 1), May 12 (Session 2), May 26 (Session 3)
- Time: 10:00 a.m. to 11:30 a.m. EDT (14:00 - 15:30 UTC)
- Zoom Link:
<https://columbiauniversity.zoom.us/j/94522663326?pwd=M3FxL2ZOU TE2K11UcjA4ZHdhdGNqdz09>
- Meeting ID: 945 2266 3326 Passcode: 231758

Instructor

- Dr. Robert R. Downs
- Senior Digital Archivist at CIESIN
 - Center for International Earth Science Information Network, Earth Institute, Columbia University
- <https://www.earth.columbia.edu/users/profile/robert-r-downs>

SERVIR West Africa Training Program on Data Stewardship and CoreTrustSeal Requirements

- Key topics:
 - The approach to institutional self-assessment of data stewardship
 - Data product and service management
 - Planning for CoreTrustSeal certification
- Instructional format
 - Lecture, questions, and discussion
 - Homework assignments
- Sessions will be recorded
 - Recordings will be posted on the SERVIR West Africa website



Outline for Current Session

- Review and Discussion of Principles and Concepts from Session 1
- Review of Selected Examples of Completed Assignments for R1, R2, R5, and R6
- CoreTrustSeal Requirements R7, R8, R9 and R10
- CoreTrustSeal Requirements R11, R12, R13 and R14
- Assignment: Describe How Data Repository Meets CoreTrustSeal Requirements R8, R11, R12 and R13



Review and Discussion of Session 1

- Data Stewardship Concepts, Principles, and Certification Instruments
- CoreTrustSeal Certification requirements
- Organizational aspects of Data Stewardship
- CoreTrustSeal Requirements: R0, R1, R2, R3, R4, R5, R6



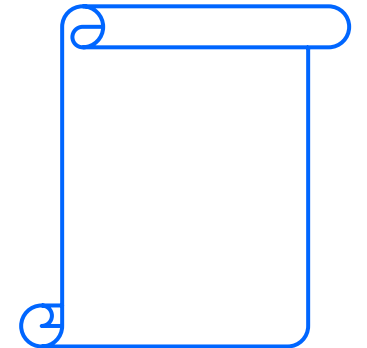
Review and Discussion of Session 1: Data Value and Open Data

- Data value is derived from use
 - Initial use by data collection team
 - Reuse by broader communities
 - Increasing use can increase data value
- Open data enables use
 - Assigning intellectual property rights to enable use
 - Use by anyone for any purpose without restrictions
 - Clearly communicate rights in to potential data users and users



Review and Discussion of Session 1: Data Stewardship, Concepts, Principles, and Certification Instruments

- OAIS (ISO 14721:2012)
- Audit and Certification of Trustworthy Digital Repositories (ISO 16363:2012)
- TRUST Principles for Digital Repositories
- FAIR Principles
- CARE Principles
- GEOSS Data Sharing Principles
- GEOSS Data Management Principles
- CoreTrustSeal Requirements



Open Archival Information System (OAIS) Functional Requirements

- **Negotiates for and Accepts Information**
 - Negotiate for and accept appropriate information from information Producers.
- **Obtains Sufficient Control**
 - Obtain sufficient control of the information provided to the level needed to ensure Long Term Preservation.
- **Determines Designated Community**
 - Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided, thereby defining its Knowledge Base.
- **Ensures Information is Independently Understandable**
 - Ensure that the information to be preserved is Independently Understandable to the Designated Community. ... Designated Community should be able to understand the information without needing special resources
- **Follows Established Preservation Policies and Procedures**
 - Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, including the demise of the Archive, ensuring that it is never deleted unless allowed ...
- **Makes the Information Available**
 - Make the preserved information available to the Designated Community and enable the information to be disseminated as copies of, or as traceable to, the original submitted Data Objects with evidence supporting its Authenticity.

T R U S T



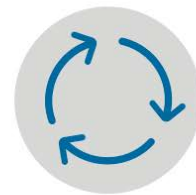
Transparency



Responsibility



User focus



Sustainability



Technology

Findable **A**ccessible **I**nteroperable **R**eusable

Collective Benefit **A**uthority to Control **R**esponsibility **E**thics

GEOSS Data Sharing Principles



GEOSS Data Sharing Principle 1.

Data, metadata and products will be shared as Open Data by default, by making them available as part of the GEOSS Data Collection of Open Resources for Everyone (Data-CORE) without charge or restrictions on reuse, subject to the conditions of registration and attribution when the data are reused;

