Assessment Information

CoreTrustSeal Requirements 2017–2019

Repository:	Research Data Archive at the National Center for Atmospheric Research (NCAR)
Website:	https://rda.ucar.edu
Certification Date:	21 June 2019
This repository is owned by:	**National Center for Atmospheric Research (NCAR), Managed by University Corporation for Atmospheric Research (UCAR)**

Research Data Archive at the National Center for Atmospheric Research (NCAR)

Notes Before Completing the Application

We have read and understood the notes concerning our application submission.

True

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background & General Guidance

Glossary of Terms

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type. Select all relevant types from:

Domain or subject-based repository

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Brief Description of Repository

The Computational and Information Systems Laboratory's (CISL) Data Engineering & Curation Section (DECS) at the National Center for Atmospheric Research (NCAR) maintains the Research Data Archive (RDA) to support Atmospheric and related Geosciences research. The RDA is developed to serve the research needs at NCAR and in the associated University Corporation for Atmospheric Research (UCAR) community, However, since the RDA is an open resource, the global community also frequently accesses it.

The RDA provides a rich information resource through a large and growing collection of data sets that support scientific studies in climate, weather, and Earth System modeling. To meet research community needs, the RDA continuously adds new and augments existing data content. Additionally, the RDA strives to minimize the researchers' data work burdens by hosting the needed reference data sets with added personal consulting services.

Access to the RDA's data content is frequently improved with new tools, web services, and high performance computing (HPC)-driven workflows that can extract data specified by the users and from multi-terabyte data sets. The RDA also provides easy access pathways, including:
1)Local connection to the CISL HPC from a directly connected central file system.
2)Via the web through dedicated user interfaces
3)Standard APIs that support machine-to-machine interoperability, and creating on demand for individuals various customized data packages from large and heterogeneous collections.

All efforts to improve the RDA focus on enhancing the productivity of the weather and climate research communities. For additional details, see:
1)"About the RDA" and "Mission Statement": https://rda.ucar.edu/#!about
2)CISL Mission Statement: https://www2.cisl.ucar.edu/org/about
3)NCAR Mission Statement: See page 1 of
https://ncar.ucar.edu/sites/default/files/documents/related-links/2017-06/NCAR_Strat_Plan_Final_102014.pdf

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Brief Description of the Repository's Designated Community.

The RDA designated community consists of the NCAR and UCAR user communities (https://nar.ucar.edu/2017/cisl/expand-content-and-access-rda). DECS leadership and staff engage our designated community through a variety of mechanisms as detailed in R6, including CISL sponsored surveys, CISL user outreach and educational seminars, and international meetings, such as the annual American Meteorological Society and American Geophysical Union meetings. DECS staff also directly engage users by email and phone to respond to questions and provide support for data service inquiries. For additional background on the RDA user communities and their research goals, see:
1)NCAR Strategic Plan:
https://ncar.ucar.edu/sites/default/files/documents/related-links/2017-06/NCAR_Strat_Plan_Final_102014.pdf
2)UCAR Strategic Plan:
https://www.ucar.edu/sites/default/files/documents/related-links/2017-05/UCAR_Strat_Plan_2015.02.25.pdf
3)NCAR overview: https://ncar.ucar.edu/who-we-are
4)UCAR overview: https://www.ucar.edu/who-we-are

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Level of Curation Performed. Select all relevant types from:

B. Basic curation – e.g. brief checking; addition of basic metadata or documentation, C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation, D. Data-level curation – as in C above; but with additional editing of deposited data for accuracy

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Accept

Comments

The following are estimates of the percentage of total number of RDA datasets according to curation level performed. An updated inventory of curation level performed on all RDA datasets is currently being completed.
B. Basic curation: 100% of the RDA's dataset collections
C. Enhanced curation: estimated to be 20% of the RDA's dataset collections
D. Data-level curation: estimated to be 5% of the RDA's dataset collections

As part of the curation process, data submitters are asked to:
1)Validate dataset metadata.
2)Verify that dataset collection's data files are organized and accessible according to their expectations.
3)If data or metadata need to be modified due to systematic problems that are discovered, DECS staff maintain communications with data submitters as needed to notify them of these changes, and ask them to validate these changes.
Additional information on the curation process can be found in R7 and R11.

For additional background on the rationale used to determine dataset curation level, see:
https://rda.ucar.edu/rdadocs/RDA_Dataset_Curation_Level.pdf

For additional background on the data ingest to dissemination workflow as detailed in R12, see:
https://rda.ucar.edu/rdadocs/RDA_data_ingest_to_dissemination_workflow_overview.pdf

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Outsource Partners. If applicable, please list them.

Not Applicable

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Other Relevant Information.

Additional background on the RDA, including metrics on data holdings and access, is provided in the CISL Annual Report: https://nar.ucar.edu/2017/cisl/expand-content-and-access-rda

RDA’s record at re3data.org: https://www.re3data.org/repository/r3d100010050

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

ORGANIZATIONAL INFRASTRUCTURE

I. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The Data Engineering & Curation Section’s (DECS) core mission (see Reference #1 below) of “supporting atmospheric and related sciences research by maintaining and developing the Research Data Archive, and reinforcing its utility with expert consulting” is driven by the mission components from both the National Center for Atmospheric Research (NCAR) and NCAR’s overarching sponsor, the National Science Foundation (NSF). These components include:

1)Computational and Information Systems Lab (CISL) mission (see Reference #2 below): “CISL's mission is to support and advance the geosciences with … data management ...”
2)NCAR mission (see Reference #3 below): “To support, enhance, and extend the capabilities of the university community and the broader scientific community, nationally and internationally”
3)University Corporation for Atmospheric Research (UCAR) mission (see Reference #4 below): “UCAR's mission is to empower our Member Institutions, our National Center, and our Community Programs by ... managing Unique resources”
4)NSF NCAR facility statement (See Reference # 5 below): “Available to university and other scientists, as well as NCAR scientific personnel, the facilities at NCAR serve the entire atmospheric sciences research community and part of the ocean sciences community. These facilities include a computing center that provides supercomputer resources and services ... for archiving, manipulating, and visualizing large data sets”

DECS leadership works with CISL management to review its mission statement periodically to ensure that it remains aligned with the organizational strategic plan updates and the broader NCAR and NSF missions.

References:
1)The DECS Mission statement: https://rda.ucar.edu/#!about
2)The CISL Mission statement: https://www2.cisl.ucar.edu/org/about
3)The NCAR Mission statement: See page 1 of https://ncar.ucar.edu/sites/default/files/documents/related-links/2017-06/NCAR_Strat_Plan_Final_102014.pdf
4)The UCAR Mission statement: See page 5 of https://www.ucar.edu/sites/default/files/documents/related-links/2017-05/UCAR_Strat_Plan_2015.02.25.pdf
5)NSF sponsorship information: https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=12809&org=AGS&more=Y#more

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

II. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The Research Data Archive (RDA) is a component of the National Center for Atmospheric Research (NCAR), which is managed by the University Corporation for Atmospheric Research (UCAR). Therefore, all of RDA's legal agreements are with UCAR.

All RDA datasets are publically accessible according to the Terms of Use for UCAR Data Repositories (https://rda.ucar.edu/index.html#repository_terms_conditions) unless there are additional requirements put forth by an institutional data submitter.

The additions of extra requirements are rare instances. They only occur to support hosting of valued international dataset collections, where the datasets need to abide by additional conditions before public sharing can take place. Users for these specific collections must agree to supplementary terms of use, in addition to the standard UCAR terms of use.

To access RDA data, all users are required to register for accounts with the RDA, which then allows the Data Engineering & Curation Section (DECS) to provide them with the appropriate license settings in their user profiles. When a user agrees to the supplementary terms of use, an authorized role is added to their RDA user profile to give them access to the restricted dataset collection. Additional background on this scenario can be found in a related RDA blog post: http://ncarrda.blogspot.com/2016/08/how-to-access-restricted-data-set.html

Additional, specific details related to R2 requirements are provided below.

For both “License agreements in use” and “Conditions of use (distribution, intended use, protection of sensitive data, etc.)”:
The RDA requires users to agree to the Terms of Use for UCAR Data Repositories (link provided above) in order to complete the RDA user registration process. The Terms of Use for UCAR Data Repositories grants a licensed subject to the terms and conditions of the Creative Commons Attribution 4.0 International license (https://creativecommons.org/licenses/by/4.0/legalcode). This is a “Click-through” agreement, and the text of Terms of Use can be viewed using the following link: https://rda.ucar.edu/index.html#repository_terms_conditions

Documentation on measures in the case of noncompliance with conditions of access and use:
The following examples of relevant statements are included in the UCAR website terms of use (https://www.ucar.edu/terms-of-use), which are also applicable to the Terms of Use for UCAR Data Repositories:
1)“Some of these Materials are governed by their own specific terms of use and copyrights. In the absence of specific terms of use, these Terms of Use shall apply. The burden of determining that use of any Materials on or linked to a Site is permissible, rests with you, the user.”
2)“In the event you breach these Terms of Use or any Material-specific terms or you infringe any copyrights, UCAR may pursue any and all legal and equitable remedies available to it, although UCAR is under no legal obligation to do so.”

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

III. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The Research Data Archive (RDA) guarantees data preservation and access for a minimum of five years as described in the “Rights / Terms, Conditions for Use, collaboration, and Ownership” section of the following page:
https://rda.ucar.edu/#!daas/terms-and-conditions

Justifications for this minimum preservation period include:
1)The 5-year preservation guarantee aligns with the National Center for Atmospheric Research (NCAR)/University Corporation for Atmospheric Research (UCAR) cooperative agreement funding cycles as described in R5, and is supported by the RDA's history of long-term institutional stability, also described in R5.
2)Storage and computational infrastructure used to maintain and enable the RDA's services have been consistently provided and renew at NCAR by various precursors of Computational and Information Systems Lab (CISL) since the RDA's inception in 1965, as detailed in R5.
3)Staffing levels have been maintained at desired levels to support the minimum 5-year preservation commitment since the RDA's inception in 1965, as described in R5.
RDA data holdings are maintained by qualified data curation staff who have educational backgrounds in relevant disciplinary fields, and participate in continuing education opportunities as described in R5 and R15.

Ongoing stakeholder engagement is essential to ensure the continued accessibility and availability of RDA data holdings for the medium term (three- to -five years) and long-term (> 5 years). NCAR, CISL, and the Data Engineering & Curation Section (DECS) actively engage with stakeholders at a variety of organizational levels in order to align strategies and services (including the RDA) with the stakeholders' needs. Specifically, formal and informal avenues for stakeholders to provide feedback have assisted the DECS in keeping the RDA aligned with the community's needs, and allowed the RDA to continue to be an essential resource for NCAR and the broader Earth Systems Sciences community. Detailed examples of stakeholder engagement are provided in R6.

In addition to the examples of regular stakeholder engagement described in R6, the National Science Foundation (NSF) has recognized that NCAR, with a notable focus on the RDA, has data management skills that are not only advanced, but also evolving. As an NSF-requested action, NCAR held the GeoDaRRS Workshop (http://dx.doi.org/10.5065/D6NC601B) in August of 2018 to help define NSF research community data management needs. The goal of this large community workshop was to help the NSF and related data repositories better support geoscience researchers when they ask the question: "Where do I put my data?". Geoscience researchers are now being asked by funding agencies and scientific publishers to archive and cite data to support open access, but often struggle to understand and fulfill these requirements. The workshop brought together over sixty individuals from multiple stakeholder groups to discuss data management and archiving challenges and opportunities within the geosciences. The relevant stakeholder communities represented by the attendees included geoscience researchers, technology experts, scientific publishers, funders, and data repository personnel. The RDA's supportive role for the community and collaboration with NSF on data management issues will continue, and hence, will likely allow the RDA to receive long-term ongoing reciprocal support from the NSF.

In an effort to remain sustainable, the RDA has a policy in place to purge non-observing based datasets (e.g. model outputs) from the archive after five years if certain requirements are not met as described under the "Dataset Withdrawal Policy" section of the following page: https://rda.ucar.edu/#!daas/terms-and-conditions . Additionally, the RDA has cost recovery mechanisms in place to support long term sustainability for specific use cases as described in the "Cost Recovery" section of the following page: https://rda.ucar.edu/#!daas/terms-and-conditions . For additional information on

RDA’s long-term preservation policy and sustainability strategies, please see R10.

If stakeholders determine that the RDA is no longer a community valued resource and funding is cut or ceased, a strategy is in place to wind down the service as described in the following document: https://rda.ucar.edu/rdadocs/RDA_strategy_to_wind_down_services.pdf . Portions of this plan are still a work in progress. For example, it is estimated that over 80% of RDA dataset collections are either a copy of or are derived from dataset collections that are maintained at other repositories. Examples of these datasets include copies of institutionally produced reanalysis datasets, such as European Center for Medium Range Weather Forecasts (ECMWF, https://www.ecmwf.int/), National Centers for Environmental Prediction (NCEP, http://www.ncep.noaa.gov/), and Japanese Meteorological Agency (JMA, https://www.jma.go.jp/jma/indexe.html). These datasets are archived in the producing institution’s repository in addition to the RDA. Similarly, copies of real-time observational, model analysis, and forecast data are archived in national weather agencies in addition to the RDA. DECS already maintains the dataset source information in the dataset metadata (e.g. point of contact for initial dataset deposit), and is in the process of adding information to the dataset metadata to define authoritatively which datasets have a second copy maintained at an alternate repository. Once this information is complete, the DECS will move forward and create an inventory of datasets detailing custodial information on a yearly basis. For additional details, see the “Maintain relevant dataset background information" section of the following document: https://rda.ucar.edu/rdadocs/RDA_strategy_to_wind_down_services.pdf

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

IV. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The Research Data Archive (RDA) has not hosted human subjects data or data that contains restricted, private, confidential, or otherwise protected information and does not plan to do so in the future. Consequently, many aspects of R4 are not applicable to the RDA.

Currently, data submitters are asked if they had to “go through Institutional Review Board (IRB)/National Center for Atmospheric Research (NCAR) Human Subjects Committee when developing your study?” (see section 2 of 4 in the following page: https://rda.ucar.edu/#!daas/worksheet-instructions). If the answer is yes, the submitter will not be eligible to deposit the proposed dataset in the RDA. Sample data files are always examined as part of the data submission process to verify compliance with data format standards and conventions, and at this point, unexpected parameters with sensitive information could be discovered. If it is determined that the submitter gave an erroneous or a dishonest response to this question, the RDA has the right to reject the archive submission at that point. Overall, the Atmospheric and Oceanographic weather data hosted by the RDA are produced by:
1)Publicly supported and accessible observing systems
2)Weather and climate research models For additional background on climate and weather data, see:
https://www.climate.gov/maps-data/primer/processing-climate-data
In both cases, outputs do not contain “protected” information, and disclosure risks are not an issue.

In terms of curation, the RDA complies with applicable disciplinary norms by leveraging the Open Archival Information System (OAIS) model to architect the archive infrastructure and workflow from dataset ingest to dissemination (https://rda.ucar.edu/rdadocs/RDA_data_ingest_to_dissemination_workflow_overview.pdf , also see R12 and R15 for more details). Additionally, disciplinary accepted metadata and data standards are leveraged by the RDA, and submitters are requested to deposit data in a disciplinary accepted file format and standard (see R14 for details).

All users that access RDA data are required to have either 1) a Computational and Information Systems Lab (CISL) High Performance Computing (HPC) account, from where they can directly access all RDA data from the disk and High Performance Storage System (HPSS) (https://www2.cisl.ucar.edu/data-portals/research-data-archive), or 2) a RDA user account (https://rda.ucar.edu/index.html?hash=data_user&action=register), which can be used to access all data that meets only the Terms of Use for University Corporation for Atmospheric Research (UCAR) Data Repositories (https://rda.ucar.edu/index.html#repository_terms_conditions).

CISL HPC users are granted access to all RDA datasets, as there are no access restrictions in place for CISL HPC users. For additional background on the CISL HPC account application and approval process, see:
1)Allocations: https://www2.cisl.ucar.edu/user-support/allocations
2)User Accounts and Access: https://www2.cisl.ucar.edu/user-support/user-accounts-and-access

3)User Responsibilities: https://www2.cisl.ucar.edu/user-support/user-responsibilities

All users who register for an RDA account must agree to the Terms of Use for UCAR Data Repositories to be considered for registration (See R2 for details). Selected datasets are subject to additional terms of use requirements as stipulated by the original data providers. RDA users must agree to the additional, affiliated terms of use requirements for these datasets and have obtained the corresponding authorizations in their RDA user profiles before they are allowed to access these data via the RDA web services and web interfaces. For additional background, see the following page:
http://ncarrda.blogspot.com/2016/08/how-to-access-restricted-data-set.html

Finally, the RDA follows institutional guidelines and policies to ensure that data are accessed and used in compliance with disciplinary and ethical norms according to the following UCAR policies:
1)Terms of Use for UCAR Data Repositories: https://rda.ucar.edu/index.html#repository_terms_conditions
2)Website terms of use: https://www.ucar.edu/terms-of-use
3)Privacy Policy: https://www.ucar.edu/privacy-policy
4)Copyright Guidelines: https://www.ucar.edu/notification-copyright-infringement-digital-millenium-copyright-act

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

V. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

History of Long-term Institutional Stability
The Research Data Archive (RDA) has been maintained at the National Center for Atmospheric Research (NCAR, https://ncar.ucar.edu/who-we-are/history) since 1965, and fulfills a crucial role in NCAR's core mission as is discussed in R0 and R1. Since its inception in 1965, the RDA has been maintained by what eventually became NCAR's Data Support Section (DSS), founded by Roy Jenne (http://ncarrda.blogspot.com/2017/01/remembering-roy-jenne-1931-2016.html).

Historically, the DSS was organized within the NCAR Computing Facility from 1965 to 1980 (first mention of DSS activities can be found on page 121 of the 1966 NCAR Annual report in the section titled "New Data", https://opensky.ucar.edu/islandora/object/archives%3A7295/datastream/OBJ/view). In 1980, the NCAR Computing Facility was reorganized into the Scientific Computing Division (SCD) and the DSS moved into the SCD as well. The DSS position within the SCD is illustrated in the 1995 org chart (http://www.ncar.ucar.edu/asr/ASR95/SCD/scdorg.html), and 1995 SCD Annual Report (http://www.ncar.ucar.edu/asr/ASR95/SCD/dss.html). In 2005, SCD was reorganized into the Computational and Information Systems Laboratory (CISL), and the DSS moved with it into the Operations and Services Division (OSD, https://www2.cisl.ucar.edu/osd), formed within CISL. In 2018, the divisions within CISL were reorganized, and a new Information Systems Division (ISD) was formed. The Data Support Section moved into the ISD and was renamed to the Data Engineering and Curation Section (DECS) at this time. The current org. charts are included below:
1)Current NCAR org. chart: https://ncar.ucar.edu/who-we-are/org-chart
2)Current CISL org. chart: https://www2.cisl.ucar.edu/dir/cisl-org-chart
3)Current ISD org. chart: https://www2.cisl.ucar.edu/isd-org-chart
4)Current DECS org. chart: https://www2.cisl.ucar.edu/org/cisl/decs

As highlighted in this above discussion, the DECS (formerly DSS) has grown and evolved with many organizational changes over the years to fulfill its core organizational role of maintaining the RDA for the Atmospheric and related sciences research community. Additionally, as discussed in R8, the RDA has a strategy of purposeful growth by constraining the scope of what types of data are allowed to be considered for inclusion in the archive. This has enabled the RDA to be efficient and effective with its allocated resources, including its funding, staff (e.g. hiring and training), and infrastructure (e.g. storage, software development, and databases). Finally, in an effort to remain sustainable, the RDA has a policy in place to purge non-observing based datasets (e.g. model outputs) from the archive after five years if certain requirements are not met as described under the "Dataset Withdrawal Policy" section of the following page: https://rda.ucar.edu/#!daas/terms-and-conditions . The RDA also has cost recovery mechanisms in place to support long term sustainability for specific use cases as described in the "Cost Recovery" section of the following page: https://rda.ucar.edu/#!daas/terms-and-conditions . Based on the RDA's history, defined archive scope, sustainability strategy, and ongoing relevance to NCAR's core mission, it can be expected the DECS will continue to be supported and be able to fulfill this mission in the years ahead.

Funding

NCAR is funded through the National Science Foundation (NSF) and other partners (https://ncar.ucar.edu/who-we-are/funding). NSF core funding is provided through a cooperative agreement that is renewed in 5 yearly cycles (https://nsf.gov/funding/pgm_summ.jsp?org=GEO&pims_id=12809). An article highlighting details of the recently approved cooperative agreement can be found here: https://news.ucar.edu/132627/ucar-nsf-sign-agreement-ncar-management . This agreement took effect on October 1, 2018, and will be effective until September 30, 2023, with the possibility of a 5 year extension if approved by the NSF's National Science Board (https://www.nsf.gov/nsb/). NSF monitors NCAR's activities through a number of reporting mechanisms, including the NCAR annual reports (https://nar.ucar.edu/), and through the NSF site visit reviews (https://www.nsf.gov/bfa/dias/caar/sitevisits.jsp). NCAR receives feedback from these reviews, and may make strategic adjustments based on this feedback in order to continue to receive funding successfully through future cooperative agreement renewals. The most recent NSF site visit review of NCAR occured in 2016. A summary of the findings can is provided in the 2016 NCAR directors message of the annual report: https://nar.ucar.edu/2016/ncar/message-ncar-director . DECS' participation in the feedback and strategic adjustments is discussed in further detail in R6.

Staffing, Staff Expertise, and Professional Development

Staffing levels in the DECS have grown from 1 at its inception in 1965, to the current staffing of 8 members in total: 6 full time software engineers, 1 support staff, and 1 manager as illustrated in the DECS organizational chart that is linked above. NCAR and CISL have consistently supported the staffing level required to maintain the RDA from 1965 to the present through NSF base funds, as the RDA provides a core resource for NCAR and the broader Atmospheric and related research communities. Consequently, proposal/project-specific "soft" funds are not needed to support RDA data curation related activities.

Additionally, an adequate travel budget is provided to DECS to support recurring attendance to meetings and trainings as highlighted in R15, where staff remain current on community data management, technical, and research discipline practices. Finally, NCAR and University Corporation for Atmospheric Research (UCAR) provide additional opportunities for training and professional development through its Employee & Organizational Development (EOD, https://eod.ucar.edu/), and Education Assistance (https://www2.fin.ucar.edu/hr/benefits/education-assistance-program) programs. For example, DECS staff have participated in "Data Carpentry" instructor training workshops and various software training (e.g. python and Fortran) courses provided through the EOD program.

The current set of software engineers and manager (7 Full Time Equivalents (FTEs)) that maintain the RDA have undergraduate degrees in Meteorology, Math, Physics, Engineering, Computer Science, and Oceanography, and Advanced degrees in Atmospheric and Oceanographic sciences. Each staff member fulfills a specialized role in developing and maintaining the RDA software infrastructure described in R15. In particular, staff members' current software expertise includes Fortran, C++, PERL, Python, Java, web related languages and structures, domain knowledge of scientific data formats and metadata schemas, and advanced knowledge of Unix/Linux based operating systems. Educational background in the Atmospheric and Oceanographic sciences provides the foundation for the "Dataset Specialist" role highlighted in the following document:

https://rda.ucar.edu/rdadocs/RDA_data_ingest_to_dissemination_workflow_overview.pdf , and each staff member can fulfill the responsibilities to curate RDA dataset collections. Finally, all staff engage in a yearly performance appraisal where they are recognized for advancements and accomplishments in data curation and Information Technology (IT) related skill sets and activities, and are coached on areas that may need growth.

New staff are provided with a standard set of onboarding documents and exercises, where they learn: 1) to use the DECS data curation tools, 2) about the controlled vocabularies and metadata standards used to support the RDA, 3) to maintain an existing dataset, 4) to create a new dataset from ingest to dissemination, and 5) about the expectations and processes to provide user support. This onboarded process is structured with the goal of integrating new staff into the DECS team in a consistent manner.

IT Resources
Historically positioned in the NCAR Computing Facility, SCD, and now CISL, DECS has been provided with a wealth of IT resources to support the RDA infrastructure. An overview of CISL IT resources is provided here: https://www2.cisl.ucar.edu/resources/resources-overview , and additional background on RDA specific IT software and hardware infrastructure is provided in R15.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

VI. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse or external, including scientific guidance, if relevant).

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The Data Engineering & Curation Section (DECS) strives to keep the Research Data Archive's (RDA) holdings and user services, including technical and curatorial, relevant to support research needs in the Atmospheric and related research communities. To do this, DECS staff use both internal and external mechanisms to secure ongoing expert guidance and feedback on existing and future RDA holdings and services. These mechanisms include:

Internal
1)DECS management meets on a bi-monthly basis with Computational and Information Systems Lab (CISL), Information Systems Division (ISD) management to review and adjust current and future activities. These meetings include the ISD director and section managers who have relevant disciplinary, technical, and policy backgrounds (see the ISD org. chart for additional details: https://www2.cisl.ucar.edu/isd-org-chart). These meetings focus on ISD specific activities and any adjustments in technology roadmaps in the context of the broader ISD activities portfolio, which can have direct impact on the RDA's ongoing and planned activities.

2)DECS management meets with all CISL lab, division, and section managers on a monthly basis to remain aware of all current National Center for Atmospheric Research (NCAR) relevant activities, CISL specific activities, and any adjustments in technology roadmaps in the context of the broader CISL activities portfolio, which can have direct impact on the RDA's long term activities. The CISL section managers meeting includes the CISL Lab Director, all CISL Division Directors, and all section managers within each CISL division (see CISL org chart for more details: https://www2.cisl.ucar.edu/dir/cisl-org-chart).

3)DECS management organizes meetings with stakeholders from each NCAR lab (https://ncar.ucar.edu/who-we-are/org-chart) on a bi-annual basis to ensure that the RDA holdings and services are meeting their science research needs, and to seek feedback on future needs for holdings and services to support continued research. Additionally, DECS management advises stakeholders from the NCAR labs on any changes to policies or procedures that may impact their research workflows. Stakeholders typically include representatives from the lab directors office, lab administrators, and a selected group of lab Principal Investigators.

4)DECS contributes to the NCAR Data Stewardship Engineering Team (DSET) activities. DSET was created by the NCAR directorate to lead the organization's efforts to provide enhanced, comprehensive digital data discovery and access through cross organizational collaborations (see: https://internal-ncar.ucar.edu/data-stewardship-engineering-team-dset). DECS management participates in DSET's technical, curatorial, and policy discussions, and aligns RDA activities with the broader organizational directions that are set by the DSET as needed.

External
1)DECS management participates in CISL Advisory Panel (CAP,

https://www2.cisl.ucar.edu/internal/cisl-advisory-committees) review on an annual basis. During these reviews, DECS management 1) provides updates on current DECS activities, any new activities that took place to accommodate CAP suggestions from the prior year's review, and the planned future activities, and 2) answers any CAP questions. The CAP typically provides feedback with suggestions for possible near-term changes in RDA holdings and services, and a broader perspective on the RDA's long-term strategic roadmap in the context of the broader CISL/ISD's data holdings and services portfolio. The CAP is composed of experts in Atmospheric, Climate, and Ocean research, and experts in high performance computing (HPC) and data services related technologies and strategies.

2)National Center for Atmospheric Research (NCAR) management, including CISL representatives, participates in the NCAR External Advisory Panel (NAP, https://ncar.ucar.edu/who-we-are/leadership/advisory-committees/ncar-advisory-panel) review on an annual basis. During these reviews, CISL management provides updates on current CISL activities, any new activities that took place to accommodate NAP suggestions from the prior year's review, and the planned future activities. The NAP typically provides feedback with suggestions for possible near-term changes in CISL services, and a broader perspective on CISL's long-term strategic roadmap in the context of the broader NCAR data services portfolio. ISD/DECS activities are typically highlighted as part of the broader CISL update. Any strategic feedback from the NAP that directly involves DECS and the RDA is communicated to DECS management by CISL management. DECS management will occasionally participate directly in the NAP review if called upon to highlight RDA related developments and answer NAP questions. The NAP is composed of leading experts in Atmospheric, Climate, Ocean, and Solar research, who have broad experience with utilizing NCAR's many services, including HPC and data services.

3)NCAR management, including CISL representatives, participates in National Science Foundation (NSF) Site Visit Team (SVT) reviews on a five-yearly basis as part of the cooperative agreement to fund NCAR. The SVT focuses on areas such as computing/data services and observing facilities. The SVT reviews broadly investigate whether the services and facilities provided are meeting NSF research community expectations and needs. Existing service portfolios are reviewed by the SVT, and suggestions for improvement may be provided to guide future strategic roadmaps. CISL/ISD/DECS activities are typically highlighted as part of the broader NCAR data services update, and any strategic feedback from the SVT that directly involves DECS and the RDA is communicated to DECS management by CISL management. DECS management will occasionally participate directly in the SVT review if called upon to highlight RDA related developments and answer SVT questions. SVT members have expertise in Atmospheric, Climate, Ocean, and Solar research, and in Computational and Data Sciences. SVT members come from the NSF university community and other institutions that could be considered peers of NCAR such as National Aeronautics and Space Administration and Department of Energy labs.

4)NCAR management, including CISL representatives, participates in the University Corporation for Atmospheric Research (UCAR) annual members meeting (https://www.ucar.edu/who-we-are/membership-governance/member-institutions/annual-members-meetings/2018-members-meeting). During these meetings, NCAR management provides updates on current NCAR activities, which include CISL activities and planned future activities. Additionally, CISL management, and at times DECS staff, participates in breakout sessions, which focus on relevant topics such as data and HPC. The combination of formal presentations and breakout session engagement provides an opportunity for CISL management to interact with representatives from UCAR member universities and receive constructive feedback regarding CISL services and data holdings. Any strategic feedback from the UCAR members meeting that directly involves DECS and the RDA is communicated to DECS management by CISL management. The UCAR members meeting is composed of leading experts from UCAR member Universities (https://www.ucar.edu/who-we-are/membership-governance/member-institutions) in Atmospheric, Climate, Ocean, and Solar research, who have broad experience utilizing NCAR's many services including HPC and data services.

The DECS works through a variety of additional communication mechanisms to engage with both its internal NCAR community and external stakeholders. Below is the summary of each communication mechanism:
1)rdahelp email: The rdahelp email provides a generic point of contact for all types of stakeholders to engage RDA staff with data or technology related questions as well as provide suggestions for improvements on RDA holdings and services.

2)Dataset specialist email/phone: A dataset specialist is assigned to each RDA dataset, and their contact information is provided near the top right corner of the dataset home page. The DECS's full directory can also be found under the RDA's "About/Contact" page (https://rda.ucar.edu/#!about). Direct contact information allows the stakeholders to engage with the DECS staff member regarding dataset specific related questions, and provide suggestions for improvements on the RDA's overall holdings and specific services for that dataset.

3)rda-users email: The rda-users email is a subscription email service that informs subscribers with updates about RDA data holdings and services. Subscribers can also contact the full email subscription list with data or technology related questions, and also provide suggestions for improvements on RDA holdings and services.

4)Social media: The DECS maintains a RDA blog site as well as Twitter and Facebook accounts to inform subscribers with updates about RDA data holdings and services. Subscribers can send direct messages via any of these forums with data or technology related questions, and provide suggestions for improvements on archive holdings and services. It should be noted that consistent messages are presented across all social media platforms, so a user only needs to subscribe to one platform in order to remain informed on RDA activities.

5)CISL user surveys executed by the CISL User Services Section (USS): The CISL USS asks the CISL HPC community to complete a survey on a three-yearly basis to inform CISL on what its community likes and dislikes about, and would like to request from the CISL service portfolio for supporting their research needs. Questions regarding RDA holdings and services are included in this survey, and the DECS receives feedback accordingly from these surveys.

6)Engagement at professional meetings: DECS staff regularly present on RDA holdings and services at domestic and international meetings and workshops, such as those hosted by disciplinary or professional societies, to provide outreach to the designated user communities. Selected examples of these meetings and workshops that the DECS participates in include but are not limited to the European Geophysical Union (EGU), American Geophysical Union (AGU), and American Meteorological Society (AMS) annual meetings. DECS also participates in data and technology centric workshops and meetings to remain current of community driven policies, technology trends, and best practices related to data curation. Further, DECS can seek out opportunities for collaboration at these data and technology centric workshop and meetings.

Selected examples of data and technology centric meetings and workshops that DECS participates in include the Earth Science Information Partners (ESIP) bi-annual, Percona Live Database annual, Globus annual, and Research Data Alliance plenary meetings. Finally, DECS staff frequently volunteer to work at the NCAR exhibit booth during meetings where the NCAR booth is present. All of these opportunities, including the venues of presentation, participation in breakout sessions, and being available for questions while working at the booth, provide valuable channels for community engagement and feedback that is leveraged by the DECS to inform future directions for RDA holdings and services.

7)Day-to-Day in person interactions: DECS staff informally engage other NCAR staff in hallway or drop-in office discussions where constructive feedback is received on RDA holdings and services. This feedback is used to inform the DECS on the adjustments the may need to be made to the RDA in the near and long term.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

DIGITAL OBJECT MANAGEMENT

VII. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

One of the core foundational principles of the Data Engineering & Curation Section (DECS) is to guarantee the integrity and authenticity of the data archived in the Research Data Archive (RDA), and to ensure this, DECS employs the following strategies.

As described in the “Research Data Archive Dataset Ingest to Dissemination Workflow Overview” document (https://rda.ucar.edu/rdadocs/RDA_data_ingest_to_dissemination_workflow_overview.pdf) and under the “Upon acceptance” section of the following page: https://rda.ucar.edu/#!daas/decision-workflow , the DECS maintains a structured process to ensure data integrity and authenticity from ingest to dissemination. Specifically:
1)The data submitter is asked to host the data files to be transferred on a remote server and provide the DECS dataset specialist with a manifest that includes the complete list of files to be transferred and the MD5 checksum for each file. As there can be certain nuances depending on the type of systems and transfer protocol used, the data submitter will work with the DECS dataset specialist to determine the best structure for the manifest file, and the appropriate method to compute the MD5 checksums for that specific use case. The checksums are recomputed and compared for consistency once the files have been transferred from the submitters’ remote server to local DECS storage. File size consistency is also verified as a second check. Additional details are provided under the “Upon Acceptance” section’s “Submission of the actual data files” subsection of the following page: https://rda.ucar.edu/#!daas/decision-workflow
2)As discussed under the “Upon Acceptance” section’s “Collaboration, verification, and confirmation of metadata record…” sub-section on the following page: https://rda.ucar.edu/#!daas/decision-workflow , the responsible DECS dataset specialist will use the information provided in the dataset submission form to seed the initial dataset metadata. The DECS dataset specialist will then work with the data submitter to complete all required metadata fields. The protocols in place to ensure data completeness are provided under the “Upon acceptance” section’s “Submission of the actual data files” subsection of the following page: https://rda.ucar.edu/#!daas/decision-workflow . Please also see R8 and R14 for details on data and metadata requirements, including required metadata fields, for supporting data and metadata completeness.
3)Changes to metadata and the data files themselves are logged through different mechanisms. The mechanisms are structured to track any and all changes to data files and dataset metadata, as well as to support data versioning. Additionally, metadata are made available in community accepted schemas, and the data files must be structured according to community accepted standards. Details related to these topics are provided below.

Data Files:
1)If a curation level greater than Basic curation is agreed upon between the data submitter and the RDA (for the definition of “Basic Curation”, see: https://rda.ucar.edu/rdadocs/RDA_Dataset_Curation_Level.pdf), the DECS dataset specialist may make agreed upon changes to the data file structure or content. In this situation, the DECS dataset specialist is authenticated and authorized by the RDA system before any changes are made. The workflow used to create the derived products is documented and maintained as part of the dataset metadata. This workflow information is also made publicly available under the “Documentation” tab of the dataset’s landing page (for an example of the documented workflow, see: https://rda.ucar.edu/datasets/ds630.0/docs/CISL-RDA-ERA5.grib_to_netCDF4_HDF5.jpg). A reference copy of the native, unchanged data file(s) is(are) always preserved with the dataset to assure reproducibility and validation of any derived product or restructuring workflows. These files are tracked as their own fileset group within the dataset, stored only on tape, and are only accessible through offline request. The native, unchanged data files receive all of the same level of

preservation support as all other files in the dataset according to the description found under “Data Preservation Policy” section of the following page: https://rda.ucar.edu/#!daas/terms-and-conditions
2)Once a dataset has been created and a Digital Object Identifier (DOI) has been assigned to a dataset, there is a variety of strategies in place to track version changes to the data files as detailed under the following page: https://rda.ucar.edu/#!data-citation/use-cases
3)The data files included in a dataset must be provided in a community supported data format as outlined under the “Section 2 of 4: Dataset Characteristics” section’s “File Format(s)” subsection on the following page: https://rda.ucar.edu/#!daas/worksheet-instructions . This is required to support access from community developed tools and support long-term, sustainable data curation.
4)There are no circumstances where data files will be deleted from a dataset, unless the dataset is purged according to the “Dataset Withdrawal Policy” found on the following page: https://rda.ucar.edu/#!daas/terms-and-conditions

Metadata:
1)All changes to data description metadata are tracked in a Concurrent Versions System (CVS) version control system (https://www.gnu.org/software/trans-coord/manual/cvs/cvs.html). The full history of metadata changes can be accessed under the “Change History” section of the Metadata Manager, i.e. RDA’s web-based tool for performing various metadata-related activities. An overview of the metadata “Change History” capability is provided under the “Manage Datasets” section of the following page: https://rda.ucar.edu/#!rdadocs/mm_guide
2)Metadata are maintained in a native RDA schema, which can be mapped to a variety of community supported standards as detailed in R14.

The RDA strategy for data changes includes two use cases:
1)Case 1: A curation level greater than Basic curation is agreed upon between the data submitter and the RDA (for the definition of “Basic Curation”, see: https://rda.ucar.edu/rdadocs/RDA_Dataset_Curation_Level.pdf) - In this case, the DECS dataset specialist may make agreed upon changes to the data file structure or content. The DECS dataset specialist is authenticated and authorized by the RDA system before any changes are made. The workflow used to create the derived products is documented and maintained as part of the dataset metadata. This workflow information is also made publicly available under the “Documentation” tab of the dataset’s landing page (for an example of the documented workflow, see: https://rda.ucar.edu/datasets/ds630.0/docs/CISL-RDA-ERA5.grib_to_netCDF4_HDF5.jpg). A reference copy of the native, unchanged data file(s) is(are) always preserved with the dataset to assure reproducibility and validation of any derived product or restructuring workflows. These files are tracked as their own fileset group within the dataset, stored only on tape, and are only accessible through offline request. The native, unchanged data files receive all of the same level of preservation support as all other files in the dataset according to the description found under the “Data Preservation Policy” section on the following page: https://rda.ucar.edu/#!daas/terms-and-conditions . The data submitter is made aware of this strategy under the “Upon acceptance” section’s “Collaboration on data curation level and any related data transformations or restructuring” subsection on the following page: https://rda.ucar.edu/#!daas/decision-workflow

2)Case 2: Updates are made to data files after the dataset has been created and a DOI has been assigned to the dataset - Descriptions of these use cases are provided on the following page: https://rda.ucar.edu/#!data-citation/use-cases . This information can also be made available for the data submitter if an applicable use case arises.

The RDA maintains provenance information to support the following use cases:
1)Case 1: A curation level greater than Basic curation is agreed upon between the data submitter and the RDA (for the definition of “Basic Curation”, see: https://rda.ucar.edu/rdadocs/RDA_Dataset_Curation_Level.pdf) - In this case, the DECS dataset specialist may make agreed upon changes to the data file structure or content. The DECS dataset specialist is authenticated and authorized by the RDA system before any changes are made. The workflow used to create the derived products is documented and maintained as part of the dataset metadata. This workflow information is also made publicly available under the “Documentation” tab of the dataset’s landing page (for an example of the documented workflow, see: https://rda.ucar.edu/datasets/ds630.0/docs/CISL-RDA-ERA5.grib_to_netCDF4_HDF5.jpg). A reference copy of the native, unchanged data file(s) is(are) always preserved with the dataset to assure reproducibility and validation of any derived product or restructuring workflows. These files are tracked as their own fileset group within the dataset, stored only on tape, and are only accessible through offline request. The native, unchanged data files receive all of the same level of preservation support as all other files in the dataset according to the description found under the “Data Preservation Policy” section on the following page: https://rda.ucar.edu/#!daas/terms-and-conditions

2)Case 2: User needs to retrieve a data file that has been replaced in a dataset versioning operation - Data access history is available for every RDA user. By using the Data Citation tool, found in the RDA user dashboard (see: http://ncarrda.blogspot.com/2017/03/the-rda-user-dashboard.html), a user can find the exact date/time that a data file was downloaded. If there have been any changes to the file since that download occured, it will be noted in the dataset file list, and the DECS dataset specialist can provide a copy of the file that was originally downloaded.

3)Case 3: User specified request provenance tracking - A receipt is provided with each user request detailing request specifications. Additionally, a unique persistent identifier is assigned to each user request, so all aspects of that request can be tracked through this mechanism, including request timestamp, request processing steps and platform, operating system version, and software versions used to support request processing. Currently, a data user needs to ask the DECS dataset specialist to provide all of these details based on the request ID, but it is in the future DECS roadmap to add a programmatic capability for RDA users to query this information by request ID.

The RDA maintains links to related RDA datasets, and the metadata associated with those datasets in its native RDA dataset metadata under the “Related RDA Datasets” section of the dataset’s landing page (examples of related datasets can be found using this sample dataset: https://rda.ucar.edu/datasets/ds131.2/#!description). Additionally, this section is mapped to other capable community metadata standards such as ISO 19115-3 under the “AssociatedResource” metadata element. Tools to support dataset relationships by DOI are also now available, but this information still needs to be populated by the DECS dataset specialists on relevant dataset collections.

To support file version changes, the RDA compares the MD5 checksum of a new file with the MD5 checksum of the replaced file. The RDA also scans the file content metadata (see: the “About Data File Content Metadata” section on the following page: https://rda.ucar.edu/#!rdadocs/dsmaint) of the new file to determine if there are any differences in the file contents. New metadata information is integrated as needed.

The repository verifies the identity of submitters by requiring them to register for a RDA user account in order to submit a request to archive data (see: https://rda.ucar.edu/index.html?hash=data_user&action=register). As part of the registration process, the potential registrant is sent a verification email at the provided email address to confirm their registration. Additionally, because data submitters must engage a dataset specialist to submit their data to the RDA, there is follow-on interpersonal communication that occurs in-person, via email, phone, or video-chat that can also be used to confirm the identity of the data submitter.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

VIII. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The Data Engineering & Curation Section (DECS) follows a structured process to appraise whether submitted “requests to archive datasets” are appropriate for inclusion in the Research Data Archive (RDA) as described in the “Terms and Conditions of Request Submission and Archive Data to the RDA” section (https://rda.ucar.edu/#!daas/terms-and-conditions) of the

RDA Data Submission System (https://rda.ucar.edu/#!daas). Specifically, the “Motivation” and “Scope of RDA Dataset Collection” sections highlight the general collection development principles that datasets must support “climate and weather research”, which is aligned with the DECS mission statement found on the following page: https://rda.ucar.edu/#!about

Once a dataset is approved by the DECS manager for inclusion in the RDA according to the “Appraisal and Selection Process” (https://rda.ucar.edu/#!daas/decision-workflow), a dataset specialist (DS) is assigned (based on relevant disciplinary background and workload) to work with the data submitter iteratively in order to work through the process of validating and confirming the following:
1)Adherence to accepted data formats and conventions
2)Data quality
3)Required dataset collection metadata as described in the “Upon acceptance” section found under the following page: https://rda.ucar.edu/#!daas/decision-workflow , and as highlighted in Figures 1 and 2 in the following file: https://rda.ucar.edu/rdadocs/RDA_data_ingest_to_dissemination_workflow_overview.pdf

The DS must enter a minimum set of required metadata fields as highlighted in the “Metadata fields” section of the following page: https://rda.ucar.edu/#!rdadocs/mm_guide . The minimum set of required metadata fields has been selected to reflect and be mappable to metadata schemas that are commonly recognized and supported by the RDA's scientific community. By doing so, the RDA is well positioned to support long-term preservation.

The dataset collection metadata is initially seeded with the general information provided in the dataset submission system (https://rda.ucar.edu/#!daas/worksheet-instructions). The general information provided through the dataset submission system are typically insufficient to provide a complete metadata record; the DS will work with the data submitter until the metadata record is deemed sufficient according to our metadata requirements. For additional background, comprehensive descriptions of RDA dataset metadata requirements, including quality assessment and control, and curation strategies, are included in R14, R4, and R9.

A list of accepted data formats (see R14 for background on accepted data formats) for submission to the RDA is provided under the “File Formats” bullet in section 2 of the following page: https://rda.ucar.edu/#!daas/worksheet-instructions . The DS validates whether sample files provided by the data submitter adhere to one of the accepted data formats (and associated conventions where applicable, such as Climate and Forecasts (CF) http://cfconventions.org/) by scanning the files with the RDA gatherxml tool (https://rda.ucar.edu/#!rdaman/gatherxml). If the format adheres to community specifications and conventions, gatherxml will successfully extract metadata from the sample files. If not, gatherxml will produce an error report, and the DS will iteratively work with the data submitter to fix the file format issues before the data submitter can provide the full dataset for archival. In some cases, third party tools such as the United Kingdom Hadley Center supported “CF-checker” (http://puma.nerc.ac.uk/cgi-bin/cf-checker.pl) may also be applied as a second check to validate adherence to the NetCDF CF-conventions on sample data files.

In addition to validating adherence to data format and convention, the DS will run checks on the data files by plotting sample fields (see Figure 2.2 in the following document: https://rda.ucar.edu/rdadocs/RDA_data_ingest_to_dissemination_workflow_overview.pdf) to make sure the data values are physically reasonable. If issues are discovered with data values, the DS will iteratively work with the data submitter to fix the data issues before the data submitter can provide the full dataset for archival.

Once sample files pass the gatherxml (and CF-checker tests where applicable) and data value checks, the full dataset collection will be transferred from the data submitter to local RDA storage using standard web protocols, then archived as highlighted in Figure 2 of the following document: https://rda.ucar.edu/rdadocs/RDA_data_ingest_to_dissemination_workflow_overview.pdf . As part of the archival process, all files are scanned by the gatherxml software. Issues with adherence to expected data format and community conventions will be discovered when the data files are scanned by gatherxml. This ensures that all files in a dataset collection adhere to expected formats and conventions. Additionally, gatherxml validates file completeness to ensure that truncated or corrupted files are not included in the dataset archive. If any of the above issues are detected in files during the gatherxml scanning process, the data submitter will be notified by the DS, asked to fix the specified issues, and make the corrected files available to be transferred to the RDA through the previously agreed to transfer mechanism.

The RDA relies on the end user community to pass along additional data issues not discovered through this initial vetting process through direct communication with the responsible DS, or through the general rdahelp@ucar.edu email list. This is a very rare occurrence, but a valuable mechanism for feedback when a data issue is not detected by the standard RDA quality control procedures.

If a data submitter wants to archive a non-preferred data format (e.g. a unique ASCII format), sufficient descriptive documentation must be provided that describes the format specification, and it is preferred that example source code to read the data is also provided. The required documentation and source code will be archived with the dataset collection package in to support long-term preservation and re-use.

One a dataset has been successfully created, the DS will engage the data submitter as needed to make adjustments to metadata, and to ensure that metadata remain current over time. Additionally, per the terms and conditions to deposit with the RDA, the data submitter is asked to contact the responsible DS to provide updates to metadata and documentation as needed.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

IX. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The Research Data Archive's (RDA) Information Technology (IT) infrastructure, supported by the National Center for Atmospheric Research (NCAR) Computational and Information Systems Lab (CISL), provides highly available storage, backup, and disaster recovery for archive data. Workflows exist for each step of the archival storage process. Data that are stored on disk and High Performance Storage Systems (HPSS) are only directly accessible within NCAR, and write permissions for the archive storage system are only granted to the few authorized Data Engineering & Curation Section (DECS) staff who are directly involved with maintaining the RDA. The RDA's data and metadata files undergo systematic back-ups and integrity checks. Dual copies of all dataset metadata components and of each RDA data file are stored in separate physical locations. These include the NCAR Wyoming Supercomputing Center (NWSC, https://nwsc.ucar.edu/) located in Cheyenne, WY, and the NCAR Mesa Lab (see MESA LAB & FLEISCHMANN BUILDING on https://www.ucar.edu/who-we-are/contact-us), located in Boulder, CO. In the event of data corruption, data can be restored from one of the physically separated backups.

Full documentation of the ingest to dissemination workflow can be found in the following document:
https://rda.ucar.edu/rdadocs/RDA_data_ingest_to_dissemination_workflow_overview.pdf

Software components used to manage archival storage of RDA datasets:
1)dsarch - Data management tool. Used by all RDA dataset specialists to archive data files programmatically to NWSC-HPSS and ML-HPSS (see: https://rda.ucar.edu/rdadocs/dsarch/).
2)Metadata Utilities -Extract metadata from files upon data ingest, publish metadata information to web directories, insert metadata information into corresponding databases, make publicly available metadata information via web-based dataset-specific user interfaces (see: https://rda.ucar.edu/#!rdaman).
2)Dataset summary metadata are entered by RDA dataset specialists through the Metadata Manager application (https://rda.ucar.edu/#!rdadocs/mm_guide). This information is stored in a native RDA metadata format. These metadata

are used to populate description pages for each dataset and are indexed to support the RDA dataset search and “Look For Data” browse function (https://rda.ucar.edu/#!lfd?nb=y). Extensible tools are available to crosswalk the metadata information from the native metadata format into standards compliant structures including ISO (19139 and 19115-3), GCMD DIF, FGDC, and Datacite. Metadata information in these standards can all be accessed through the RDA’s user facing web interface or via an OAI-PMH web service (https://rda.ucar.edu/cgi-bin/oai).

Security is considered in terms of data preservation security. No sensitive information is hosted in the RDA. A description of required levels of security security and how they are supported can be found in the Data Security section of the following document: https://rda.ucar.edu/rdadocs/RDA_data_security.pdf

A description of how data storage is addressed by the preservation policy is provided in the Data Preservation section of the following document: https://rda.ucar.edu/rdadocs/RDA_data_security.pdf

The repository has a strategy for backup/multiple copies. Two copies of all dataset metadata components (RDA Web Server disk and RDA Metadata Databases), and every data file registered in the RDA are maintained on physically separated HPSS systems. The primary copy is housed at NWSC-HPSS, and the back-up copy is housed at ML-HPSS. For additional details, see the Data Preservation Section of the following document: https://rda.ucar.edu/rdadocs/RDA_data_security.pdf

Data recovery provisions are in place. For full details, see the Disaster Recovery section of the following document: https://rda.ucar.edu/rdadocs/RDA_data_security.pdf

Risk management techniques are used to inform the strategies employed by the RDA. For full details, see the Risk Management and Resiliency section of the following document: https://rda.ucar.edu/rdadocs/RDA_data_security.pdf

Checks are in place to ensure consistency across archival copies. For full details, see the Data Security Section of the following document: https://rda.ucar.edu/rdadocs/RDA_data_security.pdf

For a description of how deterioration of storage media is handled, see the Maintaining Storage Media Section of the following document: https://rda.ucar.edu/rdadocs/RDA_data_security.pdf

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

X. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The Data Engineering & Curation Section (DECS) has developed documented procedures and terms of reference to specify the roles, responsibilities, and expectations of the Research Data Archive (RDA) designated user community, potential data submitters, and the RDA itself. Details related to these various aspects are provided below and according to the response guidance questions for R10. Although only officially formalized in 2014, these procedures and terms of reference have been in place since 2008 and have been iterated over the last ten years. Going forward, the official version of the procedures and terms of reference will be reviewed on a bi-yearly basis.

The RDA provides a preservation plan as described in the “Rights / Terms, Conditions for Use, collaboration, and Ownership” section of the following page: https://rda.ucar.edu/#!daas/terms-and-conditions . All data submitters must agree to the language found on this page. Particularly, when the data submitters have completed and are submitting their forms to request data archival with the RDA (see the Request Submission” section on this page: https://rda.ucar.edu/#!daas/worksheet-instructions), potential data submitters are asked to verify that they agree to the “Terms and Conditions of Request Submission and Archival of Data to the RDA” before being allowed to proceed to the next steps. Completing this verification enables the potential data submitters to acknowledge and agree to all actions necessary for meeting the preservation responsibilities.

All datasets archived in the RDA receive consistent digital preservation and curation support. A description of the digital preservation and support characteristics provided to all RDA datasets can be found in the “Data Preservation Policy” section of the following page: https://rda.ucar.edu/#!daas/terms-and-conditions

The process for transferring custody and responsibility from the data submitter to the RDA is detailed in the following page: https://rda.ucar.edu/#!daas/decision-workflow

As outlined in the “Terms and Conditions of Request Submission and Archival of Data to the RDA” (https://rda.ucar.edu/#!daas/terms-and-conditions), the RDA has the rights to copy, transform, and store data items as well as provide access to them.

All actions relevant to preservation including custody transfer, submission information standards, and archival information standards are described under the “Upon Acceptance” section of the following page: https://rda.ucar.edu/#!daas/decision-workflow . Additional details are provided in the following two documents:
1)https://rda.ucar.edu/rdadocs/RDA_data_ingest_to_dissemination_workflow_overview.pdf
2)https://rda.ucar.edu/rdadocs/RDA_data_security.pdf

The RDA relies on the assigned DECS dataset specialist to ensure that all required actions are completed successfully to create, ingest, and archive a new dataset into the RDA. No unrecoverable instance of data loss have been detected in RDA since its inception in 1965.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XI. Data quality

R11. 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.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The Data Engineering & Curation Section (DECS) works directly with data submitters who are approved to deposit their data in the Research Data Archive (RDA) to:
1)Ensure that dataset collection metadata is complete and accurate (including all required fields as described in R14), and
2)Document any quality issues that may exist in the data itself.

Once a request to archive data has been approved (see: https://rda.ucar.edu/#!daas/decision-workflow), a DECS Dataset Specialist (DS) is assigned to work with an approved data submitter, as described in the "Research Data Archive Dataset Ingest to Dissemination Workflow Overview" document (https://rda.ucar.edu/rdadocs/RDA_data_ingest_to_dissemination_workflow_overview.pdf).

Dataset collection metadata completeness and data quality are two areas of focus that the DS works with the data submitter to evaluate. This iterative and collaborative approach used to evaluate the quality of data and metadata, and to assess the data and metadata's adherence to relevant schema employed by the RDA is well described in the "Upon acceptance" section of the following page: https://rda.ucar.edu/#!daas/decision-workflow , and the responses to R7, R8 and R14. The information included below complements the information presented in R7, R8, and R14.

There is no interactive mechanism for RDA users to comment directly on or rate dataset data and metadata through the RDA web interface (a comment box was provided on dataset homepages in the past, but was removed due to no community use). However, there are a number of other mechanisms available for RDA users to provide data or metadata feedback as described in R6. Additionally, whenever a DS performs edits on dataset metadata (for example to add a related website or publication reference, which can happen on a yearly basis), the DS is notified by the Metadata Manager utility (https://rda.ucar.edu/#!rdadocs/mm_guide) if there are any spelling errors or broken links in the metadata that need to be fixed. Furthermore, all metadata records are submitted for inclusion in the National Center for Atmospheric Research (NCAR) Digital Asset Services Hub (DASH, https://www2.cisl.ucar.edu/dash/search) Search system. The metadata records are vetted again for completeness and consistency through the DASH Search process.

Regarding external vetting of data quality, a RDA dataset collection may include links to dataset collection commentary provided by disciplinary experts through NCAR's Climate Data Guide (CDG, https://climatedataguide.ucar.edu/). If CDG resources, such as CDG "Dataset Assessment" and "Expert Guidance", are available for a RDA dataset collection, links to the resources are provided on the RDA dataset homepage. The assessment and guidance provided in the CDG discussion address dataset collection strengths, weaknesses, progress relative to prior versions of a dataset collection, and known quality issues with a dataset collection. These are valuable and trustworthy resources for users to assess whether a RDA dataset collection may be suitable for their research use cases. For an example of the "Dataset Assessment" and "Expert Guidance" links for an RDA dataset, see the "NCAR Climate Data Guide" section on the "ERA-Interim Project" dataset collection homepage: https://rda.ucar.edu/datasets/ds627.0/#!description

In addition to the CDG resources, there are several sections on a dataset collection homepage that provide linkage to

related works and guidance on how to cite the dataset. The sections include:
1)Related RDA Datasets: Includes web links to other RDA dataset collections, which have data that may be complementary or are directly related to the dataset that is currently being viewed. RDA users can find alternative dataset collection resources through these links.
2)Related Resources: Includes web links to external (non RDA) web pages that provide supporting information or alternate access points for the dataset collection being viewed. If a RDA user wants additional information about a dataset collection that is hosted outside of the RDA, this section provides the additional resources for related information.
3)Publications: Includes citations to published works that are related to the dataset being viewed. These are typically provided by the data submitter and include description information on dataset characteristics, which can in turn be valuable in assisting RDA users in determining the dataset's applicability to their research use cases.
4)Data Citations: Includes citations of published works that have cited the dataset being viewed. This can provide valuable insight into how others have used a dataset to support various types of research. Data citations are harvested from Crossref (https://www.crossref.org/) according to the specifications outlined in the Make Data Count: COUNTER Code of Practice for Research Data (https://makedatacount.org/counter-code-of-practice-for-research-data/).
5)How to Cite This Dataset: Provides a template to cite the dataset being viewed in the following bibliographic citation styles:
a)Federation of Earth Science Information Partners (ESIP)
b)Geoscience Data Journal
c)American Geophysical Union (AGU)
d)American Meteorological Society (AMS)
e)DataCite (DC)

Examples of all of these sections can be found on the “ERA-Interim Project” dataset collection homepage: https://rda.ucar.edu/datasets/ds627.0/#!description

When possible, all of the the sections described above are mapped into standard metadata schemas, such as ISO 19115, to support the presentation of this information through federated search and discovery systems, including the NCAR Digital Asset Services Hub (DASH) Search system: https://data.ucar.edu/

When a user has questions or comments on a dataset collection, they can contact the DS whose contact information is listed near the top right corner of a dataset homepage, or they can contact the RDA via the team email address (rdahelp@ucar.edu). The direct consulting service provided by the DS to the RDA user is one of the most valuable resources served through the RDA, as the DS can provide specific insights into data related questions that might not be available through any of the other mechanisms detailed above.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Accept

XII. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The “Research Data Archive Dataset Ingest to Dissemination Workflow Overview” document (https://rda.ucar.edu/rdadocs/RDA_data_ingest_to_dissemination_workflow_overview.pdf) addresses all areas specified in R12. This document is part of the Research Data Archive’s online documentation, and is accessible using the following link: https://rda.ucar.edu/#!rdadocs

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XIII. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Community accepted search, discovery, citation, and persistent identifier mechanisms are provided by the Research Data Archive (RDA) to enable users (both humans and machines) to discover the archived data and refer to them in a persistent way through proper citation.

The RDA provides a range of search capabilities to its holdings. Search options include:

Search provided by external providers:
1)Federated supported search capabilities, such as National Aeronautics and Space Administration (NASA) Global Change Master Directory (GCMD) (https://gcmd.nasa.gov/) that machine-harvests RDA metadata from the RDA's Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) service (https://rda.ucar.edu/cgi-bin/oai), which is one of several available metadata standards from the RDA (See R15 for a list of standards).
2)Commercial search engines, such as Google Dataset Search (https://toolbox.google.com/datasetsearch), that are enabled by the Schema.org (https://schema.org) JSON-LD (https://schema.org/docs/datamodel.html) data model metadata embedded in the RDA dataset home pages.

Native RDA web application search options:
1)Free text keyword search (https://rda.ucar.edu/)
2)Faceted browse (https://rda.ucar.edu/#!lfd?nb=y)

RDA personalized user workspace search capabilities:
RDA users can "bookmark" favorite datasets by clicking on the star stencil on the header of dataset collection homepages. The list of bookmarked datasets can be accessed easily from the user's "dashboard". This capability can help in re-discovering/identifying user designated dataset collections. For additional details, see "Bookmarked Datasets" in: http://ncarrda.blogspot.com/2017/03/the-rda-user-dashboard.html

The RDA is registered in re3data.org (https://www.re3data.org/repository/r3d100010050) to provide official visibility in the repository registry space.

The RDA offers a recommended data citation on each dataset collection's homepage under the "How to Cite this Dataset" section. The citation is available to be reconfigured per several formats, including Earth Science Information Partners (ESIP), Geoscience Data Journal, DataCite, American Meteorological Society (AMS), and American Geophysical Union (AGU) styles. The citation can be downloaded in Research Information Systems (RIS) (https://en.wikipedia.org/wiki/RIS_(file_format)) or BibTeX (http://www.bibtex.org/) format. Please see an example on the following page: https://rda.ucar.edu/datasets/ds094.0/#!description

The RDA offers a "Data Citation" tool as part of the personalized user workspace. The tool generates specified citations from a user's data access history. The generated citation is available in the formats listed above, and includes the value for the "Accessed" date field. For additional details, see "Data Citation" in http://ncarrda.blogspot.com/2017/03/the-rda-user-dashboard.html

The RDA assigns Digital Object Identifiers or DOIs according to the requirements found in the "Digital Object Identifier" section of the following page: https://rda.ucar.edu/#!rdadocs/mm_guide . A general overview of the RDA DOI strategy can be found on the following page: https://rda.ucar.edu/#!data-citation . Please note that the DOI assignment is currently done through the DataCite (https://mds.datacite.org/) assignment service. The DOIs resolve to the corresponding dataset collection home (landing) pages, which are maintained to provide the most up-to-date dataset collection information. All new dataset collections receive DOIs. Efforts are also being made to update legacy dataset collections, so that they are eligible for DOIs, and to mint DOIs for these collections as resources allow. Currently, RDA dataset collections offer DOIs can be found using the following: https://rda.ucar.edu/#!lfd?nb=y&b=doi&v=Matching+Datasets

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XIV. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Overall, the Research Data Archive (RDA) leverages the following strategies to enable reuse and ensure understandability of data over time:

1)A consistent set of metadata requirements at the dataset collection level is enforced to support mapping to community supported schemas.

2)Data submitters are required to use community supported data file format standards to be eligible to archive data in the RDA.

3)When data are restructured from one format to a new format by the RDA, the original data are always retained in the archive to support provenance and reproducibility.

4)Dataset documentation and software provided by the submitter are archived as digital objects in the dataset collection along with the data in non-proprietary formats (e.g. Portable Document Format or PDF).

5)As a mediated repository, Data Engineering & Curation Support Section (DECS) staff work directly with submitters before and during data submission. DECS staff also maintain mechanisms to facilitate ongoing user feedback in order to identify and resolve issues that can help in optimizing the understandability of the datasets.

Additional details on RDA strategies to enable reuse and ensure understandability of data over time are included in the Data Preservation Policy section of the following page: https://rda.ucar.edu/#!daas/terms-and-conditions

To remain current on the Atmospheric Science (AS) community data reuse trends and needs, including metadata schemas and file formats used by the AS community, the DECS staff regularly participate in meetings hosted by and contribute to activities related to:

1)American Geophysical Union (AGU) Earth and Space Science Informatics (ESSI - https://essi.agu.org/)

2)Earth Science Information Partners (ESIP - https://www.esipfed.org/)

3)American Meteorological Society (AMS) Environmental Information Processing Technologies (EIPT - https://www.ametsoc.org/index.cfm/stac/boards/board-on-environmental-information-processing-technologies/) communities.

The knowledge gained from these engagements is used to align RDA metadata and file format requirements with the current and future needs of the AS research community, and to enhance the RDA’s capabilities to meet these needs.

When a new dataset collection is created, the workflow described in this document is followed in full: https://rda.ucar.edu/rdadocs/RDA_data_ingest_to_dissemination_workflow_overview.pdf (also see R12).

In particular, the dataset specialist responsible for the new dataset must add through the Metadata Manager tool (https://rda.ucar.edu/#!rdadocs/mm_guide) the required metadata fields (shown at the bottom of the linked web page) for the dataset collection level. Dataset collection level content metadata, derived from “file level” metadata harvested during data file archival (See “About Data File Content Metadata” in the following page: https://rda.ucar.edu/#!rdadocs/dsmaint), are populated once the files have been archived into the dataset collection. Content metadata is also automatically updated as additional files are archived in a dataset collection over time. For an example of the content metadata, or a summary metadata product derived from “file level” metadata, see:
https://rda.ucar.edu/datasets/ds083.3/#metadata/detailed.html?_do=y&view=level

Additionally,
1)Dataset collection level metadata is maintained in a native RDA schema based on ISO representations (e.g. ISO 8601) and leverages Global Change Master Directory (GCMD) controlled vocabulary keywords (https://earthdata.nasa.gov/about/gcmd/global-change-master-directory-gcmd-keywords).
2)Tools are provided to map the native RDA metadata into discipline standards/community-supported schemas according to the relevant specifications, including:
a)DataCite
b)GCMD Directory Interchange Format (DIF)
c)Federal Geographic Data Committee (FGDC)
d)nternational Organization for Standardization (ISO) 19139 and ISO 19115-3
e)JSON-LD Structured Data

An example of the available standard metadata schemas that are provided by the RDA can be reviewed by using the “Metadata Record” menu found at the bottom of this dataset homepage:
https://rda.ucar.edu/datasets/ds083.2/#!description

3)All of the listed metadata schemas, plus OAI-DC (Dublin Core) and THREDDS schemas, can be accessed through the RDA Open Archive Initiatives Protocol for Metadata Harvesting (OAI-PMH) web service:
https://rda.ucar.edu/cgi-bin/oai/https://rda.ucar.edu/cgi-bin/oai?verb=ListMetadataFormats

When the submitter provides documentation and software related to a dataset, the responsible dataset specialist must archive these resources as digital objects and as part of the the dataset collection. Once archived, these related resources can be accessed through the “Documentation” and “Software” tabs on the dataset collection homepage (For an example, see: https://rda.ucar.edu/datasets/ds131.2/#!description)

Data file formats must adhere to AS community supported standards to be considered for acceptance into the RDA (a list of community accepted file format standards is provided in section 2 of 4 of the following page:

https://rda.ucar.edu/#!daas/worksheet-instructions). For example, these formats are based on World Meteorological Organization (WMO - https://www.wmo.int/pages/index_en.html) and community defined standards, such as Climate and Forecast (CF) NetCDF (https://geo-ide.noaa.gov/wiki/index.php?title=Formats_for_delivery_of_scientific/environmental_data). Nuanced formats, such as user structured ASCII, may also accepted into the RDA when accompanied by a well described specification document. Submitters must provide the proposed file format in their request to archive data submission.

By requiring the use of well-documented, community supported data format standards as part of establishing the data's eligibility to be archived with the RDA, the RDA is well positioned to support evolution of formats, including the following cases:
1)The research community will require new versions of supporting software libraries to be backward compatible with legacy versions of the format.
2)Community supported formats are well documented. Consequently, if needed, the formats can be accessed in the future according to the native format specifications.

When data are migrated from a legacy file format/structure to a new file format/structure, the original legacy formatted data are always retained to ensure that the provenance chain is complete. Retaining data in their original format also enables reproducibility in case there is an issue discovered later with data that has been converted to the new format. Format migration typically occurs when "Enhanced" or "Data-level" curation is performed. For additional details on the rationale behind "Enhanced" and "Data-level" curation, including existing use cases for migrating from legacy file format/structure to a new file format structure, see bullets 2 and 3 in the following document:
https://rda.ucar.edu/rdadocs/RDA_Dataset_Curation_Level.pdf

Unless a decision has been made to perform "Enhanced or "Data-level" curation, data are maintained in their native formats. Currently, only new datasets are being evaluated for "Enhanced or "Data-level" curation. In the future, selected legacy datasets may be considered for conversion to modern formats if the demand is justified and resources are available to do this. Examples of legacy datasets that may be considered for future conversion include:
1)Datasets stored formats that are not self describing. For example, formats that require external tables to define data file metadata encoding such as GRIB (https://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/Introduction_GRIB1-GRIB2.pdf) or BUFR (http://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/GRIB1-Contents.html). If the external tables are not available, these types of data can be challenging to use.
2)Datasets stored in formats that are not optimized for evolving technologies. For example highly valued community datasets copied to cloud object store may need to be migrated to a format that supports efficient cloud object store access, such as Zarr (https://zarr.readthedocs.io/en/stable/).

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

TECHNOLOGY

XV. Technical infrastructure

R15. 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.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Background:
Since its inception in the mid 1960s, the Research Data Archive (RDA) has evolved with technological trends to support its data curation and end user needs. From the 1960s to 1990s, the RDA generally employed a physical data ingest and dissemination workflow. During this period, data were typically sent to the Data Engineering & Curation Section (DECS, formerly Data Support Section (DSS)) via tape media. Data were then ingested from tape into the Nation Center for Atmospheric Research (NCAR) Mass Storage System tape archive for long term preservation in the RDA. Data were also disseminated to users outside of NCAR via tape media. As network and internet technologies developed and matured in the mid to late 1990s, the RDA evolved to a network supported data ingest to dissemination workflow. In order to support this evolution, many in-house software components were developed to augment or replace legacy technologies. This trend has continued to the present as there are many specific niche needs that the in-house developed/maintained software can provide. However, the DECS regularly explores opportunities to replace in-house developed software components with community supported solutions whenever appropriate; this helps the RDA’s functionalities to be more in line with community standards and practices, and in turn, more sustainable.

Since the early 2000s, the RDA infrastructure, metadata, and software architecture has been developed, implemented, and maintained based on the current technology capabilities, user community expectations, and the concepts laid out in the Open Archival Information System (OAIS) reference model (https://www.oclc.org/research/publications/library/2000/lavoie-oais.html).

To remain current on Atmospheric Science and Informatics community software uses, trends, and needs, the DECS staff regularly participate in meetings hosted by and contribute to activities related to:
1)American Geophysical Union (AGU) Earth and Space Science Informatics (ESSI - https://essi.agu.org/)
2)Earth Science Information Partners (ESIP - https://www.esipfed.org/)
3)American Meteorological Society (AMS) Environmental Information Processing Technologies (EIPT - https://www.ametsoc.org/index.cfm/stac/boards/board-on-environmental-information-processing-technologies/) communities

Additionally, DECS staff attend application/technology specific conferences and workshops to remain abreast of current community technology trends. Examples of these conference/workshops include the Globus World (https://www.globusworld.org/) annual conference and local tour stops, Unidata (https://www.unidata.ucar.edu/) users workshops, and Percona Live (https://www.percona.com/) Open Source Database conferences. The knowledge gained from these engagements is used to align RDA infrastructure, metadata, and software implementations with the current and future needs of the Atmospheric Sciences research community, and to enhance the RDA's capabilities to meet these needs.

To support the RDA's dataset ingest to dissemination workflow (https://rda.ucar.edu/rdadocs/RDA_data_ingest_to_dissemination_workflow_overview.pdf), the RDA employs community supported metadata schemas, controlled vocabularies, and in-house developed as well as community maintained software components.

An overview of the software components used in the RDA's dataset ingest to dissemination workflow is provided below:

The following software components are used to support the tools in Figure 1 (i.e. "RDA Dataset Selection and Dataset Metadata Creation") of the workflow:
1)Dataset submission and appraisal tools (1.1) are in-house developed PHP-based (http://php.net/) web applications.
2)The Metadata Manager metadata data entry and validation tool (1.3) is an in-house developed C++-based (http://www.cplusplus.com/) web application. The Metadata Manager maps metadata into a RDA native schema based on International Organization for Standardization (ISO) representations (e.g. ISO 8601), and requires the use of Global Change Master Directory (GCMD) controlled vocabulary keywords (https://earthdata.nasa.gov/about/gcmd/global-change-master-directory-gcmd-keywords) to describe dataset collection parameters.

The following software components are used to support the tools in Figure 2 (i.e. “RDA Dataset Ingest and DOI creation”) of the workflow:

1)The in-house developed dsupdt tool (2.1), used to support automated data ingest for dynamic datasets, is written in PERL (https://www.perl.org/). dsupdt interfaces with the RDA Metadata Database to save configuration preferences using the PERL supported database modules (e.g. DBI, https://dbi.perl.org/). dsupdt uses community supported software, including wget (https://www.gnu.org/software/wget/) and ncftp (https://www.ncftp.com/) to transfer data from remote servers to local RDA/Computational and Information Systems Lab (CISL) servers.

2)Data preparation steps (2.2) are typically performed using community supported data manipulation tools. Examples of data preparation tools used by the DECS include:

a)wgrib2 (http://www.cpc.ncep.noaa.gov/products/wesley/wgrib2/)

b)NetCDF operators (http://nco.sourceforge.net/)

c)Climate Data Operators (https://code.mpimet.mpg.de/projects/cdo/)

d)NCAR Command Language (https://www.ncl.ucar.edu/)

e)ECMWF ECCODES (https://confluence.ecmwf.int/display/ECC/ecCodes+Home)

3)The in-house developed dsarch tool, used to archive data to RDA dataset collection disk and CISL High Performance Storage System (HPSS) systems (2.3), is written in PERL. It uses PERL supported database modules to record file location and description information in the RDA Metadata Database.

4)The in-house developed gatherxml tool, used to extract format specific file level metadata and write that metadata into the RDA Metadata Database and to RDA Web Server disk (2.3), is written in C++ and uses the community supported C++ MySQL connector (https://dev.mysql.com/doc/connector-cpp/8.0/en/) to interface with the RDA Metadata Database.

5)The Metadata Manager metadata data entry and validation tool (2.4) is an in-house developed C++-based (http://www.cplusplus.com/) web application. The Metadata Manager includes a C++ module that maps native RDA metadata into required DataCite (https://support.datacite.org/docs/schema-40) metadata elements and calls the DataCite API (https://mds.datacite.org/) to mint RDA dataset DOIs.

The following software components are used to support the tools in Figure 3 (i.e. “RDA Dataset Publication and Dissemination”) of the workflow:

1)The in-house developed publish_filelist tool (3.1), used to publish dataset collection file inventories for user access, is written in PERL and interfaces with the RDA Metadata Database using PERL supported database modules.

2)The in-house developed scm tool (3.1), used to generate content metadata summaries for inclusion in dataset collection level metadata on RDA Web Server Disk, is written in C++ and and uses the community supported C++ MySQL connector to interface with the RDA Metadata Database.

3)Several web applications provide interfaces for users to search, discover, and access archived data and metadata (3.2) including the following:

a)In-house developed faceted and free text search applications that are written in C++. Both of these applications use the community supported C++ MySQL connector to interface with the RDA Metadata Database.

b)In-house developed subset request applications that are written in C++, PHP, and Javascript. These use the community supported C++ MySQL connector and PHP PDO driver (http://php.net/manual/en/ref.pdo-mysql.php) to interface with the RDA Metadata Database.

c)In-house developed OAI-PMH server, used to distribute standards structured dataset metadata (3.3), is written in C++. This uses the community supported C++ MySQL connector to interface with the RDA Metadata Database. The OAI-PMH server uses community metadata specifications to map RDA native metadata into multiple schemas (see R14 for additional details on provided metadata schemas).
d)The community supported Unidata Thematic Real-time Environmental Distributed Data Services (THREDDS - https://www.unidata.ucar.edu/software/thredds/current/tds/) is used to support Open-source Project for a Network Data Access Protocol (OPeNDAP) data access (3.5).
e)The externally supported Globus Connect Server (https://www.globus.org/globus-connect-server) (3.8) is used to support Globus maintained GridFTP data transfers (https://www.globus.org/#transfer) (3.9).
f)The in-house developed dsrqst tool (https://rda.ucar.edu/rdadocs/dsrqst/), which automatically manages user data request processing, is written in PERL, and coordinates user request processing workflows that run on CISL High Performance Computing (HPC) systems (https://www2.cisl.ucar.edu/resources/resources-overview) (3.6, 3.7). dsrqst uses PERL supported database modules to interface with the RDA Metadata Database.

General software/server/infrastructure components used across all components of the RDA dataset ingest to dissemination workflow include the following:
1)The RDA Metadata Database that runs on the open source MySQL 5.7 database server (https://www.mysql.com/).
2)The web applications that currently run on Apache 2.4.x HTTP server (https://httpd.apache.org/). The Apache 2.4.x HTTP server is also used to support HTTP based data transfer activities (3.4).
3)The RDA Web and Metadata Databases that run on virtual machines, which use the CentOS 7 operating system. The virtual machines operate on a CISL maintained VMWare (https://www.vmware.com/) server cluster.
4)The RDA Dataset Collection Disk that runs on a IBM Spectrum Scale General Parallel File System (https://www.ibm.com/support/knowledgecenter/en/SSFKCN/gpfs_welcome.html).
5)The CISL HPSS, which hosts RDA data, consists of robotic tape libraries with storage capacity of more than 320 petabytes at the NCAR-Wyoming Supercomputing Center (NWSC) Cheyenne, WY and another 15 petabytes of disaster-recovery data storage at the NCAR Mesa Lab in Boulder, CO. These scalable systems comprise four StorageTek SL8500 tape libraries. File metadata are maintained in DB2, IBM's real-time database system. For additional information, see: https://www2.cisl.ucar.edu/resources/storage-and-file-systems/hpss
6)CISL’s Network Engineering and Telecommunications Section (NETS, http://nets.ucar.edu/nets/intro/introduction.shtml) maintains high volume and high availability network connectivity to support programmatic/automated RDA data ingest workflows effectively. Additionally, auto retry capability is integrated into the RDA dsupdt tool (https://rda.ucar.edu/rdadocs/dsupdt/) to support data ingest recovery as needed after system/network outages. The RDA does not maintain “real-time” datasets, where immediate access is essential to support user needs. All RDA assets are considered to be for research use only, so although NETS typically provisions around-the-clock connectivity to public and private networks at a bandwidth that is sufficient to meet the global and/or regional responsibilities, 24x7 connectivity is not essential to support the needs of the RDA user community.

Additional details on Metadata, Software, and Infrastructure not captured above are included below:

Metadata:
As highlighted above, dataset collection level metadata is maintained in a native RDA schema, which is based on ISO representations (e.g. ISO 8601), and leverages Global Change Master Directory (GCMD) controlled vocabulary keywords (https://earthdata.nasa.gov/about/gcmd/global-change-master-directory-gcmd-keywords). Tools are provided to map the native RDA metadata into community standards based schemas according to the relevant standard specifications, including DataCite, GCMD Directory Interchange Format (DIF), Federal Geographic Data Committee (FGDC), International Organization for Standardization (ISO) 19139, ISO 19115-3, and JSON-LD Structured Data. An example of the available standard metadata schemas provided by the RDA can be viewed by using the the "Metadata Record" menu found at the bottom of the following dataset homepage: https://rda.ucar.edu/datasets/ds083.2/#!description

Additionally, all of the listed metadata schemas, plus OAI-DC (Dublin Core) and THREDDS schemas, can be accessed through the RDA Open Archive Initiatives Protocol for Metadata Harvesting (OAI-PMH) web service: https://rda.ucar.edu/cgi-bin/oai/https://rda.ucar.edu/cgi-bin/oai?verb=ListMetadataFormats

Software:
As detailed above, the RDA employs a combination of in-house developed software and community supported software components to support data curation, data discovery, and data access workflows. An inventory of in-house developed software components is maintained in the 33 repositories organized under the NCAR/RDA team institutional GitHub space. Due to security concerns, there is currently a mix of publicly available and restricted repositories maintained in the RDA team space, so all repositories are not visible to external parties (a listing of the 12 publicly available repositories can be found here: https://github.com/NCAR?utf8=%E2%9C%93&q=RDA). Documentation is included as READMEs in each RDA team repository. DECS staff are working to make additional GitHub RDA repositories publicly available by by removing any sensitive information as time allows.

Infrastructure:
Quarterly meetings are held between relevant DECS staff to develop estimates of future storage requirements based on the regular, automated ingest stream volumes, and on estimated future product volumes that will be coming into the archive. Based on this information and on a yearly basis, CISL allocates HPSS and RDA dataset disk resources as needed to support future RDA growth.

Load usage and performance is actively monitored on all RDA supported servers and services to ensure performance continues to meet expectations. New servers are procured based on forecast usage metrics and recorded load usage on a 4-5 yearly basis.

DECS management participates in CISL strategic planning exercises on a bi-yearly basis to ensure that RDA service offerings evolve to meet current and future user expectations. Additionally, DECS management meets with CISL management on a monthly basis to review current services offerings, and determine whether or not these need to be adjusted to meet existing user expectations.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XVI. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The Research Data Archive's (RDA) Information Technology (IT) infrastructure, supported by the National Center for Atmospheric Research (NCAR) Computational and Information Systems Lab (CISL), provides highly available storage, virtual machine (VM) supported web and database services, and backup and disaster recovery for archived data (see R9 for additional details on data preservation and security, including disaster recovery). Additionally, robust security and monitoring infrastructure is maintained by CISL and NCAR, including physical building and cyber infrastructure (CI) security and monitoring. Further, risk analysis exercises and business continuity planning is led institutionally by the University Corporation of Atmospheric Research (UCAR), which manages NCAR/CISL. Publicly available details on all of these topics can be found in the links provided below.

RDA web and database servers run on VMware virtual servers in a cluster environment at the NCAR Wyoming Supercomputing Center (NWSC, https://nwsc.ucar.edu/) located in Cheyenne, WY. In the event of hardware failures, the VMs will be restored by CISL Enterprise Services Section (ESS, https://staff.ucar.edu/orgs/ess) staff on working cluster nodes. The ESS service level agreement supports maintenance on RDA VM's between 7AM and 7PM Monday - Friday, excluding holidays.

RDA web, Metadata Database, Dataset Collection Disk, and Cheyenne (NWSC) High Performance Storage System (HPSS) are maintained on the NWSC uninterruptible power supply (UPS) backup. In the event of a utility power issue, all RDA infrastructure remains available on the UPS system. This infrastructure is maintained by the CISL Cheyenne Operations Section (https://www2.cisl.ucar.edu/cos-org-chart).

CISL CI security and monitoring:
An overview of CISL CI security related procedures can be found on the following page:
https://www2.cisl.ucar.edu/user-support/authentication-and-security

All CISL staff are required to complete CI security training modules. These modules include training on how to recognize various types of phishing/whaling strategies that typically occur through email.

The CISL Cheyenne Administration Support Group (CASG) provides 24x7 monitoring of CISL cyberinfrastructure in coordination with CISL ESS, according to UCAR’s Computer Security Advisory Committee (CSAC) guidelines (https://www2.fin.ucar.edu/it/security). An overview of CISL user support for cyberinfrastructure can be found on the following page: https://www2.cisl.ucar.edu/user-support/cisl-resource-status

CISL ESS provides additional monitoring of RDA VM’s, and applies system updates/patches as new stable operating system (OS) versions become available and security implications require.

CISL Workstation Support Services Team (WSST) provides support for staff desktops and laptops, and performs regular virus scans and provides email monitoring for viruses according to UCAR’s CSAC guidelines (https://www2.fin.ucar.edu/it/security). Additional information can be found on the following page:
https://www2.cisl.ucar.edu/wsst/antivirus

Additional background on UCAR cybersecurity can be found in the 2017 Annual report (https://nar.ucar.edu/2017/cisl/formalize-and-enhance-ucar%E2%80%99s-cybersecurity-capabilities).

NCAR/CISL physical site security and active monitoring:
An overview of physical site security and active monitoring strategies can be found on the following page:
https://www2.fin.ucar.edu/security

NCAR/UCAR Risk analysis planning:
Publicly available information on this topic can be found in the “Risk Management and Resiliency” section of the following document: https://rda.ucar.edu/rdadocs/RDA_data_security.pdf

NCAR/UCAR Disaster and business continuity plan:
UCAR's business continuity plan is based on the following standards:

1)U.S. Department of Homeland Security, Federal Emergency Management Agency (FEMA)

2)NFPA 1600:2007 Standard on Disaster/Emergency Management and Business Continuity Programs

3)ISO 22301 • NIST SP 800-34 Contingency Planning Guide for Information Technology Systems

4)DRII/DRJ GAP Generally Accepted Practices for Business Continuity Practitioners

Publicly available information on disaster and related business continuity planning can be found on the following page: https://internal.ucar.edu/president/business-continuity

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

APPLICANT FEEDBACK

Comments/feedback

These requirements are not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response:

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments: