Assessment Information

CoreTrustSeal Requirements 2017–2019

Repository: Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC)
Website: http://paradisec.org.au/
Certification Date: 21 November 2019

This repository is owned by: University of Sydney, University of Melbourne, Australian National University
Core Trustworthy Data Repository Requirements

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type.

Domain or subject-based repository
Library/Museum/Archives
Research project repository

Reviewer 1:
Accept

Reviewer 2:
Accept

Comments

PARADISEC operates as one of the Pacific region’s most significant digital repositories of linguistic, ethnomusicological and cultural data. While we house valuable data from all over the world, our main focus is on Australia and the Pacific region, within which, since our inception in 2003, we have come to assume a major role of stewardship of cultural data from the region, as well as regional leadership in the field of digital archiving.


Reviewer 1:
Accept
Brief Description of the Repository’s Designated Community.

PARADISEC is primarily a research repository that provides curated primary data in a citable form and so is crucial research infrastructure for linguists, musicologists and anthropologists, who create archival collections in the course of their research. An important additional function of this collection is that it houses ethnographic material that is of significant cultural and historical interest to the speakers recorded and their descendants.

In summary, PARADISEC’s designated community is:

* researchers in the humanities and social sciences
* community members (individuals and organisations) represented in our records

Level of Curation Performed.

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

Comments
The level of curation we provide varies by collection depending on a number of factors. In general, our depositors provide us with (at minimum) a basic level of metadata, which is archived along with the data they provide. We provide format conversion for quite a number of our collections, including digitisation from analog formats, as this is a focus of PARADISEC’s, to save records that are potentially at risk of being lost due to physical degradation of media. Additionally, we often provide some level of metadata enhancement that may happen at various stages of our processing of the data and metadata.

Reviewer 1:
Accept

Reviewer 2:
Accept

Outsource Partners. If applicable, please list them.

Intersect (independent eResearch support agency, providing PARADISEC with data storage and support; more information https://intersect.org.au/); see attached document “PARADISEC Infrastructure Documentation” [ANNEX A].

DAMsmart (independent contracted digital media conversion services, providing PARADISEC with video digitisation and transcoding services. More information https://damsmart.com.au/)

Reviewer 1:
Accept

Reviewer 2:
Accept

Other Relevant Information.

PARADISEC’s operations are predominantly spread across three institutions:

* University of Sydney - major hosting/funding partner; Director Amanda Harris; Chair of Steering Committee Linda Barwick; data storage; catalog and website hosting; technical support, including catalog and websites; professional development; main audio digitisation facilities
* University of Melbourne - major hosting/funding partner; Director Nick Thieberger; audio digitisation lab; technical development

* Australian National University - PARADISEC is the major designated repository for the Centre of Excellence for the Dynamics of Language, which is based at Australian National University, and where various aspects of specific collections are managed. (http://www.dynamicsoflanguage.edu.au/)

Reviewer 1:
Accept

Reviewer 2:
Accept
ORGANIZATIONAL INFRASTRUCTURE

I. Mission/Scope

Compliance Level: 4

Reviewer 1:

4

Reviewer 2:

4

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

PARADISEC (the Pacific And Regional Archive for Digital Sources in Endangered Cultures) is a digital archive of records of some of the many small cultures and languages of the world. Our research group has developed models to ensure that the archive can provide access to interested communities (such as the Data Loader application, which can be used to view metadata associated with files, in areas where access to our online catalog may be limited [https://language-archives.services/about/data-loader]; and utilising open source tools such as LibraryBox http://librarybox.us/ and Raspberry Pi https://www.raspberrypi.org/); and to ensure that the archive makes effective use of international digital archiving schemas (such as the Dublin Core Metadata Element Set, and many aspects of the Open archival information system [OAIS] Reference Model [ISO 14721:2012]).

We have established a framework for accessioning, cataloguing and digitising audio, text and visual material, and preserving digital copies. A primary goal is to safely preserve material that would otherwise be lost. In this way we can make field recordings available to the people and communities recorded, and to their descendants.

We have distributed copies of recordings to the Vanuatu Cultural Centre, the University of New Caledonia, the Institute of Papua New Guinea Studies, the Solomon Islands National Museum and to Rapa Nui.

All relevant information is on our website: http://www.paradisec.org.au/about-us/

Reviewer 1:

Accept
Reviewer 2:

Accept
II. Licenses

Compliance Level: 4

Reviewer 1:
4

Reviewer 2:
4

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

All METADATA held by PARADISEC are licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (https://creativecommons.org/licenses/by-sa/4.0/).

