



Movebank Data Repository

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:

True

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 Movebank Data Repository (MDR, <http://www.datarepository.movebank.org>), established in 2011, publishes data packages of animal-borne sensor data, often and hereafter referred to as “animal tracking” data, that are associated with a written publication. The MDR is a close collaboration between the Communication, Information, Media Centre of the University of Konstanz (KIM, <https://www.kim.uni-konstanz.de/en>) and the Max Planck Institute of Animal Behavior (MPIAB, <https://www.ab.mpg.de>). The MDR is a public digital archive administered by the KIM. The MDR has close connections to the Movebank platform (Movebank, <https://www.movebank.org/cms/movebank-content/about-movebank>), an e-infrastructure for animal tracking data administered by the MPIAB. The MDR enables formal archiving and publication of data originally collected and stored on Movebank, using software, database, procedures, and administration specific to the MDR. As described in this application, many data preparation steps are currently undertaken in Movebank prior to ingest in the MDR. Further, the MDR is planning to expand to include two collections: one for animal tracking datasets and one for data analysis workflows. This application for CTS certification applies only to the current status of the MDR, consisting of published packages of animal tracking data. A more comprehensive description of the repository is available at <https://www.movebank.org/cms/movebank-content/data-repository>.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Brief Description of the Repository’s Designated Community.

Our target community includes researchers, government agencies, conservation organizations and other groups that collect and/or use animal tracking and other animal-borne sensor data. This includes the fields of movement ecology, wildlife management and conservation, and biodiversity informatics. Movebank supports over 3,000 data owners, and the MDR has published data reported in 100 journals for 678 unique authors. Most depositors publish data to meet journal requirements when publishing peer-reviewed papers. We interact daily with this target community in preparing datasets to publish and through user support and development for Movebank.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Level of Curation Performed. Select all relevant types from:

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

We offer advanced data review and curation for all submissions to the MDR. Movebank currently serves as the submission platform for data published in the MDR. This ensures that all data attributes and values are transformed to and published using a consistent data model (<https://doi.org/10.1016/j.envsoft.2010.12.005>, <https://www.movebank.org/cms/movebank-content/mb-data-model>) and vocabulary (<http://vocab.nerc.ac.uk/collection/MVB>). MDR curators complete data review and quality assurance / quality control using tools on Movebank and other software applications such as R (<https://www.r-project.org>). Movebank provides tools for semi-automated data review (<https://www.movebank.org/cms/movebank-content/upload-qc>), allowing MDR curators to efficiently assess whether files were imported correctly, require or request additional relevant information that is referred to in related papers, and update data with changes and additions prior to exporting and publishing final data files in the MDR. These tools are available for all Movebank users, but for non-published data, data owners are not generally required to complete these steps or to have the results reviewed by Movebank staff. In contrast, the MDR review procedures completed in Movebank by MDR staff are a mandatory step for publication in the MDR. Because the MDR software and storage hardware are independent from Movebank, and because the data vocabulary and model are publicly described, the submission, storage, and dissemination of data through the MDR could continue if the Movebank platform became unavailable.

Following data review and curation, MDR staff download completed files from Movebank, upload the files to the MDR, and prepare and submit metadata and readme files. Every published data package includes at a minimum

- (a) dataset-level metadata in DataCite format,
- (b) a tabular text file of event data (e.g., GPS location records),

(c) a tabular text file of deployment data (e.g., tag and animal IDs and descriptors, deployment start/end dates), and
(d) a readme file based on a standard template and prepared by MDR staff.

The data package may also include additional event data tables (e.g., for multiple sensor types) or other relevant files in a generic format, for which a description is provided in the readme. Further details are provided throughout this application, in particular R8, and in the Submission Guidelines (<https://www.datarepository.movebank.org/submissionguidelines>).

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Insource/Outsource Partners. If applicable, please list them.

Insource Partners.

- Communication, Information, Media Centre (KIM) University of Konstanz: technical host of the Movebank Data Repository
- Max Planck Institute of Animal Behavior (MPIAB): curation of the submitted data sets and user management

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Summary of Significant Changes Since Last Application (if applicable).

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:
Accept

Other Relevant Information.

The following glossary clarifies references throughout this application:

- Data author: A creator of a dataset submitted to the MDR
- Data depositor: An author of a dataset submitted to the MDR who completes the Depositor Agreement serves as the primary point of contact during the review on behalf of all creators of the dataset
- Data owner: A creator or manager of animal tracking data who may submit their data for publication in the MDR
- KIM: Communication, Information, Media Centre of the University of Konstanz
- MDR: Movebank Data Repository
- MPIAB: Max Planck Institute of Animal Behavior

As of spring 2021, the MDR has published 234 data packages, including over 100 million animal locations describing movements and behavior of over 10,000 animals of 170 taxa. Published data packages are discoverable through Clarivate Analytics' Data Citation Index (<https://clarivate.com/webofsciencegroup/master-data-repository-list>) and Google's Dataset Search. Our repository is discoverable at the re3data.org (<https://doi.org/10.17616/R3M024>) and FAIRsharing.org (<https://doi.org/10.25504/FAIRsharing.httzv2>) registries. Additionally, it is recommended by a number of domain-relevant journals and publishers to meet data archiving policies, including the British Ecological Society (https://besjournals.onlinelibrary.wiley.com/hub/data_archiving_policy), the Journal of Zoology (<https://zslpublications.onlinelibrary.wiley.com/hub/journal/14697998/author-guidelines>), Oikos (<http://www.oikosjournal.org/DataArchiving>), and Springer Nature (<https://www.springernature.com/de/authors/research-data-policy/repositories-bio/12327160>).

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

ORGANIZATIONAL INFRASTRUCTURE

1. 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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Accpet

Response:

The mission statement for the MDR is available online at <https://www.datarepository.movebank.org/missionstatement> and includes the aim to provide access to and preserve data, as well as an explicit description of how the MDR implements FAIR principles. This mission statement was approved by the director of KIM on 25 March 2021.