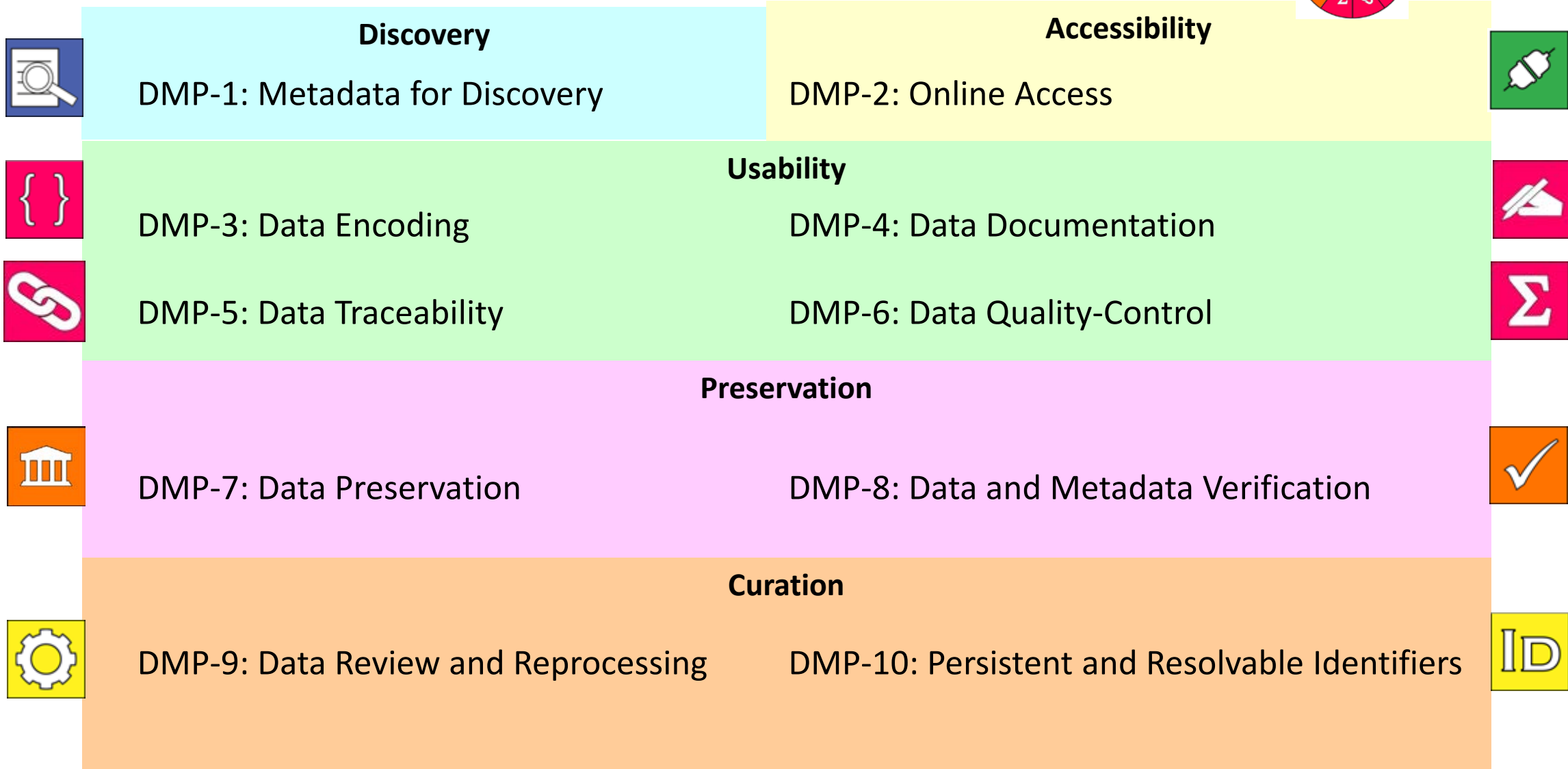
GEOSS Data Sharing Principle 2.

Where international instruments, national policies or legislation preclude the sharing of data as Open Data, data should be made available with minimal restrictions on use and at no more than the cost of reproduction and distribution;

GEOSS Data Sharing Principle 3.

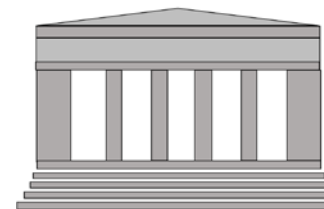
All shared data, products and metadata will be made available with minimum time delay.

GEO Data Management Principles

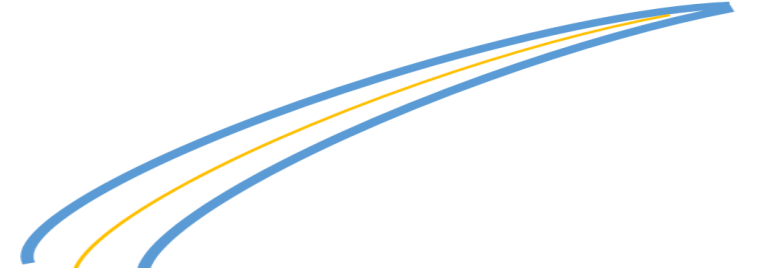


Types of Repositories Seeking Certification

- Domain Repositories
 - Disciplinary
 - Theme-based
 - Content includes data and relevant research-related information
 - Data and metadata formats reflect community practices
- Generalist Repositories
 - Cross-domain
 - Institutional (may be limited to affiliates, for deposits or access)
 - Content may include various types of research output
 - Data and metadata formats may reflect depositors' choices

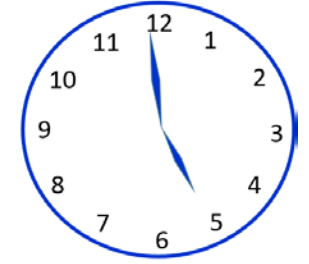


Sustainability for the Future



- Sustainable stewardship of scientific data
 - Long-term access to data for current, returning, and future users
- Sustainability of the data repository
 - How will the repository continue to operate?
- Portfolio approach to a sustainable business model
 - Diversifying sources of support
- Case study on sustainability of a scientific data center
 - Assessing and planning sustainability options

