

OPEN SCIENCE

RESEARCH DATA MANAGEMENT PLAN (DMP)¹

Project number	
Funder	
Project title	
Principal investigator (PI)	
PI ORCID identifier	
Beneficiary institute	
Project starting date	
Project end date	
Project duration	
Data manager/contact	
DMP 1st version/date	
DMP last update²/date	

SUMMARY (*dataset³ reference and name; origin and expected size of the data generated/collected; data types and formats*)

Provide a summary of the data addressing the following aspects:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

Dataset reference and name:

- Dataset reference: KUT-PRJ-2024-05
- Name: Urban Wildlife Monitoring Data

Purpose of the data collection/generation:

The purpose of the data collection is to monitor and analyze the presence and behavior of wildlife in urban areas. This data will help in understanding how urbanization impacts local ecosystems and wildlife movements.

Relation to the objectives of the project:

The primary objective of the project is to assess the biodiversity within urban settings and identify key factors that influence wildlife activities. The collected data will directly support this by providing empirical evidence

¹ Template for the Open Science Research Data Management Plan (DMP). The sections should describe how you plan to make the project data Findable, Accessible, Interoperable and Reusable (FAIR).

² DMP is to be regularly updated.

³ Several datasets may be included into a single DMP.

of species presence, movement patterns, and habitat usage within the city.

Data types and formats:

- Data types:
 - Video recordings of wildlife
 - Photographs
 - Structured observation reports
 - GPS coordinates and timestamp data
 - Species identification metadata
- Formats:
 - Video files: MP4, AVI
 - Image files: JPEG, PNG
 - Text files: CSV, JSON
 - Geospatial data: GeoJSON, KML

Existing data re-use:

No existing data is being re-used. All data collected for this project will be newly generated.

Origin of the data:

The data will originate from:

- Field observations conducted by research teams.
- Motion-activated cameras placed in strategic locations across the city.
- Reports and sightings submitted by the public through a dedicated mobile app.

Expected size of the data:

The expected size of the data is estimated to be around 500 GB per year. This includes high-definition video and images, as well as text and geospatial data.

Data utility:

The collected data will be useful to:

- Urban ecologists and conservation biologists studying the impacts of urbanization on wildlife.
- City planners and policymakers aiming to create wildlife-friendly urban environments.
- Educational institutions for research and teaching purposes.
- Community organizations and citizen scientists interested in local biodiversity and conservation efforts.

Each of the following six issues should be addressed with a level of detail appropriate to the project. Some guiding expressions with explaining guidance help in elaboration. Please note that not all guiding thoughts are to be taken into consideration, depending on the project.

1. MAKING DATA FINDABLE (dataset description: metadata, persistent and unique identifiers e.g., DOI)

Making data findable, including provisions for metadata:

Outline the discoverability of data (metadata provision)

Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

Outline naming conventions used

Outline the approach towards search keyword

Outline the approach for clear versioning

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Guidance:

The Research Data Alliance provides a [Metadata Standards Directory](#) that can be searched for discipline-specific standards and associated tools.

Metadata provision:

- We will ensure that the dataset is easily discoverable by creating comprehensive metadata for each data entry. The metadata will include information such as the date and time of data collection, the geographical location (latitude and longitude), species observed, type of data (video, image, text), and any relevant observational notes.

Identifiability of data and standard identification mechanism:

- To enhance identifiability, we will utilize persistent and unique identifiers for each dataset entry. We will assign Digital Object Identifiers (DOIs) to the datasets to ensure they are uniquely and persistently identifiable. This will facilitate long-term accessibility and citation.

Naming conventions used:

- The naming conventions for files and entries will follow a structured format: [ProjectID]_[DataType]_[Date]_[Location]_[Identifier]. For example, a video file collected on May 1, 2024, in Central Park, New York, might be named KUT-PRJ-2024-01_Video_20240501_CP_NY_001.mp4.

Approach towards search keywords:

- We will implement a standardized set of keywords to facilitate efficient searching and filtering of the dataset. Keywords will include species names, types of behavior observed (e.g., foraging, nesting), habitat types (e.g., park, urban area), and data types (e.g., video, image).

Approach for clear versioning:

- Version control will be maintained by appending version numbers to the dataset and individual files. The versioning format will follow the pattern v[Major].[Minor].[Revision]. For example, the first version of a dataset might be labeled KUT-PRJ-2024-01_v1.0.0. Any subsequent updates will increment the version numbers accordingly.

Standards for metadata creation:

Metadata will be created following the Darwin Core (DwC) standard, which is widely used in biodiversity informatics. The metadata fields will include:

- OccurrenceID (a unique identifier for each occurrence)
- ScientificName (the name of the species)
- EventDate (the date of the observation)
- EventTime (the time of the observation)
- Location (latitude and longitude coordinates)
- Habitat (description of the habitat)
- Behavior (observed behavior)
- AssociatedMedia (links to related media files, such as images and videos)

If no standards exist within a specific discipline for certain data types, custom metadata fields will be created to ensure all necessary information is captured and described comprehensively.

2. MAKING DATA OPENLY ACCESSIBLE *(which data will be made openly available and if some datasets remain closed, the reasons for not giving access; where the data and associated metadata, documentation and code are deposited (repository?); how the data can be accessed (are relevant software tools/methods provided?)*

Making data openly accessible:

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

Specify how the data will be made available

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

Specify where the data and associated metadata, documentation and code are deposited

Specify how access will be provided in case there are any restrictions

Guidance:

Participating in the open research data management (ORDM) does not necessarily mean opening up all your research data. Rather, the ORDM follows the principle "**as open as possible, as closed as necessary**" and focuses on encouraging sound data management as an essential part of research best practice.

The NRDIO recognises that there are good reasons to keep some or even all research data generated in a project closed. Where data need to be shared under restrictions, explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

The [Registry of Research Data Repositories](#) provides a useful listing of repositories that you can search to find a place of deposit.

Which data will be made openly available?

- All non-sensitive data collected as part of the Urban Wildlife Monitoring project will be made openly available to the research community and the public.
- Sensitive data, such as specific location information for endangered species or personally identifiable information, may be kept closed to protect the privacy and conservation interests of wildlife and individuals involved.

Rationale for closed data:

- Sensitive data will be kept closed to adhere to ethical and legal considerations regarding privacy, conservation, and data protection.
- Access to closed data may be granted on a case-by-case basis, subject to appropriate data access agreements and approvals from relevant authorities or ethics committees.

How data will be made available:

- The data, along with associated metadata, documentation, and code, will be deposited in a designated data repository.
- A public repository, such as the Zenodo platform or a discipline-specific repository for ecological data, will be utilized for data deposition.
- Data will be organized into datasets corresponding to specific research projects or monitoring periods, with detailed metadata provided for each dataset.

Access methods and software tools:

- Access to the data will be provided through standard web browsers, without the need for specialized software tools.
- Relevant documentation on data structure, file formats, and any necessary processing steps will be included alongside the deposited data.
- Open-source software tools, if used for data processing or analysis, will be referenced and, where possible, provided alongside the data to facilitate reproducibility and transparency.

Deposition of data and associated materials:

- The data, metadata, documentation, and code will be deposited in a reputable data repository known for its commitment to open access and data preservation.
- The chosen repository will provide persistent identifiers (e.g., DOIs) for deposited datasets, ensuring long-term accessibility and citability.

Access restrictions and provision:

- Access to openly available data will be provided without restrictions, allowing users to download, analyze, and reuse the data for non-commercial purposes.
- For closed data, access may be granted upon request, subject to review and approval processes outlined in data access policies and agreements.
- Access procedures, including contact information for data requests and any necessary forms or agreements, will be clearly outlined on the project website and in associated documentation.

3. MAKING DATA INTEROPERABLE (*which standard or field-specific data and metadata vocabularies and methods will be used*)

Making data interoperable:

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow interdisciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

Guidance:

Interoperability means allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins).

Assessment of interoperability:

- Interoperability of the data is crucial to facilitate its exchange, integration, and re-use by researchers, organizations, and stakeholders across different disciplines and domains.
- The data and metadata will adhere to established standards and vocabularies to ensure interoperability and compatibility with existing data systems and software applications.

Data and metadata vocabularies and standards:

- For data types such as video recordings, images, and structured reports, standardized formats will be used to promote interoperability. For example:
 - Video recordings: MP4 or AVI format
 - Images: JPEG or PNG format
 - Structured reports: CSV or JSON format
- Geospatial data, including GPS coordinates and spatial metadata, will follow standards such as GeoJSON or KML to facilitate integration with geographic information systems (GIS) and spatial analysis tools.
- Metadata will be structured according to established standards such as Dublin Core or Ecological Metadata Language (EML), ensuring consistency and interoperability across datasets.
- Where applicable, standardized vocabularies and ontologies will be utilized to describe data attributes and enhance interoperability. For example:
 - Taxonomic vocabularies (e.g., Integrated Taxonomic Information System - ITIS) for species identification metadata.
 - Environmental ontologies (e.g., Environment Ontology - ENVO) for describing habitat characteristics.
- Mapping to commonly used ontologies will be provided where necessary to bridge disciplinary or domain-specific terminology and facilitate interdisciplinary interoperability. For example, mapping species names to the Global Biodiversity Information Facility (GBIF) taxonomy to ensure consistency across biodiversity datasets.

Standard vocabulary for interdisciplinary interoperability:

- Where feasible, a standard vocabulary will be used for all data types present in the dataset to promote interdisciplinary interoperability.
- This includes adopting common standards for data representation, such as using widely accepted file formats and metadata schemas.
- In cases where discipline-specific vocabularies or terminology are used, mapping to more commonly used ontologies will be provided to enable cross-disciplinary integration and data exchange.

4. INCREASE DATA RE-USE (what data will remain re-usable and for how long, is embargo foreseen; how the data is licensed; data quality assurance procedures)

Increase data re-use (through clarifying licenses):

Specify how the data will be licenced to permit the widest reuse possible

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

Describe data quality assurance processes

Specify the length of time for which the data will remain re-usable

Guidance:

The [EUDAT B2SHARE](#) tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.

Reasons for embargoes may include time to publish or seek patents. If an embargo is sought, specify why and for how long, bearing in mind that research data should be made available as soon as possible.

Data licensing:

- The data will be licensed under open access licenses, such as Creative Commons licenses, to permit the widest possible re-use by the research community and stakeholders.
- A Creative Commons Attribution (CC BY) license will be applied to the data, allowing users to freely share and adapt the data for any purpose, provided proper attribution is given to the original source.

Availability for re-use and embargo:

- The data will be made available for re-use as soon as possible following the completion of data collection, processing, and quality assurance procedures.
- No embargo period is foreseen for the release of the data. The goal is to make the data accessible without delay to maximize its impact and potential for re-use by other researchers and stakeholders.

Usability by third parties:

- All data produced and/or used in the project will be usable by third parties, both during and after the end of the project.
- No restrictions will be placed on the re-use of the data, except where necessary to protect sensitive information or comply with legal and ethical requirements.

Data quality assurance processes:

- Data quality assurance procedures will be implemented throughout the data lifecycle to ensure the accuracy, completeness, and reliability of the collected data.
- Quality assurance measures will include:
 - Regular calibration and maintenance of monitoring equipment (e.g., motion-activated cameras).
 - Validation and verification of species identification through expert review and validation protocols.
 - Cross-referencing of data with external sources (e.g., biodiversity databases) to ensure consistency and accuracy.
 - Documentation of data collection protocols, processing steps, and any quality control

measures applied.

- Peer review of datasets and metadata prior to publication or release to verify data integrity and validity.

Length of time for data re-usability:

- The data will remain re-usable indefinitely, with no predetermined end date for its availability.
- By licensing the data under open access terms and ensuring its long-term preservation in designated repositories, it will be accessible to future generations of researchers and stakeholders for ongoing analysis, replication, and innovation.

5. ALLOCATION OF RESOURCES and DATA SECURITY *(estimated costs for making the project data open access and potential value of long-term data preservation; procedures for data backup and recovery; transfer of sensitive data and secure storage in repositories for long term preservation and curation)*

Explain the allocation of resources, addressing the following aspects:

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

Clearly identify responsibilities for data management in your project

Describe costs and potential value of long term preservation

Guidance:

Note that costs related to open access to research data are eligible as part of the grant (if compliant with the Grant Agreement conditions).

Costs are eligible for reimbursement during the duration of the project under the conditions defined in the Grant Agreement.

Address data recovery as well as secure storage and transfer of sensitive data.

Also consider whether the data is safely stored in certified repositories for long term preservation and curation.

Estimated costs for making data FAIR:

- **Data Collection and Storage:**
 - High-quality motion-activated cameras and field equipment: Ft
 - Data storage solutions (cloud services, physical storage): Ft /year
- **Data Processing and Quality Assurance:**
 - Software licenses for data processing and analysis: Ft /year
 - Personnel costs for data entry, cleaning, and quality assurance: Ft /year
- **Metadata Creation and Documentation:**
 - Development and maintenance of metadata standards and documentation: Ft
 - Personnel costs for metadata creation and curation: Ft /year
- **Open Access and Repository Fees:**
 - Fees for data deposition in open access repositories: Ft /year
 - Costs for DOI registration and maintenance: Ft /year
- **Total Estimated Annual Costs:** Ft

Covering Costs:

- Project grant funding will cover the initial costs of data collection, processing, and storage.
- Additional funding sources, such as institutional support, partnerships with conservation organizations, and research grants, will be sought to cover ongoing costs.
- Costs related to open access data management are eligible for reimbursement under the grant agreement, ensuring compliance with funding requirements.

Responsibilities for Data Management:

- Data Manager: Responsible for overseeing data collection, storage, processing, and quality assurance. Ensures adherence to data management protocols and standards.
- IT Support Team: Provides technical support for data storage solutions, backup systems, and repository integration.
- Metadata Specialist: Develops and maintains metadata standards, ensuring consistent and comprehensive documentation of datasets.
- Project Lead: Oversees overall data management strategy, allocates resources, and ensures compliance with funding and ethical requirements.

Costs and Potential Value of Long-term Preservation:

- Long-term Preservation Costs:
 - Annual fees for secure, certified repository storage: Ft /year
 - Costs for periodic data audits and integrity checks: Ft /year
 - Personnel costs for ongoing data curation and metadata updates: Ft /year
 - Total Estimated Annual Long-term Preservation Costs: Ft
- Potential Value:
 - Ensures the longevity and accessibility of valuable urban wildlife data for future research and policy-making.
 - Facilitates longitudinal studies and comparative analyses over extended periods, enhancing the understanding of urban ecosystem dynamics.
 - Supports educational initiatives and public engagement by providing accessible data resources for learning and advocacy.

Procedures for Data Backup and Recovery:

- Regular Backups:
 - Daily incremental backups and weekly full backups of all data stored on secure servers.
 - Backup copies stored in multiple geographic locations to prevent data loss due to local incidents.
- Recovery Plan:
 - Comprehensive data recovery procedures in place to restore data from backups in case of data loss or corruption.
 - Regular testing of backup and recovery systems to ensure reliability and effectiveness.

Secure Storage and Transfer of Sensitive Data:

- Sensitive Data Handling:
 - Encryption of sensitive data during storage and transfer to prevent unauthorized access.
 - Access controls and user authentication mechanisms to restrict access to sensitive data.
 - Secure transfer protocols (e.g., SFTP, HTTPS) for data transmission between field sites and

storage servers.

- Long-term Preservation and Curation:

- Storage in certified repositories that comply with best practices for data security and long-term preservation (e.g., ISO 27001 certification).
- Regular audits and updates to data security protocols to address emerging threats and ensure compliance with evolving standards.

6. DATA COMPLIANCE (*data compliance is the formal governance structure in place to ensure an organization complies with laws, regulations, and standards around its data*)

Explain legal compliance of the beneficiary institute:

Data policy and/or strategy

Data governance

Describe what legal items (national and EU) and how they are followed concerning data protection

Guidance:

Data policy sets broad, high level principles that governs data management, data interoperability and standards, data quality, data protection and information security.

Data governance entails defining, implementing and monitoring strategies, policies and sharing the management and use of data assets.

Main legal items include, e.g., laws on the protection and management of personal data at the national level, and GDPR at the EU level

To what extent the personal data management of the project and/or institute fulfil the requirements prescribed by GDPR?

Data policy and/or strategy:

- The beneficiary institute has established a comprehensive data policy that governs data management, interoperability, quality, protection, and security.
- The policy sets out principles for responsible data handling, ensuring compliance with legal and ethical standards, and promoting open access and data sharing.

Data Governance:

- The institute has a formal data governance framework in place, which includes the following key elements:
 - Data Steward: Responsible for day-to-day data management activities, including data collection, processing, storage, and sharing.
 - Data Protection Officer (DPO): Ensures compliance with data protection laws, oversees data privacy practices, and acts as the point of contact for data subjects and regulatory authorities.

Legal Compliance with National and EU Regulations:

- General Data Protection Regulation (GDPR):
 - The institute complies fully with the GDPR, ensuring the protection of personal data and the rights of data subjects.
 - Data processing activities are conducted based on lawful grounds as specified in the GDPR, such as consent, performance of a contract, and legitimate interests.
 - Data subjects are informed about the collection and use of their data through clear and transparent privacy notices.
 - Personal data is anonymized or pseudonymized where possible to enhance privacy protection.
 - Data subjects' rights, including access, rectification, erasure, restriction, and data portability, are upheld, and mechanisms are in place to facilitate these rights.
- National Data Protection Laws:
 - The institute adheres to national data protection laws, ensuring compliance with specific requirements that may extend beyond the GDPR.
 - National regulations concerning the handling of sensitive data, data retention periods, and data security measures are strictly followed.

Data Protection and Information Security:

- Data Protection Policies:
 - Comprehensive data protection policies outline the institute's commitment to safeguarding personal data and ensuring compliance with legal requirements.
 - Policies cover areas such as data collection, use, sharing, storage, and disposal, emphasizing the importance of data minimization and purpose limitation.
- Information Security Measures:
 - Robust information security measures are implemented to protect data against unauthorized access, alteration, and loss.
 - Staff are trained in data protection and information security best practices, fostering a culture of compliance and vigilance.
- Data Sharing and Third-party Compliance:
 - Data sharing agreements are established with third parties to ensure they adhere to the same data protection standards.
 - Regular audits and assessments are conducted to verify third-party compliance with data protection and security requirements.

DISCLAIMER

It is the responsibility of the Principal Investigator to inform the NRDIO of any ethics issues/concerns regarding the collection, processing, sharing and storage of data in relation to the project.