The access and use of FILES in the archive are covered under our "Conditions of Access" document (http://www.paradisec.org.au/deposit/access-conditions/). Only registered users can access files in the catalog, so this means that in order to access files it is necessary to agree to these Conditions of Access. Further, at the point of downloading any file, the same Conditions of Access are displayed and the user must again agree (by checkbox) to the Conditions in order to proceed with the download.

Individual items within a collection can also have their own specified access levels/restrictions. This allows a depositor to make some items/files open access, but keep others within the same collection closed or restricted. Additionally, each collection and item has a 'data access narrative' field (at collection level this field is called 'data access details', but the purpose is the same), which is used to display details of the access conditions to catalog users. For instance, if the data is currently closed but is due to be made open on a specified date, that information will be displayed in the 'data access narrative/details' field.

These licensing and user agreements, differentiated between metadata and data (e.g. audio and video files within the archive), is specified at the footer of each page of the catalog, and also in the PARADISEC website http://www.paradisec.org.au/deposit/access-conditions/

Reviewer 1:
Accept
Reviewer 2:

Accept
III. Continuity of access

Compliance Level: 3

Reviewer 1:
3

Reviewer 2:
3

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

We have been operating for 15 years and have a commitment from the University of Sydney to continue to house the collection into the future.

Following extensive discussion on this topic in our 2018 Steering Committee Meeting, we are planning to make progress in this area over the next 12 months, and can provide further documentation once we have it in place. Discussions have commenced with University of Sydney to provide us with relevant documentation, which we will pass on to ICSU-WDS when it is obtained, to hopefully upgrade our compliance to level 4 in this area.

Reviewer 1:
Accept

Reviewer 2:
Accept
IV. Confidentiality/Ethics

Compliance Level: 4

Reviewer 1:

4

Reviewer 2:

4

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

We pay careful attention to ethical norms and are highly responsive to feedback from people whose recordings are in the collection. To date we have only received praise and gratitude from speakers or their descendants for housing these recordings that would otherwise have been lost.

Our deposit form (http://www.paradisec.org.au/PDSCdeposit.pdf), which must be filled in and signed by a depositor for each of their collections, acknowledges PARADISEC’s “duty to safeguard the interests and sensitivities of relevant indigenous people,” and its “curatorial and custodial responsibilities,” and to this end the form requests details from the depositor of their understanding of the attitude of the people recorded to this material being made accessible to other people. The next section on the form asks the depositor about any special conditions required in handling the material, “e.g. ceremonial or gender restricted material, sensitive genealogical material, photographs or recordings of deceased people.” This is specifically relevant to Pacific Indigenous materials, which represents much of PARADISEC’s data, and staff are trained to handle data with these special access conditions.

The vast majority of our data are “Open” access, as specified by the depositor. This means the depositor has agreed that they (and to the best of their awareness, the people and communities represented in the recordings) are happy for the material to be freely accessed, and used in compliance with the Creative Commons Licence, and PARADISEC’s user agreement (http://www.paradisec.org.au/deposit/access-conditions/)

For data that has been collected in the last few decades by university-affiliated researchers, these have been subject to the relevant Research Ethics standards and procedures of the researcher’s university or institution, such as fully informing participants of the nature of the research they are partaking in, and gaining appropriate written or verbal consent.
For material that was collected more historically, PARADISEC relies on the honest written declaration of the depositor (in their signing of the PARADISEC deposit form) that the access level they assign is, to the best of their understanding, appropriate to the material and in keeping with the wishes of the relevant individuals and communities. PARADISEC always remains open to community input and feedback with respect to such matters.

Reviewer 1:
Accept

Reviewer 2:
Accept
V. Organizational infrastructure

 Compliance Level: 3

 Reviewer 1:
 3

 Reviewer 2:
 3

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.

PARADISEC is formally hosted by the University of Sydney. Along with Sydney University’s commitment to support PARADISEC into the future, they provide:

* our Sydney offices in the Sydney Conservatorium of Music;
* local IT infrastructure & support;
* access to regular and ongoing training and professional development modules in relevant areas.

Our close affiliation with Sydney University, along with the University of Melbourne (where one of our Directors, Nick Thieberger, is based), and the Australian National University (where the Centre of Excellence for the Dynamics of Language is hosted, of which PARADISEC is a primary nominated repository), ensures that we maintain excellent ties with the researchers who form a significant part of our designated community.

PARADISEC currently has 8 dedicated staff members working out of our Sydney office, including Sydney unit Director Dr Amanda Harris. Amongst these staff are trained and experienced operations/project officers, and audio preservation specialists who deal with analog and digital media including digitization. Many of our staff combine academic and technical backgrounds to bring a richness of experience and skill to their roles in PARADISEC. Further details: http://www.paradisec.org.au/about-us/our-staff/

Whilst we are always seeking funding, we have a digitisation service that generates some income to support other activities. Between 2014-2021 we are part of the Australian Research Council funded Centre of Excellence for the Dynamics of Language that ensures a continual steam of funds for part of our operation. (http://www.dynamicsoflanguage.edu.au/home/)
Reviewer 1:
Accept

Reviewer 2:
Accept
VI. Expert guidance

Compliance Level: 4

Reviewer 1:
4

Reviewer 2:
4

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

PARADISEC is overseen and advised by a Steering Committee (http://www.paradisec.org.au/about-us/steering-committee/). This Committee has members around the world in relevant disciplines of archiving, data management, linguistics, and Pacific languages, music & culture. The Committee meets every 12 months, and this is a valuable forum for receiving feedback on PARADISEC's work from this research arm of our Designated Community.

We have regular contact with academic and community/cultural organizations in several of the Pacific countries where much of our data originate, such as the Divine Word University (Madang, PNG), the Vanuatu Cultural Centre (Port Vila, Vanuatu), and the Sydney Wantok Society (PNG community in Sydney, Australia). Through these ties we maintain open lines of communication with communities to receive input and feedback on our work.

We are a member of the International Association of Sound and Audiovisual Archives (IASA) (https://www.iasa-web.org/), where we utilise and contribute to forums.

Our staff regularly attend and contribute to (including present at) various relevant national and international conferences in the field of digital archiving. See our publications list for further details: http://www.paradisec.org.au/resources/

We have always had the support of the National Library of Australia (https://www.nla.gov.au/) and the National Film and Sound Archive (https://www.nfsa.gov.au/) for our work.

Reviewer 1:
Accept
Reviewer 2:

Accept
DIGITAL OBJECT MANAGEMENT

VII. Data integrity and authenticity

Compliance Level: 4

Reviewer 1:
4

Reviewer 2:
4

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

We have regular script-automated checking of md5-checksums for all files, that ensures integrity. Checksums are cross-checked at several points of data transference, and also periodically throughout the entire archived collection, to audit for any errors.

We have a daily offsite backup and we engage in disaster recovery exercises to ensure our backup systems are functioning.

We use Dublin Core metadata standards (http://dublincore.org/specifications/). A minimal standard of metadata is ensured (sufficient to allow harvesting by OLAC and RDA) by our catalog’s inbuilt system of rejecting files going in to the archive that do not meet this minimum metadata requirement. (In practice almost all of our collections actually have a much higher standard of metadata, but this system ensures at least a ‘minimal standard’.) OLAC displays our average metadata score as 9.4 /10 (http://www.language-archives.org/metrics/paradisec.org.au), with an overall rating of five-stars.

Our online catalog logs all changes to metadata, including which logged-in user performed the change, date and time of the change, and details of the amendment made. This allows us to track and check any changes made in the collection. Status of recordings as 'born-digital' or 'digitized' is recorded in the metadata but is only viewable to admin users (not to general users).

We are a reasonably small operation and need to do whatever it takes to ensure the data is kept safe, but we cannot undertake some of the functions listed below (data versioning, audit trails etc) as we do not have the resources. Much of our data is digitised from analog tapes and has no versioning.

Data authenticity is asserted by depositors on their signed deposit forms (http://www.paradisec.org.au/PDSCdeposit.pdf).
Reviewer 1:
Accept

Reviewer 2:
Accept
VIII. Appraisal

Compliance Level: 4

Reviewer 1: 4

Reviewer 2: 4

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


We assess all incoming data for relevance and prioritise for ingestion those records most at risk of being lost. In practice, this means collections on old analog media (such as reel-to-reel tapes), which are at high risk of degradation, are prioritised. There are few repositories that cater to this kind of data and we are continually asked to archive material from researchers all over the world. Our ingestion workflow model (formerly referred to as our "Operations Manual") can be found here: http://www.paradisec.org.au/ingestion_workflow_feb2017/

All data and metadata are checked by one of our trained staff, to ensure only complete records are archived. Any incompleteness in data or metadata are referred back to the depositor, and resolved in partnership between PARADISEC and the depositor. We do not allow data to be uploaded directly by depositors, which ensures we maintain quality-control oversight, as well as format standardisation, over all data that we archive. If data is sent to us in non-preferred formats, we first request that the depositor re-format the data, and then re-submit to us, and if that is not possible (e.g. a deceased depositor with no active management of the data by their estate) then we are usually able to manage the re-formatting of the data, either in-house or using one of our outsource partners. For metadata, we allow depositors to add or edit the metadata for their own collection (not including "Admin only" metadata fields), but we also oversee this to ensure that metadata being added or amended adheres to our standards.

We publish detailed information on formats, and other data and metadata requirements (e.g. preferred audio file bit and sample rates) on our website, here http://www.paradisec.org.au/deposit/, and also as downloadable information sheets, here http://www.paradisec.org.au/resources/downloads/. We refer all new depositors to these information pages.
Reviewer 1:
Accept

Reviewer 2:
Accept
IX. Documented storage procedures

Compliance Level: 3

Reviewer 1:
3

Reviewer 2:
3

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

We publish our Ingestion Workflow model, and other documents relating to our operations here http://www.paradisec.org.au/resources/downloads/, and periodically review procedures. We have a data management system that takes care of our workflow, sending reports on items ingested and alerting us to any issues in the process.

We have off-site backup of the entire collection. We have run data recovery drill operations, simulating loss of sections of the archive's data, and enacting restoration from the back-up copy, which were completed successfully. We have checksum scripts that periodically cross-check archived data against the files as originally ingested, to confirm consistency.

Further information on data recovery provisions and risk management can be found in the attached document PARADISEC Infrastructure Documentation, provided by our data storage partner Intersect [ANNEX A].

Reviewer 1:
Accept

Reviewer 2:
Accept
X. Preservation plan

Compliance Level: 3

Reviewer 1:
3

Reviewer 2:
3

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

All items held by PARADISEC are understood to be archived for long-term preservation (i.e. there is no differentiation amongst items in terms of 'preservation level').

Upon accepting a deposit into the archive, we require a signed deposit form from the depositor (http://www.paradisec.org.au/PDSCdeposit.pdf). Returning of the signed deposit form to PARADISEC enacts a clear transfer of custody and responsibility (as laid out in the deposit form) between PARADISEC and the depositor. The document (deposit form) grants PARADISEC the necessary rights and powers to meet its responsibilities with regards to preservation of the materials.

Reviewer 1:
Accept

Reviewer 2:

In the future, evidence should be provided about PARADISEC’s strategy to preserve the accessibility/usability of its holdings in the face of cultural and technological change. Note that the deposit agreement does not explicitly specify that PARADISEC has the right to copy and transform (e.g., format conversion) any digital files received (i.e., recordings that were not digitized by PARADISEC). It is recommended to include such a statement in the agreement.
XI. Data quality

Compliance Level: 4

Reviewer 1:
4

Reviewer 2:
4

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.

We have developed a metadata schema that is sufficient to allow discovery of items in the collection, but is not too detailed to prevent depositors from entering metadata (which is the case for other repositories we have experience of). We require a minimal set of metadata terms, represented in this spreadsheet http://www.paradisec.org.au/wp-content/uploads/2018/06/PDSCMinimalMetadata2018.xlsx, recognising that, for some of our collections, all we know about a recording is what is written on the tape box. For current collections we allow rich metadata that is also harvested by several services and made available via RIF-CS and OAI-PMH feeds.

Our metadata is harvested by Open Language Archives Community www.language-archives.org (OLAC) and receives an average metadata score of 9.4 (94%), with no known 'integrity problems': http://www.language-archives.org/metrics/paradisec.org.au

Each Collection, as well as each Item within each Collection, contains a citation link for ease and accuracy or citation. These citation links include Digital Object Identifiers (DOIs), for persistence of identifiability and citation. For example:


All collections in PARADISEC have publicly available metadata (even in cases where the data themselves have access restrictions), and our catalog allows users to submit comments. These comments are sent to the PARADISEC admin team, who review and act upon the comments. This feedback mechanism has been used by community members on several occasions to make comments on data or metadata, which we have then - upon positive verification - been able to incorporate into the relevant collections for their improvement.
Reviewer 1:  
Accept  

Reviewer 2:  
Accept
XII. Workflows

Compliance Level: 4

Reviewer 1:

4

Reviewer 2:

4

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

We have a data management system that enforces an ordered set of steps for ingestion of data. Our website details our processes for depositors in a number of downloadable resources here:


Of particular relevance for this area of compliance, our archiving workflow documents:


The first (Paradisec_ArchivingWorkflow.pdf) is simplified and geared towards explanation of processes for general users and depositors. The second (Ingestion_Workflow_Feb2017.pdf) is more detailed and technical, used more for internal management and auditing of procedures, and for an appropriately technically-skilled audience.

Also see attached Audio workflow model [ANNEX B].

Reviewer 1:

Accept

Reviewer 2:

Accept
XIII. Data discovery and identification

Compliance Level: 4

Reviewer 1:
4

Reviewer 2:
4

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


Our catalog provides a ‘cite as’ field for each record and we also provide a Digital Object Identifier (DOI) for all Items and Collections within PARADISEC.

The Cite As field includes all contributors to the referenced collection or item.

Reviewer 1:
Accept

Reviewer 2:
Accept
XIV. Data reuse

Compliance Level: 4

Reviewer 1: 
4

Reviewer 2: 
4

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.

It is a basic tenet of our repository that data should be available for re-use, and we encourage the deposit of any enriched material arising from that re-use. Re-use is furthered by incorporation of Dublin Core metadata into our basic metadata requirements. Our basic metadata set is here: 
http://www.paradisec.org.au/deposit/metadata/

Our metadata makes use of a number of international standard codes, such as ISO 3166-2 country codes, ISO 639-3 language codes, and Australian and New Zealand Standard Research Classification (ANZSRC) Field of Research Codes.

Our choices of formats for archiving are informed by internationally agreed upon standards, including (but not limited to) those found in IASA (International Association of Sound and Audiovisual Archives) Technical Guides:

https://www.iasa-web.org/tc04/audio-preservation

https://www.iasa-web.org/tc06/guidelines-preservation-video-recordings

We monitor the evolution of standard archiving formats via open channels of contact with professional archiving organisations and bodies, such as IASA (who we are active members of), and the UK Digital Preservation Coalition (from whom we received a Special Commendation in their 2016 awards: https://www.dpconline.org/news/finalists-for-the-digital-preservation-awards-safeguarding-the-digital-legacy-category-announced).

Because part of our Designated Community may be in remote areas with only low internet bandwidth available, we provide data in formats that are more easily downloaded or streamed (e.g. mp3, mp4), as well as having an "online viewer" capability built in to our catalog. This allows community members who might be
accessing our records on a mobile device (or otherwise) with minimal download space/capability to nevertheless access the data.

Reviewer 1:
Accept

Reviewer 2:
Accept
TECHNOLOGY

XV. Technical infrastructure

Compliance Level: 4

Reviewer 1:

4

Reviewer 2:

4

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.

We have migrated the collection a number of times to move to new platforms, and have updated the catalog and data management systems several times in the life of the project.

In our main office, in Sydney, we have ICT support from the University of Sydney, including currently a dedicated systems engineer (part-time) who assists us with keeping our systems optimised, and making system documentation available.

Details of the data storage technical infrastructure provided by Intersect can be found in the attached PARADISEC Infrastructure Documentation [ANNEX A].

Additionally, PARADISEC utilises our own custom-built open source catalog software named NABU. System documentation is available in GitHub: https://github.com/nabu-catalog/nabu

Reviewer 1:

Accept

Reviewer 2:

Accept
XVI. Security

Compliance Level: 4

Reviewer 1:

4

Reviewer 2:

4

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

We have secure services that allow users to designate others who can access their items. We apply access restrictions as specified by depositors. We have no specific systems as described below, but use normal sftp/ssh protocols to access records as admins.

Because our archival storage is provided through the University of Sydney, we benefit from the security measures carried out by such a large and well-established institution. University of Sydney ICT Security is outlined here in their "Information Security Policies":


The attached PARADISEC Infrastructure Documentation [ANNEX A] has sections on Security System and Backup Procedures (page 2), and Disaster Recovery (page 3).

Reviewer 1:

Accept

Reviewer 2:

Accept
APPLICANT FEEDBACK

Comments/feedback