In addition, Movebank's goals statement (<https://www.movebank.org/cms/movebank-content/about-movebank#goals>), written in 2011, provides evidence that the goals of preservation ("archive") and accessibility ("promote open access") extend back to the founding of the MDR. This statement was written by the curator and founders of Movebank (Martin Wikelski, Roland Kays, Sarah Davidson) and has been publicly visible at movebank.org and in conference presentation slides since 2012.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

2. 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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Response:

Each data package and file published in the MDR is licensed with a Creative Commons Public Domain (CC0) license (<https://creativecommons.org/publicdomain/zero/1.0>). This requirement is publicly described in the Submission Guidelines (<https://www.datarepository.movebank.org/submissionguidelines>), and the rationale for using the CC0 license is described in the FAQ (https://www.datarepository.movebank.org/faq#how_are_datasets_in_the_repository_licensed). The CC0 licenses are documented in the DataCite metadata for every published object using the properties rightsList and rightsURI, and in the readme files and on the landing pages associated with every published data package (e.g., <https://www.datarepository.movebank.org/handle/10255/move.1067>). The use of CC0 licenses eliminates the need for compliance monitoring since there are no restrictions on data reuse.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

3. 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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Response:

The Communication, Information, Media Centre (KIM) is the central service provider for IT and library services at the University of Konstanz. In this role it is one of the core institutions of the University. As a library one main purpose is to guarantee long-term access to information. The University is permanently state funded resulting in indefinite time of support. Movebank has long-term (>20 years) funding through the Max Planck Society, with additional support through several active grants and partnerships, which secure continuity of the development of Movebank and MDR within our Designated Community over at least the next two decades

(<https://www.movebank.org/cms/movebank-content/about-movebank>). The partnership between the KIM and Movebank demonstrates a commitment to both preserving data holdings and the relevance of the holdings to our Designated Community.

The MDR is part of the KIM's service portfolio. As a memory institution, the continuity of the services, especially repositories and archive systems, is a main goal of the institution. The KIM is the primary custodian of data published by the MDR (<https://www.kim.uni-konstanz.de/en/openscience/research-data-management/repositories-and-databases>). The KIM guarantees continued access to data over a 10-year preservation period, as stated in the Preservation Policy (<https://www.datarepository.movebank.org/preservation#preservationpolicy>), with the intent of preserving the data in perpetuity barring unforeseen circumstances.

In case of an abandonment of service, the MDR is obliged to offer all datasets to the University archive as defined in the preservation policy of the repository (<https://www.datarepository.movebank.org/preservation>). The University archive is part of KIM and functions as the central institution responsible for the preservation and documentation of the history of the University and its archival material. To ensure a complete transfer of the data to the archive, MDR has extensive documentation, including how and by whom the data was stored, how and by whom metadata was added, and how and with which technology the data can be accessed and the repository recovered. If there is a gap in the documentation, this will be closed before archiving.

By archiving with and maintaining comprehensive information about the history and the necessary steps to restore the MDR and its content, the holdings could be moved to another provider, e.g., the MPIAB or another library or university.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

4. 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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Response:

Because all data published by the MDR are made public, access limitations are not possible. Further, data publication is not required by the MDR, University of Konstanz, or MPIAB. Therefore, data owners can choose not to submit or publish their data if they pose a disclosure risk, and do not need to report or justify this decision to the MDR.

References to disciplinary and ethical norms regarding data use are discussed in several places in the MDR FAQ (<https://www.datarepository.movebank.org/faq>; see “How does the repository address disclosure risk?”, “How are datasets in the repository licensed?”, “Can I publish my data but restrict access to it?”, and “How do I cite data from the Movebank Data Repository?”). Ethical implications of publishing in the MDR in comparison to other options for sharing data through Movebank are described at <https://www.movebank.org/cms/movebank-content/archiving-best-practices>.

In most cases, data owners must obtain permits and approved animal care and use protocols in order to collect animal-borne sensor data published by the MDR. This is the standard way to ensure ethical procedures are used to capture animals to attach or remove sensors as well as other attachments, sampling, or measurements completed as part of the research or monitoring program. However, these requirements and relevant authorities vary widely within our international user community, e.g., based on political jurisdiction, research method, species, and the year in which fieldwork was done. Confirmation that the owner has obtained appropriate permits and followed approved protocols is a standard reporting requirement for the universities and agencies through which the research is funded and conducted, and for the peer-reviewed papers, theses and government reports in the field that underlie the data published by the MDR.

Due to the wide variability in local requirements, and given these relevant procedures in earlier stages of the research process, it is not feasible for MDR staff to independently guarantee that there is no disclosure risk as part of the review process. We rely on data depositors to define whether the public disclosure of their data poses a risk to animals. In signing the MDR Depositor Agreement (<https://survey3.gwdg.de/index.php?r=survey/index&sid=912274>), depositors declare they are authorized to make the data public, that they have “considered any relevant risk disclosure and ethical implications of making the data publicly available” and that “data were collected in compliance with any relevant ethical guidelines, including permits and approved animal care and use protocols”.

MDR staff take steps to stay informed of current disciplinary and ethical norms through peer-reviewed literature, blog posts, and presentations, as well as through discussions with our Designated Community (see R5, R6). We offer guidance on deciding whether and how to publish data, both online (<https://www.movebank.org/cms/movebank-content/archiving-best-practices>) and through personal correspondence. The following examples illustrate how potentially sensitive information have been addressed to enable publication in the MDR in compliance with disciplinary and ethical norms:

- There are cases in which data owners have modified data in order to make it public and avoid disclosing sensitive information. For example, an owner could exclude records that identify sensitive breeding sites for an endangered species, or reduce data resolution in order to publish true measurements that document reported results but mask behavioral or other details that pose a potential threat.
- We have heard from our Designated Community that in many cases, the most sensitive information is the current or very recent whereabouts of individuals. Because of the review process, and because we are often publishing data underlying papers following months or years of analysis and peer review, data are no longer recent by the time they are published. We also provide a one-year embargo upon request (<https://www.datarepository.movebank.org/submissionguidelines>).