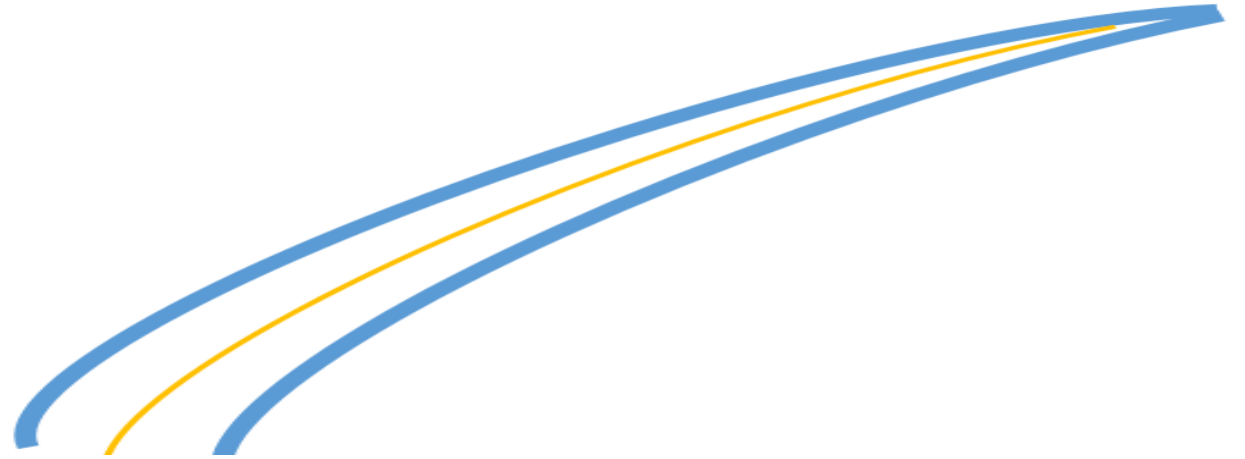
Sustainable Stewardship of Scientific Data



- Scientific data are valuable intellectual assets
 - Resources are needed to collect scientific data
 - Scientific data of recorded observations may be irreplaceable
 - Value of some scientific data may increase over time, especially in the context of long-term environmental and social change
- Stewardship enables reuse of scientific data
 - Data that have contributed to science can be reused if curated
 - Curate scientific data that have been determined to be valuable
 - Preserve and disseminate valuable scientific data that have been curated
- Scientific data stewardship must be sustainable
 - Valuable scientific data should be available for future use
 - Resources are needed to manage and disseminate scientific data
 - Sustain entities responsible for enabling future use of scientific data (preserving the bits is not enough)

Source: Downs & Chen, 2016. A Portfolio Approach to a Sustainable Business Model for Scientific Data Stewardship. SciDataCon 2016.

Sustainability of the Data Repository



- Sustainability
 - Enabling the repository to continue providing data for current and future users
- Achieving Sustainability
 - Continuing challenge for many data repositories
 - Project funding is usually limited to a specific duration
 - Repository improvement projects are more attractive than operations projects
- How can data repositories plan for organizational sustainability?

Developing A Portfolio Approach to a Sustainable Business Model for Scientific Data Stewardship

	Discrete Revenue Stream Models	Cooperative Models
Current Business Model	<ul style="list-style-type: none"> • Fees • Subscriptions • Grants and contracts • Advertising or sponsorship • Donations / endowments • Subsidies 	<ul style="list-style-type: none"> • Institutional commitments • Network development • Commitments from stakeholder communities • Incentives from funders
Long-Term Business Model	<ul style="list-style-type: none"> • Fees • Subscriptions • Grants and contracts • Advertising or sponsorship • Donations / endowments • Subsidies 	<ul style="list-style-type: none"> • Institutional commitments • Network development • Commitments from stakeholder communities • Incentives from funders

Source: Downs & Chen, 2016. A Portfolio Approach to a Sustainable Business Model for Scientific Data Stewardship. SciDataCon 2016.



SOCIOECONOMIC DATA AND APPLICATIONS CENTER (SEDAC)

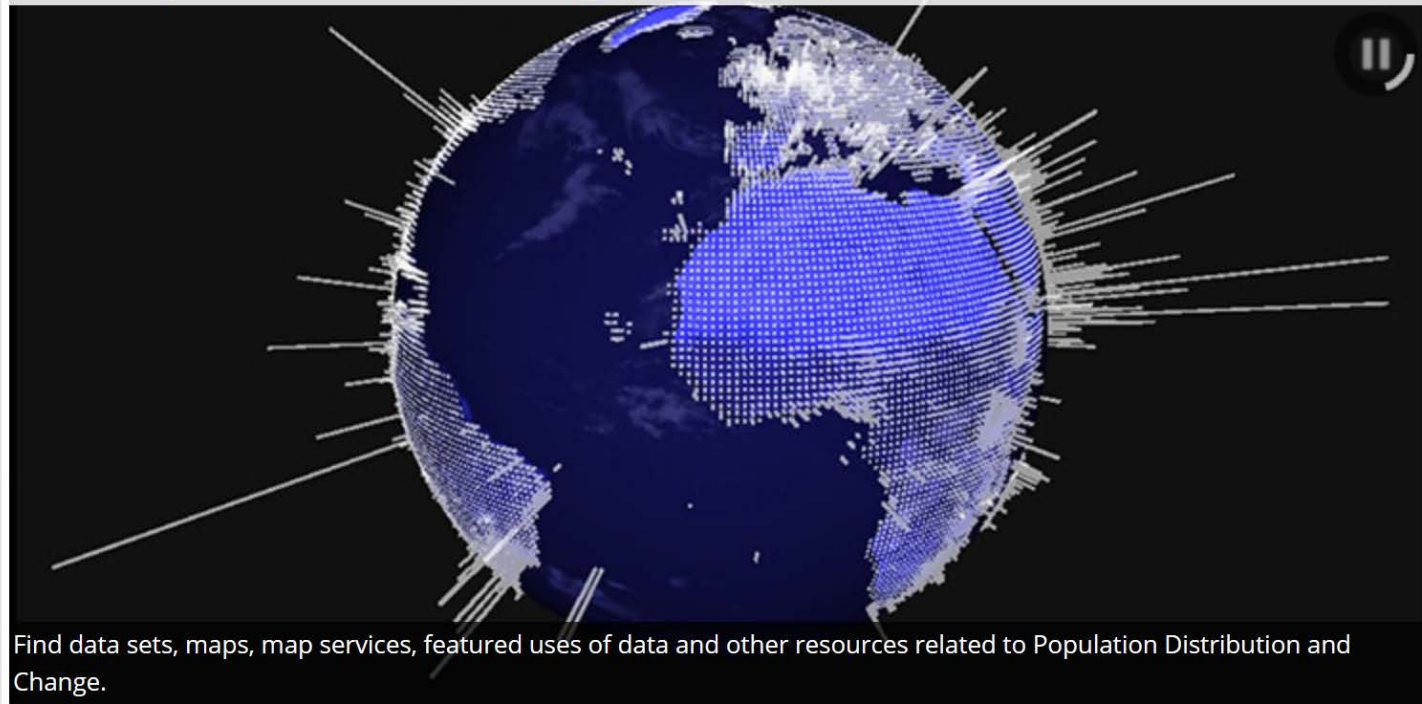
A Data Center in NASA's Earth Observing System Data and Information System (EOSDIS) — Hosted by [CIRES](#) at [Columbia University](#)



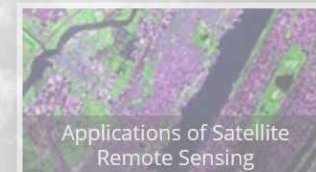
In the Spotlight

Data Citations | Follow Us: | Share:

Theme - Population Distribution and Change



Find data sets, maps, map services, featured uses of data and other resources related to Population Distribution and Change.



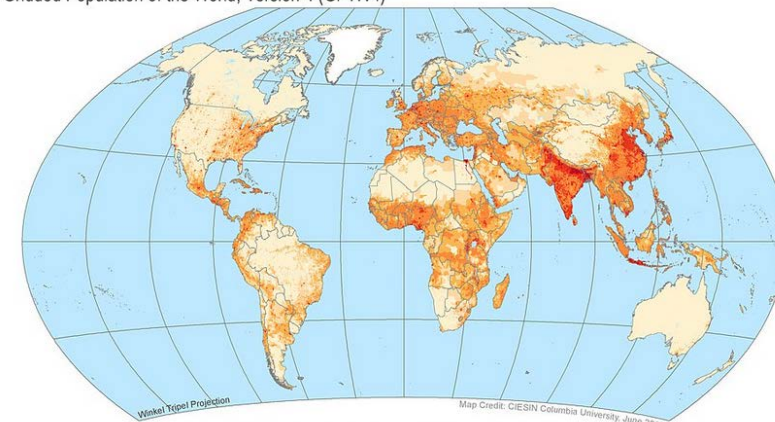
Case Study on Sustainability of a Scientific Data Center

A Case Study of a Sustainable Business Model Portfolio for a Scientific Data Center

- NASA Socioeconomic Data and Applications Center (SEDAC)
 - One of the Distributed Active Archive Centers (DAACs) in NASA's Earth Observing System Data and Information System (EOSDIS)
- Focus on human interactions in the environment
- Developing, freely disseminating, and supporting ~ 200 data sets in ~ 40 collections with more than 2000 map products and a variety of data services
- SEDAC data regularly cited in peer-reviewed journals across multiple disciplines
- SEDAC maps frequently reproduced in textbooks, newspapers, magazines, television, and online media

Population Density Grid, 2015: Global

Gridded Population of the World, Version 4 (GPWv4)



Gridded Population of the World, Version 4 (GPWv4) Population Density consists of estimates of human population density based on counts consistent with national censuses and population registers, for the years 2000, 2005, 2010, 2015, and 2020. A proportional allocation gridding algorithm, utilizing approximately 12.5 million national and sub-national administrative units, is used to assign population values to 30 arc-second (~1 km) grid cells. The population density grids are derived by dividing the population count grids by the land area grids. The pixel values represent persons per square kilometer.

Center for International Earth Data Source: Center for International Earth Science Information Network - CIESIN - Columbia University, 2016. Gridded Population of the World, Version 4 (GPWv4): Science Information Network Population Density. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <http://dx.doi.org/10.7927/H4NP22DQ>.
© 2016. The Trustees of Columbia University in the City of New York.

Map Source: <http://sedac.ciesin.columbia.edu/>

Source: Downs & Chen, 2016. A Portfolio Approach to a Sustainable Business Model for Scientific Data Stewardship. SciDataCon 2016.

SWOT Analysis of Current SEDAC Business Model

<p style="text-align: center;">Strengths</p> <ul style="list-style-type: none">• NASA committed to scientific data stewardship and dissemination• EOSDIS community support for continuous improvement• High levels of use, demonstrated through multiple metrics	<p style="text-align: center;">Weaknesses</p> <ul style="list-style-type: none">• NASA contract is single primary source of support for data center• Growth is limited if other funders assume no need for other support• Those who use data are not those who control funding
<p style="text-align: center;">Opportunities</p> <ul style="list-style-type: none">• New data management technologies and approaches• Partnerships with external communities on new data, services• Potential funding from non-traditional sources	<p style="text-align: center;">Threats</p> <ul style="list-style-type: none">• Government budget cuts affecting NASA and EOSDIS• Government shutdown for a sustained period of time• Reduced interest in SEDAC data and services

Source: Downs & Chen, 2016. A Portfolio Approach to a Sustainable Business Model for Scientific Data Stewardship. SciDataCon 2016.

A Portfolio Approach to a Sustainable Business Model for Scientific Data Stewardship

	Discrete Revenue Stream Models	Cooperative Models
Current Active Archive Business Model	<ul style="list-style-type: none"> • Fees • Subscriptions • Grants and contracts • Advertising or sponsorship • Donations / endowments • Subsidies 	<ul style="list-style-type: none"> • Institutional commitments • Network development • Commitments from stakeholder communities • Incentives from funders
Contingency Long-Term Archive Business Model	<ul style="list-style-type: none"> • Fees • Subscriptions • Grants and contracts • Advertising or sponsorship • Donations / endowments • Subsidies 	<ul style="list-style-type: none"> • Institutional commitments • Network development • Commitments from stakeholder communities • Incentives from funders

Source: Downs & Chen, 2016. A Portfolio Approach to a Sustainable Business Model for Scientific Data Stewardship. SciDataCon 2016.

Why Audit Research Data Centers?



- **Data producers** need to know where to deposit their data
 - Trust that their data will be preserved, curated, and disseminated
- **Data users** need to know where they can find data
 - Data that are vetted, described for use, and available in the future
- **Funders** need to know who to support for data management
 - Where services are reviewed routinely for continuous improvement
- **Publishers** need to know who to recommend for archiving data
 - Where referenced data will be persistently accessible and usable
- **Data professionals** need to know where they can practice
 - Apply their data management skills
 - Obtain professional development in data stewardship
- **Data centers** need to know how they are performing
 - Policies, procedures, and practices that need improvement

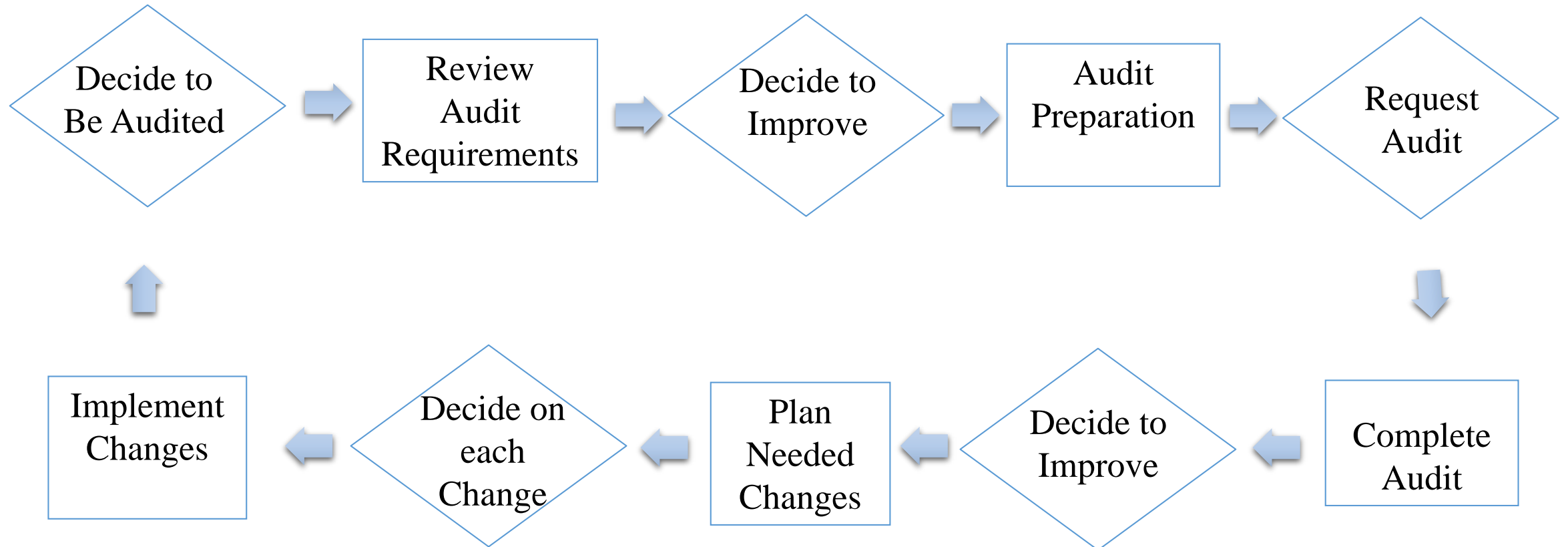
Data Repository Audit Considerations

Authority

		Internal	External
Frequency	Once	Internal one-time self-assessment	External one-time audit
	Periodic	Internal periodic self-assessments	External periodic audits

- Scope of audit:
 - Holistic vs targeted to specific capabilities, functions, or collections
- Approach should be based on objectives for the assessment
 - Why is the repository seeking an audit?
 - Which stakeholders are encouraging the audit?
 - Is improvement the primary objective or a credential?

Continuously Improving the Scientific Data Archive



Introduction to the CoreTrustSeal Certification Requirements: Using Requirements for Improvement, Self-Assessment and Certification



- Improvement
 - Identify opportunities for improving capabilities
- Self-Assessment
 - Identify areas to be improved and prepare for certification
- Certification
 - Apply for external review of repository capabilities and request certification
 - Validation period is 3 years from the date when certification is obtained

Review - Selected Examples of Completed Assignments for CoreTrustSeal Requirements: R1, R2, R5, and R6

Focus on Organization and Governance

- R1 Mission/Scope
 - Publicly accessible mission statement
- R2 Licenses
 - Intellectual property rights and restrictions on data for repository and users
- R5 Organizational Infrastructure
 - Sufficient funding, staffing, governance, and structure to operate
- R6 Expert Guidance
 - Guidance received to ensure that the repository meets community needs



CoreTrustSeal Requirement: R1 Mission/Scope



- The repository has an explicit mission to provide access to and preserve data in its domain.
- Responses should include evidence related to the following:
 - Your organization's mission in preserving and providing access to data, and include links to explicit statements of this mission.
 - The level of approval that the mission has received within the organization.
- Evidence for this Requirement could take the form of
 - *an approved public mission statement, roles mandated by funders, policy statement signed off by governing board.*

Submitted CoreTrustSeal Assignment: R1 Mission/Scope

- Eugene Brown - The main aim of a data repository is to ensure the accessibility and preservability of any data that will be stored in it. CERSGIS has therefore chosen to store its data on AWS because it is easily accessible, easy to configure and also has proven to be secured.
- Glory ENARUVBE, Afrigist - As an academic institution in earth observation, we manage a lot of data. One important system is
 - The Library management system where data preservation is important to avoid redundancy and to facilitate research. The system is made of a repository that gather student's projects information and data, most essentially EO data.
 - The system is managed by the Library Unit with technical support from ICT unit. With various access levels.

CoreTrustSeal Requirement: R2 Licenses



- The repository maintains all applicable licenses covering data access and use and monitors compliance.
- Responses should include evidence related to the following:
 - License agreements in use.
 - Conditions of use (Intellectual Property Rights, distribution, intended use, protection of sensitive data, etc.).
 - Documentation on measures in the case of noncompliance with conditions of access and use.



Submitted CoreTrustSeal Assignment: R2 Licenses

- Eugene Brown - Amazon ensures that all users are bound to its user and licenses agreement, these licenses encompass data ownership and distribution rights. CERSGIS also has its own data sharing policy that helps the distribution of data internally.
- Glory ENARUVBE , Afrigist - PMB is under a CECIL free license
 - the freedom to run the program, for any purpose;
 - the freedom to study how the program works, and change it so it does the computing as the user wishes. Access to the source code is a precondition for this;
 - the freedom to redistribute copies so users can help others;
 - the freedom to distribute copies of modified versions to others. By doing this, the whole community has a chance to benefit from changes made by other users. Access to the source code is a precondition for this.
 - Lack of documentation and demand of more customization

CoreTrustSeal Requirement: R5 Organizational Infrastructure



- The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission
- Responses should include evidence related to the following:
 - The repository is hosted by a recognized institution (ensuring long-term stability and sustainability) appropriate to its Designated Community.
 - The repository has sufficient funding, including staff resources, IT resources, and a budget for attending meetings when necessary. Ideally this should be for a three- to five-year period.
 - The repository ensures that its staff have access to ongoing training and professional development.
 - The range and depth of expertise of both the organization and its staff, including any relevant affiliations (e.g., national or international bodies), is appropriate to the mission.

Submitted CoreTrustSeal Assignment: R5 Organizational Infrastructure

- Eugene Brown - AWS has qualified professionals who ensuring that data kept on their web services are well preserved and protected yet easily accessible. They also have a lot of hands on materials and resources available that help users to manage data on the infrastructure. We leverage on this strength of AWS to build the capacity of our staffs to ensure that our data is appropriate.
- Glory ENARUVBE , Afrigist - Staff is competent to handle the repository but there is inadequate funding and trainings

CoreTrustSeal Requirement: R6 Expert Guidance



- The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).
- Responses should include evidence related to the following:
 - Does the repository have in-house advisers, or an external advisory committee that might be populated with technical, curation, data science, and disciplinary experts?
 - How does the repository communicate with the experts for advice?
 - How does the repository communicate with its Designated Community for feedback?

Submitted CoreTrustSeal Assignment: R6 Expert Guidance

- Eugene Brown - CERSGIS has a team of data managers and administrators as well as data analysts whose duty is to ensure that all data produced serve it purpose.
- Glory ENARUVBE, Afrigist - In-house adviser (ICT Unit)

Selected Quotes from Certified Repositories: R1 Mission/Scope



- The Center is committed to assuring the Center’s data producers and user base that the largest collection of polling and survey interview data in the world will be preserved and made accessible for the long term. The Roper Center’s mission statement is released under the authority of the Roper Center Board of Directors (Roper Center: Mission Statement <https://www.ropercenter.cornell.edu/about-center/roper-center-mission-statement>
 - Source: https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research_2020-22.pdf
- As written in the Odum Institute Data Archive Digital Preservation Policy, the mission of the Odum Archive is to “provide trusted long-term preservation and stewardship of research data assets to broaden scientific inquiry, promote research reproducibility, and foster data fluency now and into the future.” This mission complements the broader Odum Institute mission, which is “to foster groundbreaking social research that improves the lives of people in North Carolina and around the world.”
 - Source: <https://www.coretrustseal.org/wp-content/uploads/2020/10/Odum-Institute-Data-Archive.pdf>



Quotes from Certified Repositories: R2 Licenses

- Ongoing data providers sign a data provider agreement letter (Roper Center Sample Data Provider Agreement <https://ropercenter.cornell.edu/sites/default/files/wp-content/uploads/2014/12/Roper-Center-Data-Provider-Agreement-sample.pdf> Accessed 12/23/2020) that outlines the ongoing relationship between the organization and the Roper Center and gives the Roper Center the rights to archive and disseminate the data.
 - Source: https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research_2020-22.pdf
- The UNC Dataverse Terms of Use also describes the application of data usage license agreements to data submissions. Users who submit data materials to UNC Dataverse must accept the UNC Dataverse default Creative Commons “No Rights Reserved” (CC0) license that places the materials in the public domain, or provide an alternative custom data usage license agreement that establishes the acceptable uses, conditions, and/or restrictions of the submitted data materials.
 - Source: <https://www.coretrustseal.org/wp-content/uploads/2020/10/Odum-Institute-Data-Archive.pdf>

Quotes from Certified Repositories: R5 Organizational Infrastructure

- Externally audited financials and three and five-year financial projections present a solid case for long-term sustainability based on sensible membership growth, grant-supported special projects, and cost containment. Currently, annual income ... is derived primarily from the support of individual member organizations who pay an annual service subscription fee. With over 280 members in academic and non-academic or “affiliate” categories ... the outlook for continued operation ... beyond 5 years is positive
 - Source: https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research_2020-22.pdf
- The Odum Archive employs three permanent FTE who include the Assistant Director for Archives and two Research Data Archivists. All have LIS graduate degrees and A full-time temporary staff member, who holds a BSIS, provides additional archival support as needed. ... Archive staff work closely with Research Data Information Systems staff, which provides IT and research computing support. They include the Assistant Director for Research Data Information Systems, two Systems Programmers/Analysts, an Applications Analyst, and a Systems Administrator
 - Source: <https://www.coretrustseal.org/wp-content/uploads/2020/10/Odum-Institute-Data-Archive.pdf>



Quotes from Certified Repositories: R6 Expert Guidance



- The Board meets twice a year, but key members of the Board provide regular guidance to staff via phone calls and email In particular, members of the Archiving and Technology Committee provide essential guidance for archival planning ... through email, phone calls, and in-person meetings. In addition, the Transparency and Acquisitions Committee serves an essential role in guiding acquisition policy based on the evolving needs of the polling community served by the archive, in particular by setting minimum requirements for methodology disclosure.
 - Source: https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research_2020-22.pdf
- The Odum Archive operates with support from other Odum Institute units that provide in-house expertise to ensure Research Data Information Systems staff develop and maintain the underlying technical infrastructure The Odum Archive consults regularly with Odum Institute social science subject experts in qualitative, quantitative, mixed methods, data science, and survey research to identify trends
 - Source: <https://www.coretrustseal.org/wp-content/uploads/2020/10/Odum-Institute-Data-Archive.pdf>

CoreTrustSeal Requirements: Digital Object Management

- R7. Data integrity and authenticity
- R8. Appraisal
- R9. Documented storage procedures
- R10. Preservation plan
- R11. Data quality
- R12. Workflows
- R13. Data discovery and identification
- R14. Data reuse



CoreTrustSeal Requirements R7, R8, R9 and R10

Focus on Digital Object Management



- R7. Data integrity and authenticity
- R8. Appraisal
- R9. Documented storage procedures
- R10. Preservation plan

R7. Data integrity and authenticity

The repository guarantees the integrity and authenticity of the data

- Key issues:
 - Description of checks to verify that a digital object has not been altered or corrupted (i.e., fixity checks)
 - Documentation of the completeness of the data and metadata.
 - Details of how all changes to the data and metadata are logged.
 - Description of version control strategy.
 - Usage of appropriate international standards and conventions (which should be specified).
- Responses should include evidence related to the following questions:
 - Does the repository have a strategy for data changes? Are data producers made aware of this strategy?
 - Does the repository maintain provenance data and related audit trails?
 - Does the repository maintain links to metadata and to other datasets? If so, how?
 - Does the repository compare the essential properties of different versions of the same file? How?
 - Does the repository check the identities of depositors?



R8. Appraisal

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



- Responses should include evidence related to the following questions:
 - Does the repository use a collection development policy to guide data selection?
 - Does the repository have procedures in place to determine that the metadata required to interpret and use the data are provided?
 - What approach is used for data that do not fall within the mission/collection profile?
 - What is the approach if the metadata are insufficient for long-term preservation?
 - Does the repository publish a list of preferred formats?
 - Are checks in place to ensure that data producers adhere to the preferred formats?
 - What is the approach towards data that are deposited in non-preferred formats?
 - What is the process for removing items from your collection, also keeping in mind impact on existing persistent identifiers?

R9. Documented storage procedures


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



- Responses should include evidence related to the following questions:
 - How are relevant processes and procedures documented and managed?
 - Does the repository have a clear understanding of all storage locations and how they are managed?
 - Does the repository have a strategy for multiple copies? If so, what is it?
 - Are risk management techniques used to inform the strategy?
 - What checks are in place to ensure consistency across archival copies?
 - How is deterioration of storage media handled and monitored?

R10. Preservation plan

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

- Responses should include evidence related to the following questions:
 - Does the repository have a documented approach to preservation?
 - Is the level of responsibility for the preservation of each item understood? How is this defined?
 - Are plans related to future migrations or similar measures to address the threat of obsolescence in place?
 - Does the contract between depositor and repository provide for all actions necessary to meet the responsibilities?
 - Is the transfer of custody and responsibility handover clear to the depositor and repository?
 - Does the repository have the rights to copy, transform, and store the items, as well as provide access to them?
 - Are actions relevant to preservation specified in documentation, including custody transfer, submission information standards, and archival information standards?
 - Are there measures to ensure these actions are taken?

CoreTrustSeal Requirements R11, R12, R13 and R14

Focus on Digital Object Management

- R11. Data quality
- R12. Workflows
- R13. Data discovery and identification
- R14. Data reuse



R11. Data quality

The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality related evaluations



- For this Requirement, please describe:
 - The approach to data and metadata quality taken by the repository.
 - Does the repository have quality control checks to ensure the completeness and understandability of data deposited? If so, please provide references to quality control standards and reporting mechanisms accepted by the relevant community of practice, and include details of how any issues are resolved (e.g., are the data returned to the data provider for rectification, fixed by the repository, noted by quality flags in the data file, and/or included in the accompanying metadata?).
 - The ability of the Designated Community to comment on, and/or rate data and metadata.
 - Whether citations to related works or links to citation indices are provided.

R12. Workflows

Archiving takes place according to defined workflows from ingest to dissemination

- For this Requirement, responses should include evidence related to the following:
 - Workflows/business process descriptions.
 - Clear communication to depositors and users about handling of data.
 - Levels of security and impact on workflows (guarding privacy of subjects, etc.).
 - Qualitative and quantitative checking of outputs.
 - The types of data managed and any impact on workflow.
 - Decision handling within the workflows (e.g., archival data transformation).
 - Change management of workflows.



R13. Data discovery and identification

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

- Responses should include evidence related to the following questions:
 - Does the repository offer search facilities?
 - Does the repository maintain a searchable metadata catalogue to appropriate (internationally agreed) standards?
 - What persistent identifier systems does the repository use?
 - Does the repository facilitate machine harvesting of the metadata?
 - Is the repository included in one or more disciplinary or generic registries of resources?
 - Does the repository offer recommended data citations?



R14. Data reuse

The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data

- Responses should include evidence related to the following questions:
 - Which metadata are provided by the repository when the data are accessed?
 - How does the repository ensure continued understandability of the data?
 - Are data provided in formats used by the Designated Community? Which formats?
 - Are measures taken to account for the possible evolution of formats?



Assignment: Describe How Data Repository Meets CoreTrustSeal Requirements R8, R11, R12 and R13 by May 21st

- Assess your data repository in terms of each requirement, focusing on the questions in the Extended Guidance
 - CoreTrustSeal Requirements: <https://doi.org/10.5281/zenodo.3638211>
 - Extended Guidance: <https://doi.org/10.5281/zenodo.3632533>
 - Glossary: <https://doi.org/10.5281/zenodo.3632563>
- Identify ways in which the repository meets each requirement
 - Prepare a short description of a capability or feature of the repository that meets one or more aspects of each CoreTrustSeal requirement, or
- Identify opportunities for improving the repository to meet each requirement
 - Prepare a short description of a feature that could be improved or implemented in the repository to address one or more aspects of each CoreTrustSeal requirement



The SERVIR 2021 Training Program on Data Stewardship and the CoreTrustSeal Requirements videos are available online on the NASA SEDAC YouTube Channel:

https://www.youtube.com/channel/UCjUjAvV7M04SxxpM5wq4fMw?view_as=public

Thank you!

rdowns@ciesin.columbia.edu

Center for International Earth
Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY