

# SERVIR West Africa Training Program on Data Stewardship and the CoreTrustSeal Requirements: Session 3

Robert R. Downs, PhD

Center for International Earth Science Information Network (CIESIN)

The Earth Institute, Columbia University

Coordinated by AFRIGIST

May 26, 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

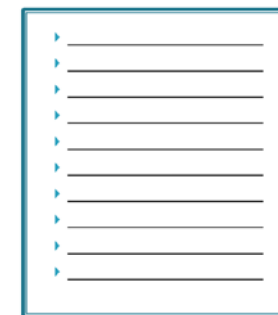
- Dates: May 5 (Session 1), May 12 (Session 2), May 26 (Session 3)
- Time: 10:00 a.m. to 11:30 a.m. EDT (14:00 - 15:30 UTC)
- Zoom Link:  
<https://columbiauniversity.zoom.us/j/94522663326?pwd=M3FxL2ZOU TE2K11UcjA4ZHdhdGNqdz09>
- Meeting ID: 945 2266 3326    Passcode: 231758

## 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 Session 2
- Review of Selected Examples of Completed Assignments for R8, R11, R12 and R13
- CoreTrustSeal Requirements R15 and R16
- Overview of selected CoreTrustSeal Requirements
- Review of Online Resources for Improving Data Stewardship and Attaining CoreTrustSeal Certification
- Closing Remarks



## Review and Discussion of Session 2

- Data Stewardship Concepts, Principles, and Certification Instruments
- Repository Sustainability
- CoreTrustSeal Certification requirements
- Organizational aspects of Data Stewardship
- CoreTrustSeal Requirements: R7, R8, R9, R10, R11, R12, R13, R14



# 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.

# GEOSS

## Data Sharing Principles

**T**



Transparency

**R**



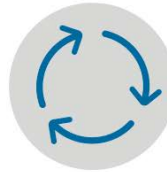
Responsibility

**U**



User focus

**S**



Sustainability

**T**



Technology

# GEOSS

## Data Management Principles

**F**

indable

**A**

ccessible

**I**

nteroperable

**R**

eusable

**C**

ollective Benefit

**A**

uthority to Control

**R**

esponsibility

**E**

thics

# GEOSS DSP, DMP & TRUST: Possible Synergies for Implementation

## Data Sharing Principles

### • DSP 1

- Open Data
- Without charge or restrictions
- GEOSS Data-CORE
- Subject to conditions
- Attribution