As discussed in R8, we are considering adding the possibility to deaccession files as part of a planned software upgrade. Such a feature could allow for deaccessioning published files if disclosure risks or ethics violations were identified.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

5. 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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Response:

The MDR is a collaboration between the Communication, Information, Media Centre of the University of Konstanz (KIM) and the Max Planck Institute of Animal Behaviour (MPIAB) and receives long-term funding from both institutions. The University was founded in 1966 and achieved status of excellency in the German Universities Excellence Initiative in 2007, 2012, and 2019. The KIM was formed in 2014 through a merging of the University's library and computing centre in 2014. In 2019, the MPIAB, previously one of two campuses of the Max Planck Institute for Ornithology, expanded to become a new Max Planck Institute within the Max Planck Society (<https://www.mpg.de/en>), one of the world's leading organizations for basic research (https://en.wikipedia.org/wiki/Max_Planck_Society). The MPIAB has close ties to the University of Konstanz, including shared buildings on the University campus, MPIAB department directors that are also professors of the University (e.g., see <https://www.uni-konstanz.de/en/university/news-and-media/current-announcements/press-releases/press-releases-in-detail/Max-Planck-Institut-fuer-Verhaltensbiologie-in-Konstanz-gegruendet-17378>), and a joint doctoral research school (https://www.orn.mpg.de/2453/Short_portrait). The strength of these host institutions and the connections between them help ensure the stability and sustainability of the MDR.

The MDR staff consists of several positions, totalling on average 3–4 individuals and 1–2 FTE positions. This includes a data curator and curation and support specialist from MPIAB and a developer from KIM. From June 2019 to May 2023, another FTE position is being financed in the context of the ongoing project "Movebank 2.0"

(<https://www.kim.uni-konstanz.de/en/the-kim/about-kim/projects/aktuelle-projekte/movebank-20/>) which aims to further enhance components of Movebank including the MDR. As a service of KIM, the MDR is part of the department for IT services research and teaching. The following part- and full-time positions are currently working on the MDR:

- product owner (KIM): the product owner is the department leader of IT services research and teaching and is responsible for the service (permanent position).
- data curator and curation and support specialist (MPIAB): funded through grants when available, otherwise through structural MPIAB funding (permanent position).

- two software developers (KIM): funded through structural and excellency funding at the University of Konstanz (permanent/temporary position).
- information scientist/librarian (KIM): funded through state project funds (temporary position).

In addition to structural funding through KIM and MPIAB, grant funding that directly supports the MDR includes

- 2010–2012: German Research Foundation (DFG) AOBJ 576687 (“MoveVRE” project)
- 2019–2021: European Union’s Horizon 2020 H2020-INFRAEOSC-2018-2020 Grant agreement 831558 (“FAIRsFAIR CoreTrustSeal Certification Support”)
- 2019–2023: Baden-Württemberg Ministry for Science, Research and Art (“Movebank 2.0” project)

As institutions employing hundreds of people, KIM and MPIAB provide additional staff and resources that support the MDR and MDR staff. This includes other staff working directly on KIM data repositories and the Movebank platform, as well as resources like human resources (e.g., hiring, benefits), IT (e.g., technical support), and facilities (e.g., office space, cafeteria).

The MDR staff have access to training and workshops, and also participate in conferences and meetings to engage in knowledge transfer and discussion with peers and experts. Events attended by MDR staff include workshops on the following topics:

- research data management
- software development
- domain-relevant research
- technology developments (e.g., movement ecology research and wildlife management),
- repository certification
- metadata
- persistent identifiers
- long-term preservation
- open science

Further, the staff at KIM work in an interdisciplinary team on the topic of Open Science

(<https://www.kim.uni-konstanz.de/en/openscience>). This enables discussion and support from a broad spectrum of perspectives and backgrounds. In addition to the MDR, KIM also hosts KOPS

(<https://kops.uni-konstanz.de/?locale-attribute=en>), an institutional publication repository with a high percentage of open access publications, providing additional colleagues with whom to exchange knowledge on general repository topics. As

part of the University of Konstanz, KIM staff also can partake in the academic staff development program

(<https://www.uni-konstanz.de/en/asd/profile/academic-staff-development>) of the university to acquire relevant skills

needed. The staff at MPIAB have access to multiple weekly seminar series that provide updates on new research and applications within the Designated Community, as well as career development programming

(<https://www.ab.mpg.de/15594/karriereentwicklung>).

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

6. 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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Response:

MDR staff have access to experts through the MPIAB, the University of Konstanz Departments of Biology and the Centre for the Advanced Study of Collective Behaviour, the Team Open Science at KIM, data depositors, and the Designated Community. The MPIAB is a world leader in movement ecology research and in the development and use of state-of-the-art sensor technology. Participation of MDR staff within other work of the KIM and MPIAB leads to ongoing discussion and interaction on topics such as movement ecology research, open access, research data management, and high-performance computing, which benefits the development of the MDR. Communication with experts, the Designated Community, and MDR staff in a variety of formal and informal ways. The MDR does not currently have a formal advisory board, but we are reviewing options and examples for establishing one in the future. To date, public forms of communication include:

- Presentations and workshops where Movebank and the MDR are presented, discussed, and used directly with MDR staff and our Designated Community (see R5).
- Public communications with or Designated Community, including annual Movebank newsletters that always include a

section about the MDR (e.g., <https://www.movebank.org/cms/movebank-content/february-2021-newsletter>) and publicly-announced user surveys (e.g., <https://ling-survey.uni-konstanz.de/sosci/movebank>). Survey results are used internally and are not public for data privacy reasons.

- Communications with relevant publishers and journal editors. For example, correspondence with Springer Nature led to workflow and metadata improvements and MDR's subsequent listing in the publisher's list of recommended repositories (<https://www.springernature.com/gp/authors/research-data-policy/repositories/12327124>).
- Communications within and beyond our Designated Community through Twitter (<https://twitter.com/movebankteam>) and Facebook (<https://www.facebook.com/Movebank>). Upon release of new MDR data packages, our final letter to data depositors includes an offer to spread the word about their work through social media (e.g., <https://twitter.com/MovebankTeam/status/1239615982877978626>), and members of our Designated Community have independently communicated about the MDR's work (e.g. <https://twitter.com/CraigTaylorViz/status/1129325284971429888>).

Other forms of communication are ongoing and occur daily and over years, through email communication, meetings, and participation in relevant initiatives. Public evidence of these relationships includes output from collaborations and is provided below where available.

- Daily support and curation for our Designated Community and data owners through email, phone, and in person. On average, the MDR publishes approximately one data package per week, each of which involves communicating directly with data owners about their data and reading one or more written manuscripts or publications describing their research (<https://www.datarepository.movebank.org/submissionguidelines>).
- Daily communications with KIM and MDR staff. For example, the development and maintenance of the MDR is coordinated and technically aligned with other repositories developed by KIM, increasing the efficiency of development and infrastructure use. This is planned to extend even further by updating the MDR repository software to a product used for other repositories of KIM during 2021.
- Daily communications with MDR staff, data depositors, engineers developing animal-borne tag technology, and Movebank developers. To serve the Designated Community, Movebank must quickly accommodate new data attributes and tag technology and growing data volumes. Because the Movebank database serves as the data submission portal for the MDR, these efforts help ensure that the MDR can continue to publish data relevant to our Designated Community even within the context of fast-developing technology (<https://www.movebank.org/cms/movebank-content/february-2021-newsletter>).
- Long-term partnerships between MDR staff and the Designated Community that have built and maintained support for specific data types and user groups, and development of technology and methods (<https://www.movebank.org/cms/movebank-content/partner-institutions>). For example,
 1. Ongoing communication with leading researchers in geolocation (i.e., the derivation of animal movements and behavior from light-level measurements recorded on animal-borne loggers) beginning with a workshop in 2012 (<https://www.nceas.ucsb.edu/workinggroups/establishing-open-source-animal-tracking-analysis-platform-archival-geolocators>) has led to methods and guidance for publishing these data in the MDR (<https://geolocationmanual.vogelwarte.ch>, see Chapter 9; Lisovski et al. (2019) <https://doi.org/10.1111/1365-2656.13036>). This effort has led to publication of many geolocation datasets by the MDR (<https://www.datarepository.movebank.org/discover?query=geocator>).
 2. The Arctic Animal Movement Archive

(<https://www.movebank.org/cms/movebank-content/arctic-animal-movement-archive>), a data collection hosted on Movebank and highlighted in Science Magazine (<https://doi.org/10.1126/science.abb7080>), involved close communications with several hundred data owners beginning in 2015 (https://above.nasa.gov/cgi-bin/inv_pgp.pl?pgid=3391), collection of feedback through a written (non-public) participant survey describing their requirements for maintaining long-term access to data, and the curation of several hundred datasets in Movebank, several of which were published in the MDR as a result (e.g., <https://doi.org/10.5441/001/1.p5bn656k>).

- Participation of MDR staff in societies and initiatives relevant to the Designated Community, including the International Bio-Logging Society, the leading global organization for animal-borne sensor data (<https://doi.org/10.3897/biss.3.38919>; <https://www.bio-logging.net/#news>; Rutz et al. 2019, <https://doi.org/10.1038/s41559-020-1237-z>); the Machine Observations Interest Group of TDWG, the leading global source of data standards for biodiversity (<https://www.tdwg.org/community/mobs>); the Data Policy Standardization and Implementation Interest Group of the Research Data Alliance (<https://www.rd-alliance.org/node/53666/members>).

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

DIGITAL OBJECT MANAGEMENT

7. 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
Accept

Reviewer 2

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

Response:

Several core features of the MDR are key to understanding how data integrity and authenticity are treated in the MDR:

- Data depositors prepare data for submission through the Movebank platform and formally initiate a submission to the MDR by submitting a Depositor Agreement. These data undergo a review by MDR staff prior to acceptance and deposition in the MDR (<https://www.datarepository.movebank.org/submissionguidelines>). This process is required for all MDR data depositors but is not a general requirement for the use of Movebank itself (see R0). The review involves extensive data checks, edits, addition of descriptive data and metadata, and communications between MDR staff and the data owner (see R8, R12).
- Upon acceptance and publication, each data package stored in the MDR consists of static files that are uploaded by MDR staff and reflect the results of the review and QC processes. These cannot be changed or updated by the depositor.
- MDR staff collect a required set of metadata from the data depositor, and subsequently publish and maintain these metadata at DataCite, following DataCite's metadata schema (<https://schema.datacite.org>). The same metadata properties are published for every published data package (e.g., <https://search.datacite.org/works/10.5441/001/1.j484vk24>) and file within a data package (<https://search.datacite.org/works/10.5441/001/1.j484vk24/1>) (see R8, R14).
- With a future software update in progress (see R15), data depositors will initiate new submissions through the repository directly, allowing them to directly enter some of the required metadata. All metadata will be checked and updated or completed as needed by MDR staff prior to publication, and this change will not replace the established curation and review workflows.

Data integrity in the MDR is ensured through checksums, data review and curation, metadata management protocols, and a versioning system that is currently being updated from a "manual" process to a core function of the repository software. To verify the integrity of published data files, MD5 checksums are generated and used if needed. As of yet this check is not done on a regular basis. This is planned to turn into a regular workflow when updating the software of the MDR during 2021. Through the submission and review process (see above), MDR staff ensure completeness of published data files. The required metadata fields must be provided before a data package can be published, i.e., the software does not allow the "accept" button to be pushed until these fields are completed. Once a data package is accepted, metadata are published by the MDR repository software as part of the data release. MDR staff, rather than data depositors, upload final data files and approve final metadata within the MDR prior to publishing data packages, further assuring completeness (see above).

Metadata are published following the internationally recognized DataCite schema (<https://schema.datacite.org>) (see R8). When needed, MDR staff make updates to published metadata at DataCite DOI Fabrica (<https://doi.datacite.org>). This service manages metadata versioning and documents the time and date that metadata are published and updated (see figure in appendix, "310321_CTS_MDR_appendix.pdf"). Currently, published data files cannot be changed or removed from the MDR, so there is not a process for logging data changes. Instead, a published data package can be superseded by publishing a new data package with updated files, which follows the same curation and documentation process as for other data packages. This is done in cooperation with the data author, and we indicate relationships between data

packages and files in the DataCite metadata (property relatedIdentifiers) and the readme files for the new data package. The MDR does not currently support an automated versioning system, but a change of the underlying software from Dryad to DSpace 7 is planned in 2021 (see R15), and MDR staff have identified versioning as a requirement for this upgrade. With the planned new versioning system, this process of cooperative updates is expected to continue, with improvements to automate documentation steps (e.g., the publication of related metadata updates) and make versioning more clear and usable (e.g., allowing citation of a specific version or all versions of a data package).

Data authenticity in the MDR is ensured through metadata maintenance and the data review. The MDR data review workflow includes extensive checks of the reliability of data attributes and values, using both automated and manual processes (see R8). Data files are downloaded from Movebank as the submission platform and uploaded to the MDR by MDR staff after the review is complete. Because data files published by the MDR cannot be updated or removed, there is currently no strategy for data changes or comparing file versions. Data owners who wish to deposit data must complete a Depositor Agreement (<https://survey3.gwdg.de/index.php?r=survey/index&sid=912274>), provide a published document or manuscript describing the data and an author list for the data, and confirm that they have “the authority to submit data for publication on behalf of co-owning individuals and institutions”. The identity of the depositor is subsequently verified through direct communications with MDR staff and review of the related written document.

Required metadata, combined with the use of persistent identifiers, define relationships between data sets and metadata (see R8, R13). These metadata require provenance information including the DataCite properties contributor (the host repository), creator (authors), publisher (the host repository), dates (the year issued), relatedIdentifier (defines relationships between data files, data packages, and other relevant resources including written publications), and a description (includes a citation for a paper or report that further describes the data and typically includes provenance information, i.e. how the data were collected). Additional provenance metadata is also automatically collected and stored within the repository using Dublin Core elements that record the date an package or file is accessioned, available, and issued, and a provenance description using checksums (e.g., <https://www.datarepository.movebank.org/metadata/handle/10255/move.1068/mets.xml>). Further, every published data file has an associated readme file, a text file containing the same provenance information that is provided in the metadata (e.g., www.datarepository.movebank.org/bitstream/handle/10255/move.1068/README.txt?sequence=1). This serves as important redundancy in documenting data authenticity because it does not require access to the internet, is human-readable, and can be easily downloaded, stored locally, and included if data files are transferred to other people or storage locations. The MDR has updated published metadata, most commonly updating relatedIdentifier to define additional relationships, for example when additional documents describing a data package are published, or when a data package is superseded.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

8. 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:

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Response:

As described in R0, the MDR is closely related to the Movebank data platform, providing a formal archiving and publication service for data initially imported to Movebank (<https://www.movebank.org/cms/movebank-content/data-repository>). This connection, and how it defines the scope of and changes to our collection development, is described in our Submission Guidelines (<https://www.datarepository.movebank.org/submissionguidelines>). We also offer guidance to help data owners decide the scope of a dataset to submit for publication in the MDR (see “What should I archive?” at <https://www.movebank.org/cms/movebank-content/archiving-best-practices>).

Further details about the processes described below are provided in the Submission Guidelines. Before data are published in the MDR they are imported to Movebank, which involves checks using database rules and transformation of all data attributes and values to a harmonized and publicly documented data model (<https://www.movebank.org/cms/movebank-content/mb-data-model>) and vocabulary (<https://www.movebank.org/cms/movebank-content/movebank-attribute-dictionary>). After import to Movebank, data can be checked by owners for errors and supplemented by flagging outliers or adding documentation, according to best practices for data organization and quality control (<https://www.movebank.org/cms/movebank-content/upload-qc>). If a data owner wishes to deposit their imported data for publication in the MDR, they must submit the imported data for review by MDR staff. No independent upload by depositors to the MDR is possible, resulting in every published dataset being intellectually verified by MDR staff. During the review process, MDR staff and depositors work together to make corrections, additions or changes based on these checks. Checks involve both data quality (e.g., see <https://www.movebank.org/cms/movebank-content/upload-qc>) and completeness

(https://www.movebank.org/cms/movebank-content/archiving-best-practices#general_best_practice_tips). We work with the data owner to add this and other information relevant to the study, based on the associated paper or report, expert knowledge of species and methods used, and what data were recorded. Where additional information outside of the Movebank data format will improve understanding of the data and related analysis results, this can be included in additional data files in open formats (e.g., a script used to derive location estimates based on sensor readings). Following the review and curation steps described above, MDR staff prepare files for upload to the MDR and subsequent publication as described in R0 and R14.

Metadata associated with published data packages are entered or enriched by MDR staff directly into the MDR or are created automatically by the repository software. Data packages can only be published after all metadata have been entered. Information is obtained through the Depositor Agreement (<https://survey3.gwdg.de/index.php?r=survey/index&sid=912274>), communication with the depositor, and review of the documents describing the data. The required metadata describe the data owners (authors/creators), publisher (MDR, University of Konstanz), year of publication, descriptive details (keywords, abstract), data license (CC0), provenance information, and relationships between files within the data package and other related documentation and resources (also see R7). As part of a planned software upgrade (see R15), users will initiate submissions within the repository by completing the Depositor Agreement and providing depositor-defined metadata properties (e.g., author list). This will not override established review and curation processes as MDR staff will still approve every submitted data set before publication.

See the following examples of metadata published at DataCite and within the MDR:

DataCite:

- Data package: <https://search.datacite.org/works/10.5441/001/1.j484vk24>
- Data file within a package: <https://search.datacite.org/works/10.5441/001/1.j484vk24/1>

MDR:

- Data package: <https://www.datarepository.movebank.org/metadata/handle/10255/move.1067/mets.xml>
- Data file within a package: <https://www.datarepository.movebank.org/metadata/handle/10255/move.1068/mets.xml>

Metadata can be updated by MDR staff following publication in the MDR, but published data files cannot be changed or removed from the MDR. Instead, a published data package can be superseded by publishing a new data package with updated files (see the Submission Guidelines, <https://www.datarepository.movebank.org/submissionguidelines>), and we are preparing a repository software upgrade that will include the addition of an automated versioning system. These processes are described further in R7. As part of this upgrade, we will evaluate adding a feature to support data file deaccession. In rare cases, where removal of the data files might be necessary, the files can be removed without deleting the corresponding metadata. This will support persistent discovery, understandability, and proof of the data, even if they are no longer available.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

9. 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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Response:

The MDR uses the Datacentre infrastructure of the University of Konstanz. The Datacentre is a sub-department of the KIM and is responsible for maintenance and development of the central server infrastructure of the University. This includes the management of the central virtual server and the storage infrastructure (see “Sachgebiet Datacenter” (<https://www.kim.uni-konstanz.de/das-kim/kontakt-und-oeffnungszeiten/abteilungen/it-infrastruktur>)). The MDR software runs on a virtual machine (VM) in the Datacentre. KIM uses the platform “VMware vSphere” to create and manage VMs. Published data files and corresponding metadata of the MDR are stored on the virtual storage drives on the VM itself. The VM is mirrored in two different physical locations on campus. During regular operation only one copy is used. If the working copy fails in case of emergency, the synchronized copy replaces the original version and ensures the seamless continuity and recovery of the service. A further backup layer can be added if needed, using the backup service of KIM (<https://www.kim.uni-konstanz.de/en/services/administration-and-maintenance/backup/general-information/>). With the software “Bacula” an additional copy of the VM can be stored and recovered at an offsite location on a regular basis. This is currently not the case for the MDR, but could be implemented if needed in the future.

To prevent data loss in case of an emergency, regular back-ups of the VM the repository runs on, including the software

and the published data, are carried out automatically. This is accomplished by using the snapshot feature of the VMware vSphere software and the software “Veeam Backup & Replication”. A complete backup of the VM is generated and stored once a week, with additional incremental backups in the days between. The VM and the data are always stored and therefore recoverable in 30 states, that have been snapshot. Furthermore “Veeam Backup & Replication” checks the integrity of the stored snapshots. This whole process of VM management runs automatically once the VM is initialized in the Datacenter.

The storage system of the Datacenter reads out the S.M.A.R.T. data of the used storage drives. This enables a monitoring of the storage hardware and leads to action in case of emergency. RAID5 is used to ensure ongoing service while defect hardware is fixed or replaced. All described processes are thoroughly documented internally in the respective department of the Datacentre.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

10. 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
Accept

Reviewer 2

Comments:
Accept

Response:

To formalize and document MDRs approach to long-term preservation, a preservation policy was created and published in 2021 (<https://www.datarepository.movebank.org/preservation#preservationpolicy>). As the repository currently holds a relatively homogenous collection of data, all data sets are treated equal in terms of long-term preservation.

Data owners who are considering depositing data are informed about the review and publication processes, including the use of the CC0 license and persistent identifiers, through the Submission Guidelines (<https://www.datarepository.movebank.org/submissionguidelines>), FAQ (<https://www.datarepository.movebank.org/faq>), and the MDR description at <https://www.movebank.org/cms/movebank-content/data-repository>. The submission and review are not initiated until the author has confirmed that they want to proceed by providing information needed for the review and accepting the terms of use, including the CC0 license, through a Depositor Agreement (<https://survey3.gwdg.de/index.php?r=survey/index&sid=912274>).

Repository rights are clarified by the use of the CC0 license, which puts data published by the MDR in the public domain (see R2). Thus, once the author has agreed to deposit data under this license, the MDR has the right to store, copy, transform, and distribute it following publication. To ensure long-term access and reusability only open formats are accepted in the repository (see the Submission Guidelines, <https://www.datarepository.movebank.org/submissionguidelines>). MDR staff are aware of the possible need to transform file formats in the future to keep the data accessible, and this will be evaluated if and when it becomes necessary.

The MDR aims for conformity towards the OAIS reference model. A high-level mapping can be found on the repository website (<https://www.datarepository.movebank.org/preservation#oais>).

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

11. 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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Response:

As a domain-specific repository that is closely linked with a larger platform (Movebank) that is developed with ongoing feedback from our Designated Community, and as part of a larger team of data experts at KIM, MDR staff have and maintain relevant technical and domain-specific expertise (see R0, R5, R6). Further, our Designated Community includes both data depositors and end users, with MDR staff obtaining regular feedback from both groups (see below). Because MDR staff review and curate data prior to accepting data for publication and finalizing and depositing data files in the MDR, data quality assurance overlaps largely with our response to R8. See R7, R8, R12, and the Submission Guidelines for further description of our approach to evaluating, ensuring, and documenting data quality and completeness.

All published data packages must relate to at least one written work describing the dataset (see R0), providing a valuable resource for potential users of published data. This primary related work is referred to in published metadata (see R7, R8), displayed on the landing page for every published data package and file (e.g., <https://doi.org/10.5441/001/1.j484vk24>), and in the readme file that accompanies every published data file. Information about additional related works, including papers or other datasets, can be included in the readme files and added to metadata (for example, see metadata for a published data package augmented with links to several additional related works, using relatedIdentifier type `IsDocumentedBy`, at <https://search.datacite.org/works/10.5441/001/1.sv6335t3>). The use of data packages, i.e., sets of related files and metadata, within the MDR also helps to ensure that potential users understand these resources and their relationships.

We do not currently offer a venue for comments or ratings on data published by the MDR. However, feedback can be provided in the following ways:

- In most cases, the primary written work that describes the published data package is a peer-reviewed journal article, and most non-peer-reviewed works are government reports and theses/dissertations. The process of completing these publications involves feedback and reviews of the study methods and results and serves as a relevant assessment of the published data.
- The mandatory use of DOIs in the MDR, and links to related citations in the DataCite metadata (see R7 and R8), allow for both the data packages and related works to be included in citation indices and other use metrics.
- Data are often shared and used following import into Movebank and prior to submission for publication in the MDR. For example, the owner might make the unpublished dataset available to reviewers as part of the review of a related manuscript, or share it for use in a regional conservation assessment while data are still being collected. These options

provide opportunities for others to contribute to assessing the data and often lead to corrections (e.g., the reviewers notice data for some animals are missing) or the addition of more information (the conservation organization requests that the owner add the animals' age and sex to the dataset for use in their assessment). Therefore, by the time data are being reviewed by MDR staff for publication, informal feedback and QA/QC has often already contributed to the quality and usefulness of the data.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

12. Workflows

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

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase
Accept

Reviewer 2

Comments:
3 – The repository is in the implementation phase
Accept

Response:

The workflows for submission, review, communication with data authors, and publication and post-publication activities are described in our Depositor Agreement (<https://survey3.gwdg.de/index.php?r=survey/index&sid=912274>) and Submission Guidelines (<https://www.datarepository.movebank.org/submissionguidelines>) and resources linked within the guidelines. These include:

- timeline and steps for the review, coordination with manuscript reviews and provision of a preliminary DOI, data release, and embargoes (also see R10),

- guidance on selecting data for publication and allowed data formats (also see R0 and R8),
- instructions for data appraisal and QA/QC (also see R8), and
- preparing final files for upload to the MDR and publication.

MDR staff use a Redmine instance hosted by KIM to manage internal communications, plan tasks, and document progress related to MDR development. This includes a Wiki and a ticketing system. Access to this knowledge base is restricted to authorized personnel using Redmine's authentication system. The use of Redmine also ensures that relevant discussions, decisions, draft documents, meeting minutes, and other non-public resources related to MDR development are archived and can be securely made available to other KIM and MDR staff. Redmine will also be used to plan and implement the upgrade of the repository software in 2021.

As all data published by the MDR are licensed under CC0 (see R2), security of published data is not a concern and no access restrictions are in place for reading and downloading data. Therefore, there are currently no user accounts available for the MDR except for those used by MDR staff. File upload, administration of submissions, and metadata management are completed only by MDR staff using secure accounts (see R7). This will change with the planned repository software upgrade (see R15). Following this upgrade, data depositors will be able to register with the repository, initiate submissions, and provide select metadata properties that are defined by the depository and available upon submission. This will not change the established data curation and review workflows, and MDR staff will remain responsible for ensuring that metadata are complete and correct. During the review of the data by MDR staff, data are stored, assessed, and accessed from within the data owner's related study on Movebank, which will have public or private access as controlled by the data owner. For a discussion of how we address cases of sensitive data (e.g., endangered species), see R2.

These procedures have remained largely consistent since the MDR was established in 2011. Significant changes have involved written communications and joint planning and decision-making with the KIM and MPIAB directors and MDR staff. Such changes include publication of the Mission Statement (<https://www.datarepository.movebank.org/missionstatement>), Preservation Policy (<https://www.datarepository.movebank.org/preservation#preservationpolicy>), Submission Guidelines (<https://www.datarepository.movebank.org/submissionguidelines>); implementation of a Depositor Agreement (<https://survey3.gwdg.de/index.php?r=survey/index&sid=912274>); and additions to the metadata and documentation to be included in the Web of Science Data Citation Index and Springer Nature's list of recommended repositories (see R0). In these cases, we took the opportunity to discuss and make decisions based on how changes requested by the third party could be implemented technically, and how they could be used to improve the overall quality, discoverability, and understanding of data published by the MDR.

To formalize and increase transparency of the workflows, BPNM-designed processes were created using the University of Konstanz process portal (<https://www.uni-konstanz.de/en/quality-management/process-management-and-organisation-process-portal>) and those directly relevant for data owners are posted on the repository website to increase transparency.

- Data review procedures (https://www.datarepository.movebank.org/faq#how_do_data_in_movebanks_data_repository_

differ_from_other_data_in_movebank)

- Submission process (https://www.datarepository.movebank.org/faq#how_do_i_prepare_and_submit_my_data)

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

13. 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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Response:

A recommended citation for each data package published by the MDR is provided on the landing page for each data package and file (e.g., <https://doi.org/10.5441/001/1.j484vk24>) offers users a recommended citation for the data package as well as a citation for the related publication based on the data. This information is also included in the readme file for each data file (e.g., <https://www.datarepository.movebank.org/bitstream/handle/10255/move.1068/README.txt?sequence=1>). The information needed to create a citation (authors, publication year, title, publisher) is also included in the metadata for every published data package and file.

To enable persistent identification, all published data packages and files have DOIs (prefix 10.5441/001/1) issued by KIM,

which is a registered data centre as part of the German National Library of Science and Technology (TIB Hannover) DOI Consortium. This enables the persistent and globally unique identification of every published data package and file in the MDR. The relationships between data files for each dataset are maintained by publishing data packages that share a DOI, with extensions to that DOI for each individual file. The DataCite metadata for each data package and file includes the DOIs of all objects in the package, the DOI for one or more related papers, authors of the data package, repository name, and other information needed for humans and machines to identify and cite the work (see R7, R8).

Published data packages and files are discoverable in a variety of ways that leverage the MDR's use of metadata standards, DOIs, and recommended citations. These discovery options target a variety of audiences such as experts in movement ecology, wildlife management, biodiversity and conservation, as well as researchers in general and software developers interested in data visualization and machine learning. We publish metadata both within the repository (using DataCite and Dublin Core properties) and at DataCite, thereby complying with widely-used metadata standards and enabling machine harvesting (see R7, R8).

- The MDR website includes options to browse data by searching metadata (<https://www.datarepository.movebank.org/discover>) and by author name (<https://www.datarepository.movebank.org/browse?type=author>).
- Metadata are machine-discoverable via the DataCite APIs (example: <https://search.datacite.org/works/10.5441/001/1.j484vk24>). This also leads to the data being discoverable via services that use the DataCite APIs, such as Google's Dataset Search ([https://datasetsearch.research.google.com/search?query=Movebank Data Repository](https://datasetsearch.research.google.com/search?query=Movebank+Data+Repository)).
- Data packages are discoverable through the Web of Science Data Citation Index (<https://clarivate.com/webofsciencelibrary/master-data-repository-list>).
- MDR data can be harvested via an OAI-PMH interface via <https://www.datarepository.movebank.org/oai/request>.
- Published data are accessible from the R package "move", a popular analysis package for movement ecologists (<https://cran.r-project.org/web/packages/move/vignettes/browseMovebank.html#download-data-from-the-movebank-data-repository-as-a-movemovestack-object>).
- DOIs and citations for published data packages are included in the "Study Details" of the related study on Movebank (https://www.movebank.org/cms/webapp?gwt_fragment=page=studies,path=study208413731) to improve discoverability in searches on this platform.

Finally, as described in R0, the MDR is included in several data and repository discovery services and recommended for use by authors publishing in domain-relevant journals, which helps to help to improve discovery of the MDR in general.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

14. 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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Response:

The mission of the MDR (<https://www.datarepository.movebank.org/missionstatement>) explicitly includes the support of data reuse. By supporting meaningful, legal, and efficient data reuse, we allow others not only to reproduce published results, but also to complete reanalyses and metaanalyses that combine datasets. This requires adequate curation and documentation (see R7, R8, R11, R12), data licenses that explicitly allow reuse and modification (see R2), and the use of a standard vocabulary and file formats (R0, R8, R14) that make the data interoperable, i.e., end users can develop software applications that automatically read published data (e.g., <https://cran.r-project.org/web/packages/move/vignettes/browseMovebank.html#download-data-from-the-movebank-data-repository-as-a-movemovestack-object>).

