TETTRIs

TETTRIs

Transforming European Taxonomy through Training, Research, and Innovations

Deliverable 9.2

Update of the Data Management Plan

WP 9/CETAF

Document: Deliverable
Version: v.0x
Date: xx/0x/202x

TETTRIs related product

Identification	Value
Title	D9.2 - First Update of the Data Management Plan
Author(s)	Marta Leon Monedero
Affiliation	CETAF
Contributors (Affiliation)	DPO, Pavla Ruzkova, Anders Tottrup
Publisher	CETAF
Identifier of the publisher	
Doc. Version	v04
Resource	Report
Publication year	2024
Sensitivity	Low – Public document
Date	025/06/2024
Citation	

Abstract: This is the first update of the Data management plan (DMP), which was initially produced on M8 of the project and has now been reviewed in M19. This iteration reviews the initial consideration, and expands in particular the data management related to Third-Party Projects (3PPs). Additionally, it presents the appointment of a Data Protection Officer (DPO). This DMP is a living document that will be updated and further elaborated as the project advances.

Keywords: Data Management, Data Management Plan, Data Storage, Data Usage, Origin of Data, Open Access, Third-Party Projects´ Data, Personal Data, Sensitive Data, Data types, etc.

Revision:

no.	Reviewer	Status	Notes
01	Anders Tottrup	Reviewed	Commented on typos/structure and approach.
02	Pavla Ruzkova	Reviewed	
03	Jean-Marc Herpers	Reviewed	
04	Marta Leon Monedero	Reviewed	Adopted comments from reviewers.

Settings	Value
DELIVERABLE	
Deliverable title:	First Update of the Data Management Plan
Deliverable n°:	D9.2
WORK PACKAGE	
Work Package	9
Work Package Leader	RBINS
Task	T9.3
Task Leader	CETAF
Type:	Document
Lead beneficiary:	Ana Casino
Citation:	
Due date of deliverable:	M19 - 30th of June 2024
Actual submission date:	
Deliverable status:	Aproved

Document Control Information

Settings	Value
Document Title:	9.2 Data Management Plan
Project Title (Acronym):	TETTRIs
Document Owners:	Marta Leon Monedero
Technical Project Coordinator	Ana Casino, CETAF
Financial Project Coordinator	Frederik Hendrickx , RBINS
Doc. Version:	v04
Sensitivity:	Public
Date:	

Document history:

The Document Author is authorised to make the following types of changes to the document without requiring that the document be re-approved:

- Editorial, formatting, and spelling
- Clarification

To request a change to this document, contact the Document Author or Owner. Changes to this document are summarised in the following table in reverse chronological order (latest version first).

Revision	Date	Created by	Short Description of Changes

Configuration Management: Document Location

The latest version of this controlled document is stored in location.

Table of contents

1. Preface

This first update of the DMP reviews the initial considerations for Data Management, and will be mainly dedicated to expand information on the data management related to the 3PP process. The next official document update is scheduled for M42.

2. Introduction

TETTRIs is a HORIZON EUROPE (HE) funded project that aims to transform European taxonomy by facilitating knowledge sharing and improving access and use of data based on Open Science principles. These principles are crucial for developing the project and are essential for the dissemination and exploitation of results. Adhering to the principle of "As open as possible, as closed as (legally) necessary," TETTRIs will implement a comprehensive and consistent Open Science approach. Project partners will make the project data available to external parties, such as next-generation taxonomists, users in need of taxonomic knowledge, and decision-makers, under applicable regulatory provisions (e.g., GDPR, IPR, Open Data Directive) and recommendations covering ethics, gender, and diversity dimensions. Structured and regular data management is vital to the transformative use of taxonomic knowledge that TETTRIs aims to achieve, ensuring successful and sustainable development beyond the project's lifetime.

This Data Management Plan (DMP) serves as the main reference document for setting the framework and provisions for collecting, producing, storing, and using data generated within TETTRIs. This document is the first iteration of the DMP and is envisioned as a living document, in accordance with the Horizon Europe Model Grant Agreement, to be updated multiple times during the project (specifically in M7-v2, M19-v3, and M40-v4). TETTRIs will establish processes to ensure effective implementation of the DMP in terms of responsibilities, data documentation, filing, and standardisation. Additionally, it will maintain confidentiality in the context of its Third-Party Projects (3PPs) and set the conditions for archiving received proposals and their resulting outcomes.

The DMP follows the research data management policies of the institutions generating the data and adheres to the FAIR (Findability, Accessibility, Interoperability, and Reusability) Guiding Principles for scientific data management and stewardship. Furthermore, it complies with the TETTRIs Grant Agreement stipulations relating to confidentiality, security, and data protection (see articles 13, 15, 16, and especially 17).

The main body of this document, divided into several sections, collates and analyses key components related to data management based on input from partners, collected via an online survey, as the first step towards drafting the DMP. The survey, completed at the Task or WP level by 17 partners, covers the entire spectrum of data anticipated in TETTRIs. Each section is further complemented by in-house expert recommendations where applicable and necessary.

3. Project Data

As covered in the original DMP the information on the data involved in TETTRIs presented here is obtained mostly through the online survey. Additionally, the first 18th months of

implementation have confirmed and matched the initial information gathered. The main new source of information relates to the 3PPs which will be covered in section 4 of this report. The main update for this section concerns the governance of data management. Following the recommendations included in the final remarks of the initial version of the DMP, a Data Protection Officer (DPO) has been appointed. Jean-Marc Herpers (RBINS) was appointed for this role upon the consolidation of the Ethics Board. As a member of the Ethics Board he participates in the beneficiaries (or 3PPs) consultations on matters related to data.

3.1 Origin of data

The origin of data used under the activities of TETTRIs is a mix of publicly available datasets, private (institutional) data and self-generated data as already stated in the initial version of the DMP. A vast majority of partners work with publicly available data, while some of them also generate new data. A smaller number of tasks will also acquire private/copyrighted data for which obtaining a permit will be required. DNA sequencing will be used for generating new data, specifically for developing and validating molecular methods.

3.2 Collection of data

No major updates in data collection were foreseen in the original DMP. The implementation of Third-Party Projects (3PPs) may include additional sources and mechanisms for data collection. If so, these new approaches will be included in future iterations of the DMP. .

3.3 Data storage

There are also no major updates in the storage of the data within the project. As outlined in the initial DMP document (M8), various formats are used for preserving scientific data in different repositories, depending on the task. These formats include CSV, XML, RDF, and TSV. Additionally, some tasks will employ highly specialised formats for specific needs, such as fasta, PKL, or TIFF. For documentation purposes, most partners will choose common file formats like PDF or DOCX. Proprietary file formats should be used only when they offer clear, proven advantages over open data formats or when open data formats are unavailable. The project aims to implement methods for uniform and sustainable data storage and publication in accordance with FAIR principles, developed within the framework of biodiversity informatics initiatives.

The size of datasets managed by the different partners is expected to range from very small (several megabytes) to substantially large (10-50 gigabytes). Storage arrangements and reporting will be handled individually by each task. However, it is important to note that these numbers could change significantly based on various factors.

Most TETTRIs partners will use either their own infrastructure or web hosting (cloud) to store data, while a few will use a combination of both. Typically, server locations are within the EU (77%), and the non-EU servers used are GDPR-compliant.

3.4 Data Sustainability

As covered in the initial DMP document (M8), in terms of sustainable data management, TETTRIs aims to integrate its activities into existing data infrastructures as much as possible. Examples are the training courses to be included under the CETAF-DEST platform, taxonomic information systems in WP2 (in particular EUNomen, Catalogue of Life and the GBIF Checklist Bank), Citizen Science platforms such as "The Herbonauts'' and DoeDat in WP6, as well as trusted repositories for long-term presentation and curation (eg. GBIF and OBIS). The aim is to strengthen sustainable infrastructures in mutual interest and to avoid parallel development as far as possible.

3.5 Personal data[1]

As explained in the initial DMP document (M8), personal data, such as contact details (e.g., name, surname, email address, location, affiliations, publications), will be collected during various tasks of the project. Necessary personal data protection measures will be implemented in accordance with EU GDPR principles. Occasionally, datasets will include personal data; in such cases a privacy statement and consent forms will be provided beforehand. Relevant information will be anonymised appropriately, especially for publication purposes. Each institutional partner collecting and storing personal data will ensure data protection and perform administrative duties to guarantee that personal data are secure and not used for purposes unrelated to the TETTRIs project.

As outlined in M8, individuals' explicit consent will be requested for the collection and use of their personal data. Translated consent forms may be required for 3PPs in non-English-speaking countries. This is particularly important for activities involving local participants, where understanding the content is crucial (e.g., training courses). WP10 has produced different templates to request consent depending on the activity and have been shared with TETTRIs and 3PPs.

3.6 Sensitive data[2]

There is no need foreseen for any sensitive data to be gathered under the TETTRIs project. If this were to change, the appropriate handling of sensitive data will be a responsibility of each institution and the research teams gathering the data. In terms of 3PPs, that

---

[1] Personal data is any information that relates to an identified or identifiable living individual. Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data. Examples of personal data are name and surname; home address; email address such as name.surname@company.com; identification card number; location data (for example the location data function on a mobile phone); an Internet Protocol (IP) address; cookie ID*; the advertising identifier of your phone; data held by a hospital or doctor, which could be a symbol that uniquely identifies a person.

[2] Sensitive data is considered: personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs; trade-union membership; genetic data, biometric data processed solely to identify a human being; health-related data; data concerning a person's sex life or sexual orientation; data about protected species, objects of cultural heritage (e.g. Dinosaur specimens or historic, prehistoric or religious archeological sites and artefacts).

responsibility remains with the leaders of the approved projects in cases where those projects involve dealing with sensitive data.

3.7 Data Embargos

Project data is still not expected to be subject to embargoes in the context of TETTRIs.

4. Third-Party Projects

As covered in the initial DMP, as per the TETTRIs Grant Agreement, data and information gathered from the applicants submitting proposals under the 3PPs Call need to be securely stored and remain private for the sole use of their reviewing and assessment. To that end, access to that information was restricted to members of the Community Implementation Board (CIB), the TETTRIs 3PP Administrator (Catalyze[3]), the Project Management Team (PMT), and Topic Leaders. The 3PP call for proposals opened on June 30th, 2023. The online submission system was set up through the TETTRIs website where applicants could enter their proposals through the completion of a questionnaire. This system gathered data regarding the project and its applicants. This data was collected and stored by Catalyze on a local drive to which only the employees of Catalyze working on this project have access. The personal information requested was minor (name, e-mail and affiliation of the principal investigator), the majority of the data was strictly related to the projects proposed. The applicants were informed of what the use of this data would be as well as their storage conditions and their right to withdraw their information at any time. The evaluation was also carried through an online system that could only be accessed by the evaluators with specific individual credentials. Each evaluator was assigned a number of proposals to review, and in their individual portal they could only access the proposals they reviewed. The data could not be downloaded and everything had to be worked through the system, therefore they did not store the 3PPs data at any point. The 12 awarded projects agreed to the publication of their project data in the TETTRIs Website.

While each of these 12 projects are responsible for the management of their own data they have been adequately informed of the ethics requirement and GDPR compliance expected and set by the HE programme, and a specific clause tackling their responsibility to comply is included in their 3PP agreement. Guidelines including the requirements and procedures to follow have been shared for their information, as well as the templates available within the TETTRIs project. Additionally, an informal reporting exercise has been established with the 12 projects on a monthly basis in which they report their activities, providing a good overview of any potential conflicts or complex situations that might arise regarding the data involved in each initiative.

5. Data usage

As previously stated in the initial DMP, data produced by the project will be open by default and there will be no restriction on access and re-use beyond the compliance of

---

[3] An external company subcontracted via a tender to manage the administration of the Third-Party Projects.

Creative Commons licence, data outputs should be CC-0 licence and publications CC-BY licence (see Creative Commons[4]).

6. Accessible Data

TETTRIs adheres to the open access guidelines for scientific publications and research data set by the Horizon Europe programme. Beneficiaries must ensure open access to peer-reviewed scientific publications and use trusted open access repositories for datasets, scripts, and other materials during and after the project. The same requirements apply to 3PPs as they produce results. TETTRIs results will be freely disseminated through suitable channels, including scientific publications and presentations at international conferences and workshops. The primary publication venues will be highly regarded open access journals such as the European Journal of Taxonomy (EJT), Biodiversity Data Journal (BDJ), Research Ideas and Outcomes (RIO), among others.

Project deliverables gathering the main data used for each task are publicly disseminated through the TETTRIs website. Currently, there are ongoing processes and discussion on opening a dedicated open access collection in the RIO Journal or an open access repository (e.g., Zenodo) for all deliverables and other important project outputs that will not be published in journals.

7. Final remarks

In the context of the TETTRIs project, it is important to differentiate between the data generated by TETTRIs partners and those produced under the 3PPs. This iteration of the DMP aims to provide a framework for the governance of data of the 3PPs, building on the baseline provided by the original DMP.

Additionally, this DMP covers the appointment of Jean-Marc Herpers (RBINS) as DPO. He is part of the Ethics Board and supports the project in addressing personal data management and ethics-related topics (e.g., involvement in the revision of the guidelines produced for Ethics procedures).

As previously stated in the first version of the DMP, as the project advances, more accurate and precise information will be available in terms of the data generated or used under the different Tasks in the project. When this occurs, the Task leader shall transfer that information into specific data management rules or protocols to be integrated in a newly iterated version of the DMP.

---

[4] https://creativecommons.org/about/cclicenses/

APPENDIX 1

ID	Reference or Related Document	Source or Link/Location
1	1st DMP	TETTRIs D9.1 Data Management Plan.pdf
2	Consent form examples	Prior Consent Forms
3	Survey Results	LINK