### • DSP 2

- Minimal restrictions on use
- Cost of reproduction and distribution



### • DSP 3

- Minimum time delay



## Data Management Principles

DMP-1: Metadata for Discovery



DMP-2: Online Access



DMP-4: Data Documentation



DMP-10: Persistent and Resolvable Identifiers



DMP-1: Metadata for Discovery



DMP-4: Data Documentation



DMP-6: Data Quality-Control



DMP-5: Data Traceability



DMP-7: Data Preservation



## TRUST Principles

Transparency



User Focus



Technology



Responsibility

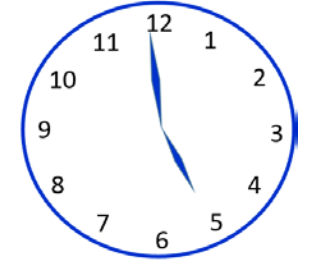


Sustainability



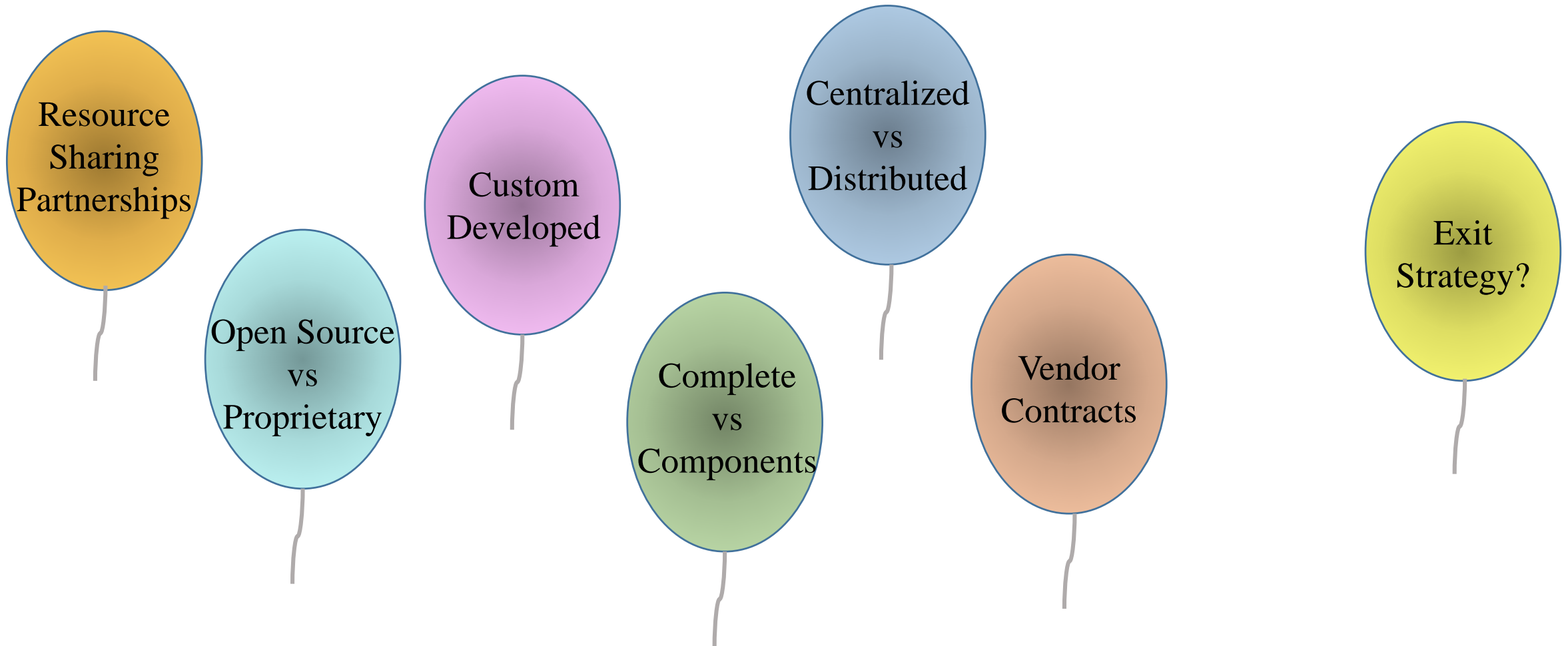


# Sustainable Stewardship of Scientific Data

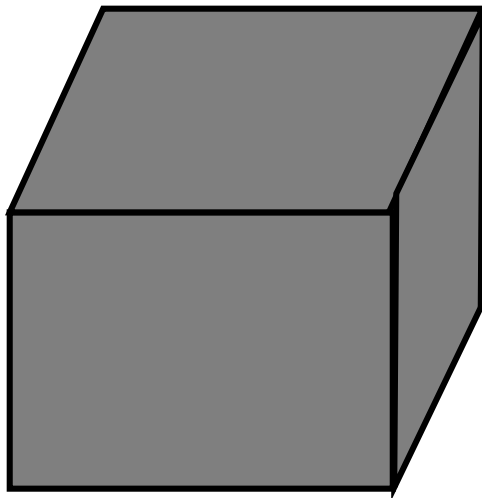


- Scientific data are valuable intellectual assets
  - Resources are needed to collect scientific data
  - Scientific data of recorded observations may be irreplaceable
  - Value of some scientific data may increase over time, especially in the context of long-term environmental and social change
- Stewardship 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)

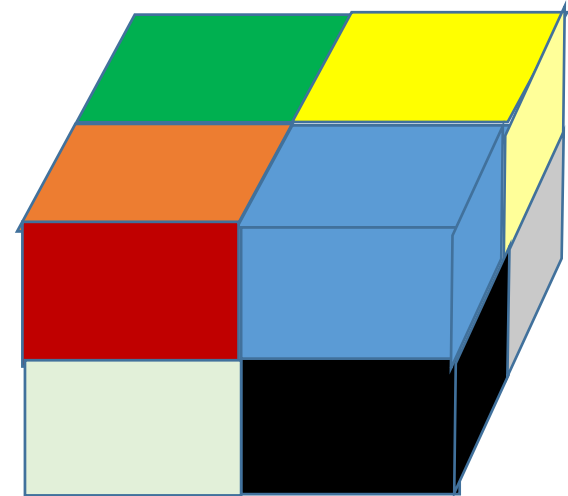
# Sustainable Selection of Data Repository Platform: Selected Options



## Repository Platforms Choices: Whole System Solutions vs Components



?



?

# Types of Shared Data Curation Services

## Repository Services

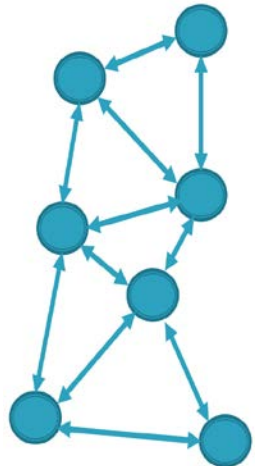
- Ingest
- Storage
- Discovery
- Access (Website)
- Analysis capabilities

## System Services

- Installation
- Hosting
- Support
- Maintenance
- Development

## Data Services

- Data Product Development
- Metadata
- Documentation
- PID Assignment
- Long-Term Preservation



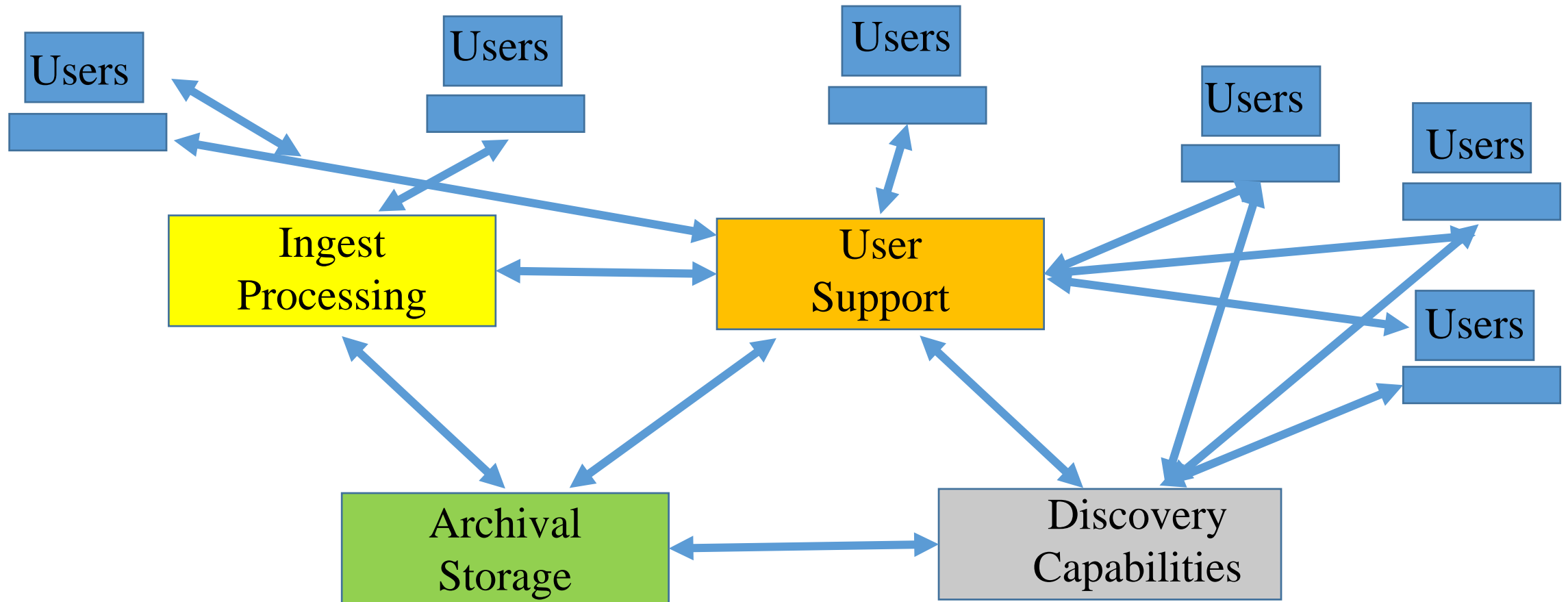
## Review Services

- Acquisition Appraisal
- Product review
- Repository Certification

## Organizational Services

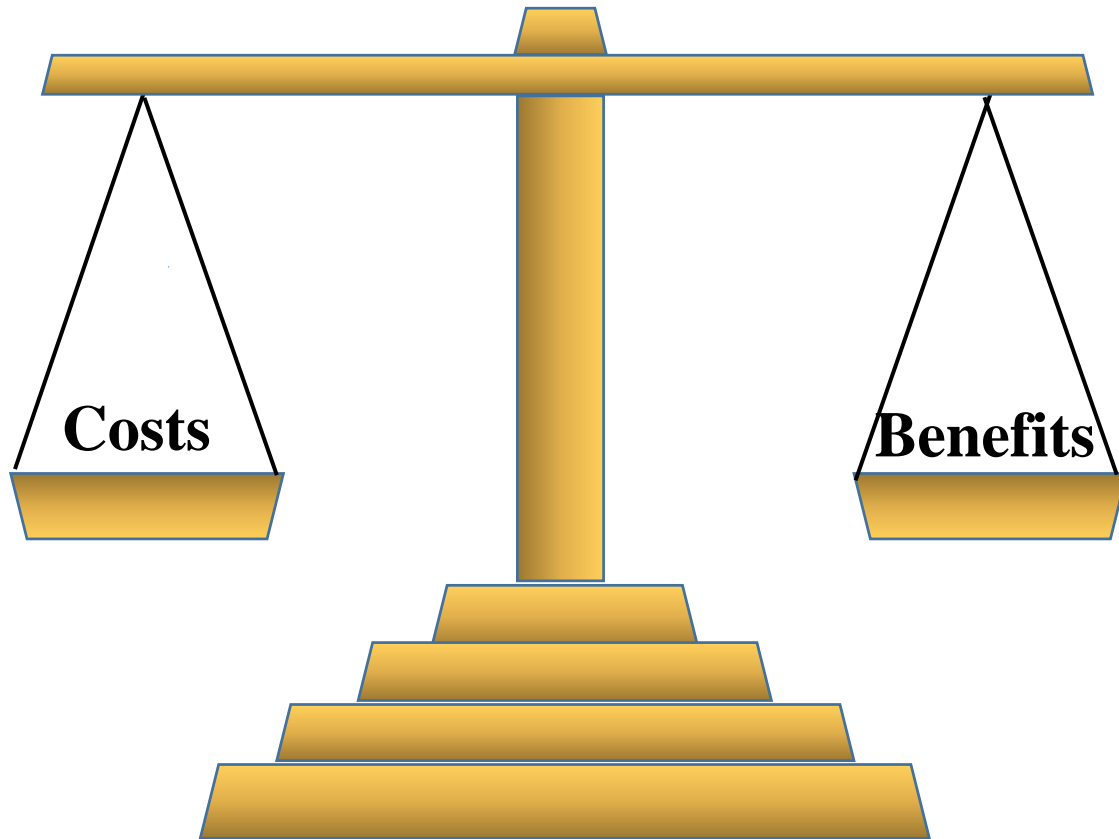
- Strategic Planning
- Technology Selection
- Staffing
- Training

# Example of a Distributed Data Repository Implementation



Source: Downs, Lehnert, Donaldson, Zhang, and Silva. 2020. Repository Implementation Approaches: Choices and Decisions. Research Data Alliance 15th Plenary (RDA P15).

# Sustainability Implications of Data Repository Services



## Selected Data Service Operations

- Acquisition
- Product Development
- Service Development
- Curation
- Peer-Review
- Support
- Preservation

# 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: R8, R11, R12, and R13

## Focus on Digital Object Management



- R8. Appraisal
  - The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.
- 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
- R12. Workflows
  - Archiving takes place according to defined workflows from ingest to dissemination
- R13. Data discovery and identification
  - The repository enables users to discover the data and refer to them in a persistent way through proper citation.



## **Submitted CoreTrustSeal Assignment: R8 Appraisal**

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

- Anastasia Wahome: "RCMRD has an existing data management policy, the policy guides all aspects from data collection, data to be archived and data to be published. Within the policy there are metadata guidelines to be employed based on ISO 19115 standards. This guides on the mandatory and optional fields to be included when creating metadata contents. There are various tools employed to ensure metadata schemas are adhered to. "
- Eugene Brown: "The repository has procedures and template for depositors to ensure that the required metadata is attributed to the data. Data that don't fall in the mission profile (such as none authoritative data) are kept and used as secondary data. Our repository could publish various formats but CERSGIS has adopted the ISO 19115 hence our Geonetwork is configured to received only this type of metadata therefore when we received any metadata other than the prescribed format, we outline it to the ISO 19115 format."

## **Submitted CoreTrustSeal Assignment: 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.

- Anastasia Wahome: The RCMRD data management policy document contains a chapter (2.3.10) on Data Integrity, Validation and Correction. In brief this chapter indicates the following: The data creators are responsible for data integrity, responding to data accuracy and inconsistencies where necessary. Data managers provide the modality for data quality checks and assurances, whereas data administrators ensure the checks are carried out for all datasets before storage and listing in the inventory. Data users on the other hand, have the responsibility to assist in data quality checks, through providing the right feedback. Citations for related works are accompanied within the metadata files."
- Eugene Brown: "There are personnel tasked to perform data quality assurance before the data is uploaded. Also, our Geo-network is also configuring in such a way that some information on the metadata (e.g., spatial reference, thumbnails, the scale of the map, contact details of data owners and some keywords) cannot be bypassed when uploading data. We are also allowing some comments from the community on our data. Owners of data and the intended use of the data is properly cited in all our data."

## Submitted CoreTrustSeal Assignment: R12 Workflows

Archiving takes place according to defined workflows from ingest to dissemination.

- Anastasia Wahome: "RCMRD data management policy has explicit roles in regard to the data collection and maintenance, availability and integration, documentation, handling procedures as well as access levels. The data committee also defines the roles and responsibilities of the various actors involved; from the data managers, administrators, creators, system administrators and the users."
- Eugene Brown: "We provide descriptions of the data in our repository, e.g., we inform users on the general purpose of data (especially open data and data for research purpose only). We perform quality and quantity checks on the data e.g. We check for attributes and their corresponding records in the data. We also check on the themes, keywords of the data, and the data constraints. The types of data we manage to include vector and raster dataset, spreadsheet, map services and others."

## **Submitted CoreTrustSeal Assignment: R13 Data Discovery and Identification**

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

- Anastasia Wahome: "RCMRD repositories offer search capabilities using defined categories, keywords and tags. There exists a metadata catalogue created under international standards that allows users to search, query and access to data. The repositories do also allow for batch harvesting of metadata from specific resources."
- Eugene Brown: "Our repository has search functionalities that allow users to search by keywords. We emphasize important keywords for our data and we ensure that our data has standard metadata that makes it easier to search for a keyword. Our repository allows data harvesting by authorized users."

## Quotes from Certified Repositories: R8 Appraisal



- The Roper Center’s Acquisition Policy defines accepted sampling design and data collection methodologies allowed for submission, as well as minimum disclosure requirements, and provides guidance in prioritization of collection development. (Roper Center Acquisition and Transparency Policy <https://ropercenter.cornell.edu/roper-center-transparency-and-acquisitions-policy> Accessed 7/29/2020.)
  - Source: [https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research\\_2020-22.pdf](https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research_2020-22.pdf)
- To ensure that data housed in the Odum Archive meet standards of quality and usability according to applicable disciplinary norms, the Odum Archive enforces the Odum Institute Data Archive Collection Development Policy, which establishes the selection and appraisal criteria that determine the suitability of data submissions for inclusion in the Odum Archive collections. These criteria, which are informed by the requirements and preferences of the Designated Community, are based on substantive value of the data as well as the quality and rigor of the data.
  - Source: <https://www.coretrustseal.org/wp-content/uploads/2020/10/Odum-Institute-Data-Archive.pdf>

## Quotes from Certified Repositories: R11 Data Quality



- The Center requires that any Submission Information Package (SIP) contain the documentation necessary for appraising the quality of deposited data, e.g., a methodology report, questionnaires, and interviewer instructions (Roper Center Data Deposit Form  
<https://ropercenter.cornell.edu/sites/default/files/pdf/DataDepositFormfillable2.pdf>  
Accessed 7/29/2020).
  - Source: [https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research\\_2020-22.pdf](https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research_2020-22.pdf)
- The Odum Archive addresses data quality requirements in the Odum Institute Data Archive Collection Development Policy based on the intended outcome that data in the Odum Archive be “independently understandable for informed reuse,” thus enabling the Designated Community to make quality-related evaluations. Therefore, the Collection Development Policy sets requirements for quality standards for data accuracy and interpretability, completeness and readability of documentation, and acceptability of file formats.
  - Source: <https://www.coretrustseal.org/wp-content/uploads/2020/10/Odum-Institute-Data-Archive.pdf>

## Quotes from Certified Repositories: R12 Workflows



- A conceptualized overview of the data life cycle, its phases, and processing workflows is available for data producers, academics, and other researchers on the Roper Center's website (Roper Center Data Curation <https://ropercenter.cornell.edu/data-archiving/data-curation-and-architecture> Accessed 7/29/2020). This informs our various stakeholders of the effort incurred to enhance studies with meaningful information, making them as complete as possible, usable, and independently understandable for future researchers.
  - Source: [https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research\\_2020-22.pdf](https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research_2020-22.pdf)
- The Odum Institute Data Archive Curation Workflow document outlines the general archiving workflow in four stages: deposit (transfer of SIP from the data submitter to the archive), triage (file review and data management planning), processing (AIP generation), and access (delivery of DIP to data consumers). This information along with Odum Archive policies and guidance documents are published on the Odum Archive website for purposes of transparency to depositors and users.
  - Source: <https://www.coretrustseal.org/wp-content/uploads/2020/10/Odum-Institute-Data-Archive.pdf>

## Quotes from Certified Repositories: R13 Data discovery and identification

- For datasets from national U.S. populations, users can also search the question-level Roper iPoll database by keyword within the text of the question, organization, controlled-language topic, and dates. Once a question of interest is found, the user can easily download the documentation and individual-level data, as well as any supporting materials available, such as topline reports or data tables.
  - Source: [https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research\\_2020-22.pdf](https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research_2020-22.pdf)
- Users can discover datasets using faceted browsing that filters dataset records based on defined categories (e.g., author name, subject, production date) mapped to standardized metadata fields. Basic search performs a full-text query across all metadata fields to return dataset records that contain specified terms. Advanced search functionality allows for more directed queries by allowing users to search for dataset records with specified terms appearing within a particular metadata field.
  - Source: <https://www.coretrustseal.org/wp-content/uploads/2020/10/Odum-Institute-Data-Archive.pdf>





# CoreTrustSeal Requirements

## Technology: R15 and R16



- R15. Technical infrastructure.
  - The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.
- R16. Security.
  - The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.



## R15. Technical Infrastructure

Repositories need to operate on reliable and stable core infrastructures that maximizes service availability. Furthermore, hardware and software used must be relevant and appropriate to the Designated Community and to the functions that a repository fulfils.

- Responses should include evidence related to the following questions:
  - What standards does the repository use for reference? Are these international and/or community standards? How often are these reviewed?
  - How are the standards implemented? Are there any significant deviations from the standard?
  - Does the repository have a plan for infrastructure development? If so, what is it?
  - Is a software inventory maintained and is system documentation available?
  - Is community-supported software in use? Please describe.
  - Are availability, bandwidth, and connectivity sufficient for the Designated Community?
  - Does the repository have a disaster plan, a business continuity plan, procedures and arrangements to provide swift recovery or backup of essential services in the event of an outage?

## Quotes from Certified Repositories: R15 Technical Infrastructure



- The Roper Center maps its archival metadata to the DDI standard. ... uses the ISO 3166-1 alpha-2 standard for geographical metadata information. Web applications produced by the Roper Center follow HTTPS protocols and transfer data using JSON. ... All systems used currently to receive and archive files, write metadata to the database, track workflow and versioning, and publish studies, as well as all external search tools, were developed internally for Roper Center by its IT software development team ... in Java.
  - Source: [https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research\\_2020-22.pdf](https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research_2020-22.pdf)
- Redundant power and network links across servers help to eliminate any single point of failure ... Cluster storage is provided by two redundant NetAPP storage appliances ... in compliance with ISO 27002: Code of Practice for Information Security Controls, these servers are protected by both a local software firewall and a UNC-maintained campus hardware firewall ... Regular security scans ... monitor for system vulnerabilities and for oversight of firewall configuration requests. ... patch management system is in place ...
  - Source: <https://www.coretrustseal.org/wp-content/uploads/2020/10/Odum-Institute-Data-Archive.pdf>



## R16. Security

The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

- Responses should include evidence related to the following:
  - Your IT security system, employees with roles related to security (e.g., security officers), and any risk analysis tools (e.g., DRAMBORA2) you use.
  - What levels of security are required, and how these are supported.
  - Authentication / authorization procedures to securely manage system access to (e.g., Shibboleth, OpenAthens).
  - What mechanisms are in place to prevent, detect, and respond to a security incident.
  - How is technical infrastructure security controlled by the repository or the host/outsource institution?
  - Are authentication and authorization procedures in place sufficient to guarantee the security of the data holdings at each stage of the workflow (e.g., by requiring two-factor authentication for sensitive data)?
  - Which company security policies are in place to govern the security of all systems, including network security, intrusion checks, physical facility security, and password policy?

## Quotes from Certified Repositories: R16 Security



- Twice-daily backup services are run to replicate the archive on multiple servers ... The web application is comprised of multiple servers load-balanced to handle a large volume of users and prevent against a server going down. ... Access to the Data Center is granted by an authorized proximity card ... Multiple data storage locations ... ensure that the data archive is protected from local disaster. Data is backed up ... three times a day ...
  - Source: [https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research\\_2020-22.pdf](https://www.coretrustseal.org/wp-content/uploads/2021/04/Roper-Center-for-Public-Opinion-Research_2020-22.pdf)
- These computing centers are designed with redundant power and cooling to ensure continued access ... Physical access controls are in place and enforced by staff 24/7 to prevent unauthorized entry ... Remote access to these systems is controlled by Virtual Private Networks, Virtual Local Area Networks, and strong password authentication ... systems are protected by both a local software firewall and ... hardware firewall to prevent unauthorized access ... Regular patch management strategies are employed ...
  - Source: <https://www.coretrustseal.org/wp-content/uploads/2020/10/Odum-Institute-Data-Archive.pdf>

# Overview of Selected CoreTrustSeal Requirements



Review of requirements that were not assigned as homework

- Context
  - R0
- Organizational Infrastructure
  - R3, R4
- Digital Object Management
  - R7, R9, R10, R14

# CoreTrustSeal Focus on Context: R0



Context enables general understanding of the repository

- Brief Description of Designated Community
  - "applicant understands the scope, knowledge base, and methodologies—including preferred software/formats—of the user community or communities they are targeting"
- Level of Curation
  - Note: a repository may offer multiple levels of curation
  - Indicate how rights and expertise enable curation
  - Describe how level of curation affects workflow
- Insource Partnerships
  - Units within host organization (institution, agency)
  - Explain the relationships among partners
- Outsource Partnerships
  - Identify all outsource partners that are described within the application

## Defining the Designated Community

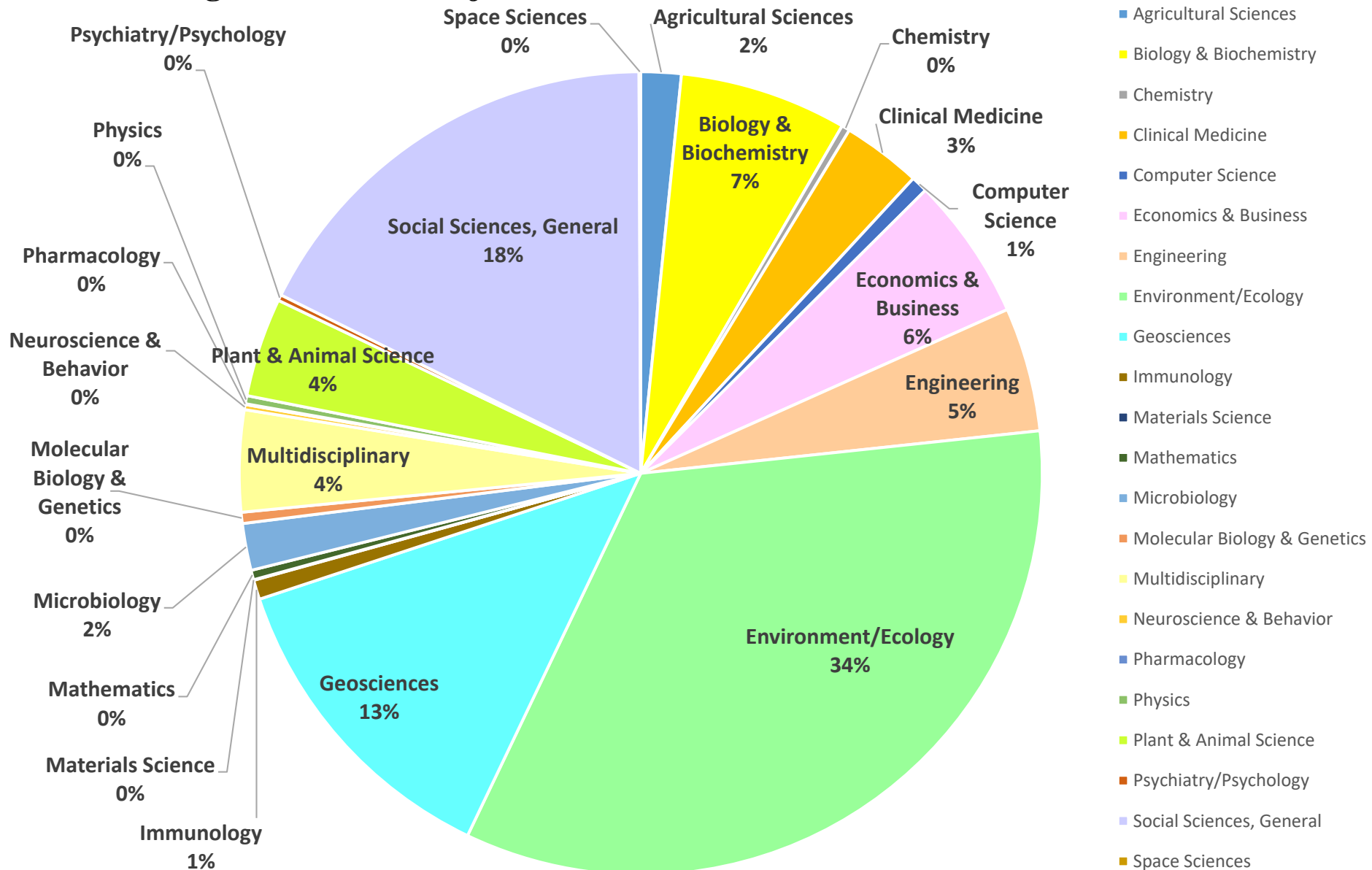


- "Designated Community: An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time." (CCSDS, OAIS, 2012, 1-11)
- Both domain repositories and generalist repositories may provide public access to data
  - Domain repositories may prioritize service for users of one or more disciplines, themes, or topic areas
  - Generalist repositories may prioritize service for researchers, generally, or prefer not to prioritize services for any segment of users



**Describing the Designated Community:  
An Example from the NASA SEDAC**

# SEDAC Communities Served: 22 Major Fields of Science (216 Categories and 92 Subject Areas) of Journals that Cited SEDAC Data 2004-2013



# CoreTrustSeal Focus on Organizational Infrastructure: R3 and R4

Organizational infrastructure enables the repository to meet its goals

- R3 Continuity of Access
  - Address how risks to repository sustainability are addressed
  - Describe contingency plans to ensure continued data access for unanticipated events or circumstances
- R4 Confidentiality / Ethics
  - Would the repository know if it received confidential or sensitive data?
  - Indicate how confidential and sensitive data are handled and protected
  - Explain how repository meets ethical norms of the Designated Community
  - Are decisions influenced by the ethical principles of the community?



# CoreTrustSeal Focus on Digital Object Management: R7, R9, R10, & R14



Digital object management facilitates operational data stewardship

- R7 Data Integrity and Authenticity
  - How are inventories, fixity checks, and version control performed?
- R9 Documented Storage Procedures
  - Are multiple storage locations designated for data?
- R10 Preservation Plan
  - Describe policies, plans, or procedures for the preservation of data holdings
- R14 Data Reuse
  - How are data described to facilitate understanding by potential future users?

# **Review of Online Resources for Improving Data Stewardship and Attaining CoreTrustSeal Certification**

# Online Resources for Improving Data Stewardship and Attaining CoreTrustSeal Certification

- CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022. (Version v02.00-2020-2022): <https://doi.org/10.5281/zenodo.3638211>
- CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022: <https://doi.org/10.5281/zenodo.3632532>
- CoreTrustSeal Trustworthy Data Repositories Requirements: Glossary 2020–2022: <https://doi.org/10.5281/zenodo.3632562>
- French translation of Extended Guidance for the initial version of the CoreTrustSeal requirements: <https://www.rd-alliance.org/system/files/documents/CoretrustsealFR.pdf>
- 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#](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#](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](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)

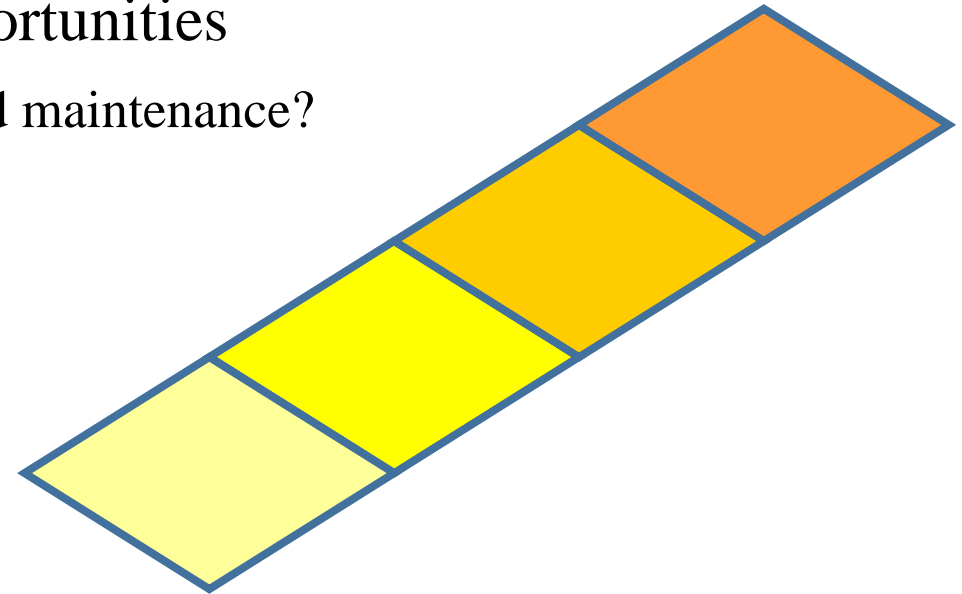
- Downs RR. 2019. International Standards for Trustworthy Data Repositories. National Institutes of Health (NIH) Trustworthy Data Repositories Workshop. Rockville, MD. <https://doi.org/10.7916/d8-h7xf-ha70>
- Downs RR. 2021. Adopting the TRUST Principles for Digital Repositories with the GEOSS Data Management 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>

## **Closing Remarks**

Improving data repository capabilities is an ongoing process

# Data Repository Improvement: Getting Started

- Consider the CTS requirements in terms of current and long-term needs
  - Establish criteria for prioritizing issues to be addressed
  - What capabilities are we trying to achieve?
- Balance resource demands by leveraging opportunities
  - What resources are needed for implementation and maintenance?
- Collaborate with other data repositories
  - Learn from the experiences of others
- Planning for improvements
  - Assess the feasibility of adoption



# Initiating Organizational Change for CoreTrustSeal Certification

- Obtain buy-in from stakeholders
  - Identify the benefits and the costs
  - Present the business case
  - Justify efforts in terms of current objectives and available resources
  - Set objectives for both short-term and long term goals
- Bring others along
  - Build on the business case, in terms of incremental steps
  - Describe the value of the impact (users)
  - Tie to current priorities
- Attaining efficiency
  - Foster learning
  - Build on current tools and services
  - Conduct iterative levels of testing proposed changes

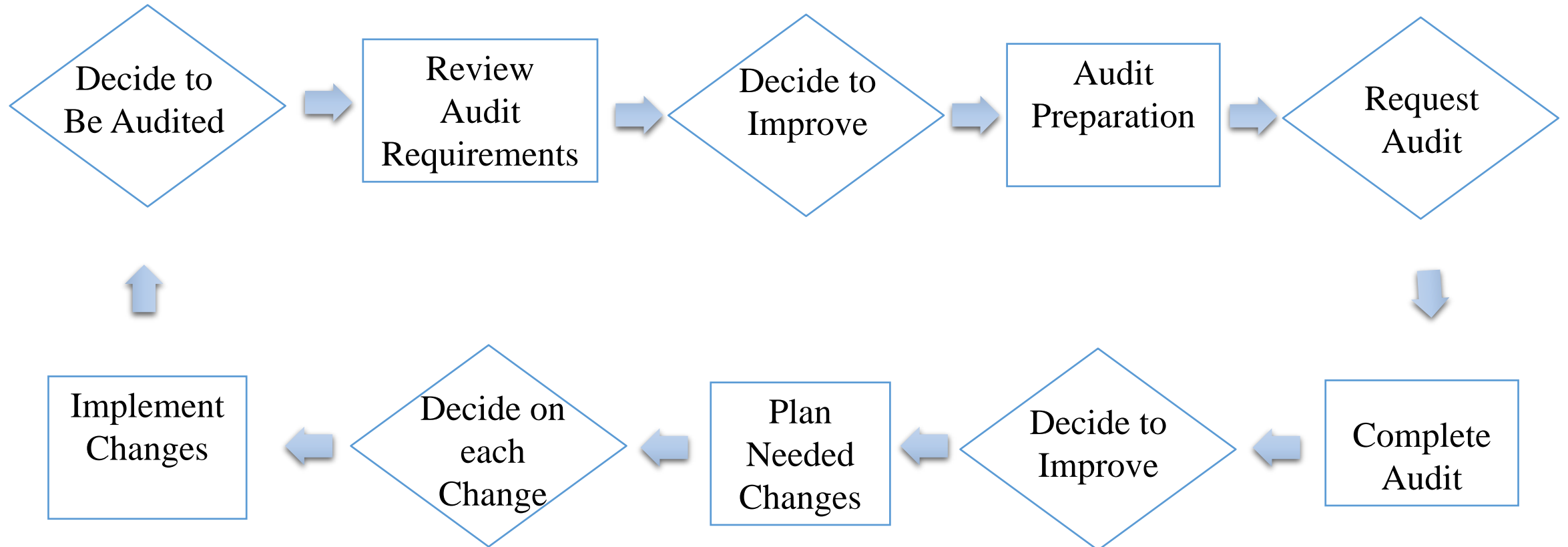


# Preparing the Data Repository for CoreTrustSeal Certification

- Perform an initial self-assessment to identify current gaps
  - Describe gaps on a worksheet
- Prioritize gaps to be addressed
  - Identify alternative opportunities to close gaps and re-prioritize as needed
  - Leverage synergistic opportunities that address multiple gaps
- Complete efforts to address gaps in prioritized order
  - Update the self-assessment to reflect completed efforts
- Submit CoreTrustSeal application after meeting each requirement
  - Every requirement does not have to be fully implemented



# Continuously Improving the Scientific Data Archive



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](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