Metadata for data packages and files follow the DataCite (<https://schema.datacite.org>) and Dublin Core (<https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>) schema (for examples and properties used, see R7, R8). These metadata identify the DOI (property identifier) and data authors (creators), link to relevant written works (descriptions, relatedIdentifiers), specify the CC0 license (rightsList), provide keywords to enable discovery (subjects), define relationships between objects in the data package (relatedIdentifiers), and record information about data provenance and checksums. Where relevant and as possible, we also include information about relationships between data packages (relatedIdentifiers).

The core files of every published data package are tabular text/csv files with contents that use a published, persistent, machine-readable vocabulary (see R0, <http://vocab.nerc.ac.uk/collection/MVB/current>, Submission Guidelines,

<https://www.datarepository.movebank.org/submissionguidelines>). This vocabulary has grown over time with the development of Movebank and through ongoing communications with our Designated Community (see R0, R6). The vocabulary is currently used to describe over 2.5 billion animal occurrence observations stored in Movebank, including over 100 million animal occurrence observations published in the MDR, and is widely used within the Designated Community. The vocabulary is managed to balance consistency (e.g., use of controlled lists for attribute values when possible, no option for users to create attributes on the fly) and flexibility (e.g., use of manufacturer-specific variables when needed, and inclusion of generic fields to encourage thorough documentation and allow inclusion of dataset-specific information). If attribute definitions change, we ensure that existing values will not be made incorrect by the change. Attributes are rarely deprecated, and the definitions of deprecated attributes remain documented in the published vocabulary, which includes versioning, and in relevant readme files.

In addition, we support data owners by allowing publication of additional files in open formats relevant to the data and analysis in the MDR, and aligned with the published vocabulary where relevant/feasible (see Submission Guidelines, <https://www.datarepository.movebank.org/submissionguidelines>). To further ensure understandability of the data, every data file has an associated readme file that provides a local, human-readable documentation of information stored in the published data and metadata (see R0, R7).

The use of a managed vocabulary and open text data files, and connection to a widely used research platform within the domain, help to ensure the data remain understandable over time and minimize future needs for an evolution of data formats. The text/csv file format of published files has remained the same since the MDR was established in 2011 and we have no plans to update it. If this format were to change in the future to some other storage type, we would ensure that the vocabulary would remain valid, and would make changes to documentation as needed to describe the new format and develop a plan to either update or continue to document existing published data.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

TECHNOLOGY

15. 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:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase

Accept

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Accept

Response:

The MDR technical infrastructure is hosted by KIM (see R9). The MDR runs on a Virtual Machine (VM) that is part of the KIM Datacentre. The VM on which MDR is run uses Ubuntu Linux 18.04-LTS as its operating system, resulting in security updates at least until 2023. Operating system updates are executed regularly. The MDR uses the Data Dryad repository software (<https://github.com/datadryad/dryad-repo>), which is an open-source fork of DSpace written in Java and mainly uses the standard features of said software. The software contains standards such as an OAI-PMH interface and basic documentation for running and maintaining the software is available internally that was created during the initial start of the repository in 2011 and has been expanded since. This enables the developer of the repository at KIM to maintain running the service and implementing new features as needed.

During 2020, the KIM and MDR staff evaluated migration to a new repository software with a self-designed criteria catalogue that was also shared and discussed with other actors in the field of research data management from other institutions in the state of Baden-Württemberg. A software upgrade is pursued to enable the implementation of new features, enable more interoperability, and simplify maintenance and improve sustainability of the system. The evaluation led to plans for using DSpace 7 as the new software for the system. Works on the software upgrade will start early 2021. Part of this upgrade will include further documentation of the new system including a software inventory. The migration to the new software is planned to be finished during 2021.

To ensure data integrity and quality control, the MDR publishes only static text/csv files and thus does not support real-time or near-real-time data streams. The MDR does not have any legal, regulatory, or administrative requirement to provide uninterrupted access to data. Our current system has been sufficient to support public access to MDR data as bandwidth and time access are ensured by the University of Konstanz being part of different national and regional infrastructure projects such as the German National Research and Education Network (<https://www.dfn.de/en/>) and BelWü (<https://www.belwue.de/>).

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

16. Security

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

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase
Accept

Reviewer 2

Comments:
3 – The repository is in the implementation phase
Accept

Response:

A general description of KIM's IT security and privacy systems is available at <https://www.kim.uni-konstanz.de/en/email-and-internet/it-security-and-privacy>. Regarding security personnel, the University of Konstanz employs a team for IT security. A security officer for information security is part of this team. Tasks of this officer include working on security guidelines, coordinating information security processes, and instructing for the implementation of security measures. This includes the possibility of running OpenVAS (<https://www.openvas.org/>) tests on IT systems if needed. An ongoing information security concept is being worked on in close cooperation between MDR staff and the IT security officer. It is planned to finish a first version in the following months and to keep it updated from this point on.

To prevent damages from system failure and human the VM of the repository runs on backed-up hardware (see R9) so the VM and the content of the MDR can be restored if needed. To prevent attacks against the service, the MDR runs in

protected networks of the University infrastructure. This guards the system from unauthorized access from outside such as DDoS, and only allows ssh access from defined IP ranges, preventing possible hacking attacks. Furthermore, unused ports are blocked. As all data on MDR is public data, temporary downtime of the website is the only identifiable risk, which is classified as low. The University recommends best practices for secure passwords

(<https://www.kim.uni-konstanz.de/e-mail-und-internet/it-sicherheit/passwoerter-und-authentifizierung/passwoerter/>). The only people who log in to the MDR using accounts are MDR staff, who are trained and informed by the KIM about password and account security, and they are obligated to these rules by their work contracts.

Regarding physical security, the MDR is hosted and run on the centralized storage and server infrastructure of the University of Konstanz as part of KIM's Datacentre (also see R9 and R15). A concept controlling physical access to the server room is available internally. Regarding hardware security, to prevent data loss by hardware deterioration SMART stats of storage units is monitored and RAID-5 technologies are used for the servers hosting the VM (see R9). The VMs are backed up regularly on two different locations on campus and the back-up can be extended to another offsite location (see R9).

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 CoreTrustSeal certification procedure. Any comments on the quality of the Requirements, their relevance to your organization, or any other contribution, will be considered as part of future iterations.

Response:

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments: