



Project Deliverable D1.2: Data Management Plan

TEMPÈST

Date: 30/11/2023

Doc. Version: 1.1



TEMPEST Project

next generation Multiple architecture battery Systems for Industry

Document Control Information

Settings	Value
Document Title:	Deliverable 1.2: Data Management Plan
Document Author:	Jeremy Warren
Project Owner:	RESCOLL
Project Manager:	Jeremy Warren
Doc. Version:	1.1
Sensitivity:	PU (Public)
Date:	30/11/2023

Document Data Sheet

Settings	Value
Project Title:	TEMPEST
Coordinator:	RESCOLL
Grant Agreement N°:	101103681
Project Start Date:	May 1, 2023
Project Duration:	36 Months
Work Package:	1
Work Package Leader:	RESCOLL
Participants:	All
Document Author:	Jeremy Warren
Document Author Organization:	RESCOLL
Project Owner:	RESCOLL
Project Manager:	Jeremy Warren
Doc. Version:	1.1
Sensitivity:	PU (Public)
Date:	30/11/2023

Document Approver(s) and Reviewer(s):

NOTE: All Approvers are required. Records of each approver must be maintained. All Reviewers in the list are considered required unless explicitly listed as Optional.

Name	Role	Action	Date
Jeremy Warren	Project Manager	<Approve/Review>	10/10/2023
Magali Clavé-Henry	AGB	Review	10/10/2023



Document history:

The Document Author is authorized 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 summarized in the following table in reverse chronological order (latest version first).

Revision	Date	Created by	Short Description of Changes
1.0	10/10/2023	Jeremy Warren	Initial Version
1.1	30/11/2023	Jeremy Warren	Changes requested by Project Officer: <ul style="list-style-type: none"> • Addition of full project title to Pg. 2. • Cleanup of front page formatting • Update of document name to correspond with deliverable names as per Annex 1. • Further elaboration on conclusions

Configuration Management: Document Location

The latest version of this controlled document is stored in the TEMPEST project's private repository, in the WP1 folder.



TEMPEST Deliverable

TABLE OF CONTENTS

1. DEFINITIONS	5
2. EXECUTIVE SUMMARY	5
3. INTRODUCTION	6
3.1. Methods for Data Collection and Overall Strategy	6
4. DATA SUMMARY	7
4.1. Data Reuse	7
4.2. Types of Data	7
4.3. Purpose of Data	7
4.4. Data Size.....	8
4.5. Data Origin.....	8
4.6. Utility of Data.....	8
4.7. Collection of Data	9
5. FAIR DATA	9
5.1. FAIR data: Making data findable, including provisions for metadata	9
5.2. FAIR data: Making data accessible	10
5.3. FAIR data: Making data interoperable.....	12
5.4. FAIR data: Increase data re-use.....	12
6. OTHER RESEARCH OUTPUTS	13
7. ALLOCATION OF RESOURCES	13
8. DATA SECURITY	14
9. ETHICS	14
10. OTHER ISSUES	15
11. EUROPEAN COMMISSION (HORIZON): HORIZON EUROPE DMP + - GDPR RECORD	15
11.1. GDPR record.....	15
12. EUROPEAN COMMISSION (HORIZON): HORIZON EUROPE DMP + - DPIA	15
12.1. DPIA.....	15
13. CONCLUSION	15



1. DEFINITIONS

API	Application Programming Interface
KPI	Key Performance Indicator
WP	Work package
DMP	Data Management Plan
DOI	Digital Object Identifier
FAIR	Findable, Accessible, Interoperable, and Reusable
GDPR	General Data Protection Regulation
ORCID	Open Researcher and Contributor ID
PAR	Parity ARchive
PID	Persistent IDentifier
PU	Public
PURL	Persistent Uniform Resource Locator
ROR	Research Organization Registry
SENS	Sensitive

2. EXECUTIVE SUMMARY

This deliverable, D1.2, represents the initial version of the DMP, and will be updated in two subsequent deliverables: D1.5 (Data Management Plan Update 1, M18) and D1.6 (Data Management Plan Update 2, M36). This deliverable encompasses the initial version of the Data Management Plan, and will describe the following elements:

- The guiding principles of the DMP
- Summary of data: its sources, reuse, gathering, and processing during the project.
- How data will be stored and processed, both at the project level, and in terms of curation and processing for FAIR compliance with Horizon Europe guidelines.
- Allocation of resources for FAIR compliance
- Security of data and archival strategies.
- Strategies for data metadata allocation
- Applicability of GDPR

The overall goal of the DMP is to put in place an environment which supports efficient and secure data collection, along with classification methods to clearly identify data for its dissemination level, and whether it is considered to be “IP sensitive”, i.e., critical to ongoing or future patent or other IP protection. Data classified for distribution to public repositories will be anonymized, if necessary, prior to uploading.

Zenodo will be used as the public repository of choice for open research data, although other repositories will also be evaluated during the project as additional sources for storing open data. TEMPEST will comply with the Horizon Europe Open Research Mandate.

The coordinator of the project, RESCOLL, has implemented a file-sharing platform based upon Nextcloud, which will be used for all file storage during the project’s 36-month duration. The part of the platform dedicated to the TEMPEST project is uniquely accessible to project partners.

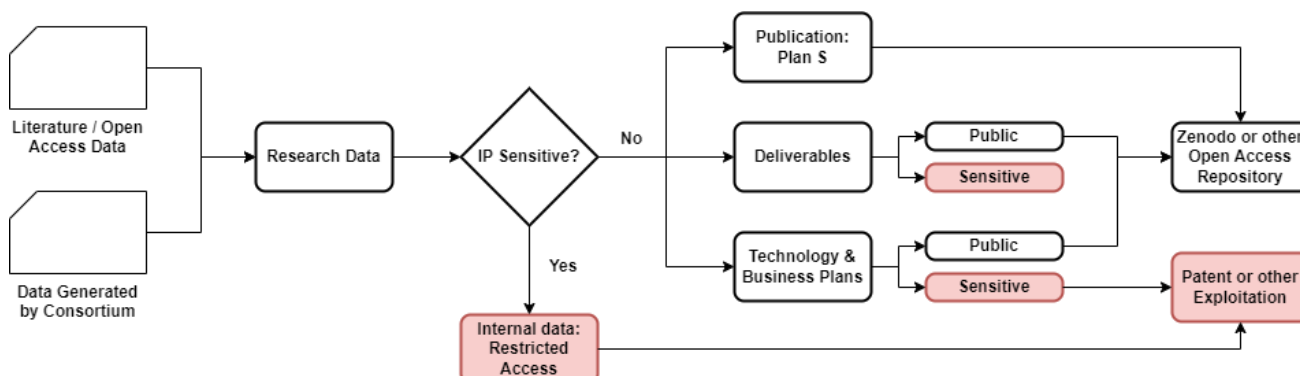
The DMP is intended to be a living document, and will be updated during the project, through D1.5 and D1.6.



3. INTRODUCTION

3.1. Methods for Data Collection and Overall Strategy

Data collection and management can best be described by the schema below:



During the project, a log will be kept of all pertinent data generated which will describe the following basic elements of information:

- Generating Organization: Name and ROR
- Generating User: Name and ORCID
- Project Repository Primary Folder location and subfolder(s)
- Filename
- Title
- Version
- Date Logged
- File Type (e.g., PDF, Word, excel, etc.)
- Unique Internal Number

In addition, supplementary metadata will be logged, which will include:

- IP Sensitive Flag (see above), either as TEMPEST foreground IP or Partner Background IP
- Classification (Public or Sensitive, see above)
- Whether it is cleared for release
- License
- Taxonomy (where relevant)
- Keywords

Once data is cleared for release and publication on Open Access Repositories, the following additional metadata will be logged:

- Open Access Repository Name
- Persistent identifier
- Date Crated on Open Access Repository

As is shown in the schema above, the IP Sensitive flag is very important for identifying what data is to be clearly separated from all other data processed during the project. IP Sensitive flagged data will be compartmentalized, with clear justification given for its classification as such. All other data will be processed as shown above, with preference to Public data wherever possible. It should be noted that this metadata will be in addition to any intrinsic metadata generated by the application and available to the file format (e.g., EXIF image data).

4. DATA SUMMARY

4.1. Data Reuse

Will you re-use any existing data and what will you re-use it for?

The TEMPEST project will use, where necessary, data from literature sources for some of its work, notably with respect to materials properties to be used for modeling. As this data represents a reliable, trusted source of information, it will serve as a baseline. In cases where data is missing in this framework, it will be completed by mechanical or physicochemical testing.

4.2. Types of Data

What types and formats of data and other research outputs will the project generate or re-use?

The TEMPEST project has identified the following types of data which will be generated during the project. The data types below represent data which may be either generated, collected or reused during the project.

Type	Native Format	Open Format	Typical Size/file
Mechanical Testing	Proprietary	CSV	Several MB
Physicochemical Testing	Proprietary	CSV	Several MB
Images	JPG, other	JPG	Several MB
Video	Varies	MKV	Several GB
Simulation Data	Proprietary	CSV or ASCII	Up to several GB
Mathematical Models	Proprietary	ASCII	Several MB
NDT/SHM	Proprietary	CSV	Several MB
3D Models	Proprietary	IGES or STEP	Up to many MB
Matlab code	Matlab proprietary	.txt	Up to MB
Simulation model	*.SPFX	.CSV	Several MB

4.3. Purpose of Data

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

Data generated or reused during the project will be used for the following purposes:

1. Generation of base-level material data for performing simulation work
2. Generation of testing data for the following purposes:
 - a. Validating the results of simulation work



- b. Evaluating the performance of materials
 - c. Characterizing materials properties
 - d. Evaluating battery and subsystem operation and performance
 - e. Calibration and/or verification of the operational state of equipment used or developed during the project
3. Cross-validation of data against existing standards or literature values for quality control

4.4. Data Size

What is the expected size of the data that you intend to generate or re-use?

In total, the current expected quantity of data to be generated or reused during the project should not exceed 500 gigabytes. This is purely a preliminary value, as the major data-generating stages of the project have not yet begun, so only initial estimations can be provided. However, the following estimations can be provided:

- Mechanical testing data: $\sim 200 \times 5 \text{ MB} = 1 \text{ GB}$
 - Physicochemical testing data: $\sim 500 \times 5 \text{ MB} = 2.5 \text{ GB}$
 - Images: $500 \times 5 \text{ MB} = 2.5 \text{ GB}$
 - Videos: $50 \times 1 \text{ GB} = 50 \text{ GB}$
 - Simulation data: $100 \times 500 \text{ MB} = 50 \text{ GB}$
 - Mathematical models: $25 \times 2 \text{ MB} = 50 \text{ MB}$
 - NDT/SHM data: $200 \times 10 \text{ MB} = 2 \text{ GB}$
 - 3D models: $20 \times 3 \text{ MB} = 60 \text{ MB}$
 - Matlab code: $25 \times 1 \text{ MB} = 25 \text{ MB}$
 - Simulation models: $75 \times 2 \text{ MB} = 150 \text{ MB}$
 - Other data (Word, Excel, PowerPoint, etc.): $350 \times 5 \text{ MB} = 1.5 \text{ GB}$
- ***Preliminary total: 135 GB***

4.5. Data Origin

What is the origin/provenance of the data, either generated or reused?

Data which is generated by the TEMPEST project will primarily be from work performed by the partners of the consortium. In one case, additional data will be generated as a result of subcontracting work related to slurry evaluation. The details of the data provenance and its handling will be described in Update 1 to the DMP.

Reused data will primarily come from the literature and open-access repositories. These will be documented on a case-by-case as they arise during the project.

4.6. Utility of Data

To whom might your data be useful ('data utility'), outside your project?

The TEMPEST project will generate a large amount of data (see above), which will be of great interest to the following organizations/entities:

- Companies developing new batteries and testing methods for batteries and their subsystems



- Research and technology organizations working in battery research
- University researchers working on batteries and related technologies
- Any organization working on functional coatings (for impact protection, fire protection, thermal management, etc.)

4.7. Collection of Data

Data generated in the TEMPEST project will be deposited on the TEMPEST repository hosted by Applus+ RESCOLL. Data uploads of data for Open Access publication to the platform should follow the following guidelines:

- Data should be placed in the “Research Data for Open Access Publication” folder
- Data should be named as follows:
 - OA-TEMPEST.{DocumentTitle}.{UniqueInternalNumber}
- All data should use the unique internal number from the Data Management log

All scientific publications, public classified deliverables, and data approved for Open Access public publication should be published on Zenodo according to the following guidelines:

- All data must be uploaded to the [TEMPEST Project Community](#) and the [European Commission Funded Research \(OpenAIRE\) Community](#).
- If as a TEMPEST Project member, you do not yet have a Zenodo profile:
 - Please create a profile. It may be easiest to link to your ORCID profile.
 - Click on the link to the TEMPEST Project Community and click on “New Upload”
 - Enter the appropriate data for the file. Please make sure to add the European Commission Funded Research (OpenAIRE) Community to the list of communities in order to make sure that the dataset is uploaded to both communities.

5. FAIR DATA

5.1. FAIR data: Making data findable, including provisions for metadata

Will data and other research outputs be identified by a persistent identifier?

Yes. As described above in § 3.2, all data cleared for release and publication on Open Access Repositories such as Zenodo will have a persistent identifier generated and logged in the Data Management log for the TEMPEST project. These will be accompanied by unique identifiers for the organizations (ROR numbers wherever possible) and ORCID identifiers for the researchers involved. In order to maintain consistency, ORCID identifiers have been chosen as the unique allowable identifier for involved researchers during the TEMPEST project.

Unique IDs for published data may include (but are not limited to):

- DOI for publications
- Repository-specific PURLs or PIDs
- Additional PURLs or PIDs generated as needed



Metadata

1. **Will rich metadata be provided to allow discovery?** Yes
2. **What metadata will be created?** Metadata will be created/translated during the Open Access publication process on Zenodo (or other approved platform) in accordance with FAIR principles. As a minimum, data will be described in compliance with the DataCite Metadata Schema.
3. **What disciplinary or general standards will be followed?** The TEMPEST project will follow the standards for OpenAIRE data accessibility, as well as compliance with all standards employed by Zenodo.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Yes: During the logging of data in the TEMPEST Data Management log, keywords will already be identified for all pertinent data. These keywords will be translated over to the metadata during publication on Zenodo, in order to enhance the metadata discovery for others.

Will metadata be offered in such a way that it can be harvested and indexed?

Yes: Through the Zenodo (or other appropriate platform), metadata will be made available through an open REST API, in JSON format.

5.2. FAIR data: Making data accessible

Will the data and other research outputs be deposited in a trusted repository?

Yes: As described above, the targeted repository for all Open Access publication will be Zenodo, though other Open Access repositories will be evaluated during the project. Any additional repositories will be vetted in order to ensure their compliance with Horizon Europe guiding principles on Open Access data.

Have you explored appropriate arrangements with the identified repository where your data and other research outputs will be deposited?

Yes: The process has not yet been completed, but it is in the process of being performed. As the project is not yet at the stage of publishing data to an Open Access repository, it is not yet critical, but it will be completed as soon as possible.

Does the repository ensure that the data and other research outputs are assigned an identifier? Will the repository resolve the identifier to a digital object?

Yes: Zenodo ensures persistent identifiers are assigned to data, and will resolve the data to a digital object. Any other resources used (e.g., DOIs) will also be chosen for the same resolution capability.

Will all data and other research outputs be made openly available?

No, certain datasets cannot be shared openly for the following reasons: As described in § 3.2, certain data may be flagged as “IP Sensitive” or “Sensitive.” Such data is not for public dissemination, though justification will be given in all cases (e.g., the Deliverable in question was flagged as SENS in the Grant Agreement).



Is an embargo applied to give time to publish or seek protection of the intellectual property (e.g., patents)?

Yes: The terms of embargo periods for publication and protection are described in the TEMPEST project's Consortium Agreement.

If an embargo is applied (see question 2.2.5), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

For publication, a pre-draft version must be supplied to consortium members at least 45 days in advance of publication, with any objections by consortium members required no later than 30 days prior to publication.

Will the data and other research outputs be accessible through a free and standardized access protocol?

Yes: As described above, Zenodo will be used as the preferred repository, and makes available data through free and standardized access protocols.

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

Any data flagged as "IP Sensitive" or "Sensitive" will not be made available to the public, either during or after the project. All other data, however, will be made available as possible.

How will the identity of the person accessing the data be ascertained?

As previously mentioned, all Sensitive or IP Sensitive data will be compartmentalized to the TEMPEST project's repository. Access to this repository is managed by Applus+ RESCOLL, with secure authentication on a per-user required for access to data.

Is there a need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?

No: This is handled at the Project Steering Committee level in the project. A separate committee is not needed.

Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why.

Yes

Will metadata contain information to enable the user to access the data?

Yes

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

As per the terms of Zenodo's service, data access is guaranteed for at least 20 years. Any additional Open Access publication sources will need to have comparable data retention lifetimes. The consortium will endeavor to find supplementary repositories to Zenodo, in order to have fallback redundancy for data and metadata accessibility.

Will documentation or reference about any software needed to access or read the data be included? Will it be possible to include the relevant software (e.g., in open source code)?



Wherever possible, software code will be open source; however, this will be subject to the exploitation strategy of the partners and the classification level of the code, as described in § 3.2.

5.3. FAIR data: Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines?

In order to ensure interoperability, the data will be compliant with DataCite metadata schema. The [DataCite Metadata Generator](#) will be used to provide a framework for XML metadata generation.

Will you follow community-endorsed interoperability best practices? Which ones?

Yes, the [DataCite Best Practice Guide](#) will be used as the foundation for all metadata and data interoperability.

In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies:

Will you provide mappings to more commonly used ontologies?

Yes, wherever possible.

Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Yes. They will be published through Zenodo.

Will your data and other research outputs include qualified references to other data (e.g., other data from your project, or datasets from previous research)?

Yes, wherever possible, using standard persistent identifiers and with explanation of the intent of the reference in order to provide a meaningful association.

5.4. FAIR data: Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use?

Data packages for publication through Zenodo will include appropriate documentation to describe the methodology, and pertinent information necessary to validate the data analysis.

Will your data and other research outputs be made freely available in the public domain to permit the widest re-use possible?

Wherever possible, yes. However, this will be subject to clearance of the data for release into the public domain by the decision of the consortium, taking into consideration the limitations imposed by the need for commercial exploitation and intellectual properties. However, it should be clear that preference will be given to licensing the data through standard reuse licenses (Creative Commons).

Will your data and other research outputs be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Again, wherever possible, yes. However, this will be subject to clearance of the data for release into the public domain by the decision of the consortium, taking into consideration the limitations



imposed by the need for commercial exploitation and intellectual properties. In compliance with Zenodo guidelines, the preferred license for Open Access publication will be Creative Commons Attribution 4.0.

Will the data and other research output produced in the project be useable by third parties, in particular after the end of the project?

Yes: The use of Open Access archives such as Zenodo will ensure that research data can be used after the end of the project.

Will the provenance of the data and other research outputs be thoroughly documented using the appropriate standards?

Yes

Describe all relevant data quality assurance processes.

All data for publication will be screened by the Data Manager prior to publication, in order to ensure that the data is appropriately logged, tagged with the correct metadata, and prepared for submission to Zenodo. Once submitted to Zenodo, the curation process will assure that data is compliant with DataCite, and that all appropriate metadata and supporting information is supplied prior to approval.

This process will be updated in the later revisions of the DMP as needed.

6. OTHER RESEARCH OUTPUTS

Do you have any additional information that was not addressed in the previous sections, which you wish to provide regarding other research outputs that are generated or reused throughout the project?

This will be updated in the two subsequent updates to the DMP.

7. ALLOCATION OF RESOURCES

What will the costs be for making data and other research outputs FAIR in your project?

For the moment, all costs related to project-related storage and archival of the data will be absorbed by Applus+ RESCOLL. During the project, approximately 4 PM will be dedicated to Data Management tasks, with a total personnel cost of approximately 30 k€.

Data hosting on the server is shared with other virtual machines; however, current measurements estimate a power consumption linked to the data storage, database and web front end (on average, 17.5 W of the total average power consumption of ~180 W) to be on the order of approximately 153 kWh/year, or an annual cost of €30/yr. The server is power-optimized in order to limit the impact upon the environment by the hosting activities.

How will these be covered?

The hosting costs are minimal: these will be absorbed by RESCOLL in operating overhead.



The personnel costs are part of the project financing, for RESCOLL's activities.

Who will be responsible for data management in your project?

Name	Dr. Jeremy Warren
Partner	Applus+ RESCOLL
Email	Jeremy.warren@applus.com
Telephone	+33 5 47 74 69 00 Extension 1106

How will long-term preservation be ensured?

Data will be archived at RESCOLL for at least five years following the end of the project, with an objective of ten or more years if possible. In addition, long-term preservation will also be assured through Open Access repositories such as Zenodo. The costs for hosting at RESCOLL are only in server operating costs, which are minimal, and represent part of the operational overhead. This will be updated with a more complete plan in the final update to the DMP.

8. DATA SECURITY

What provisions are or will be in place for data security?

Data integrity & archival will be assured at several levels on the TEMPEST Project File Repository:

- All files deposited on the File Repository have regularly scheduled, redundant snapshots to track changes and allow rollback. These snapshots also include check summing for assuring data integrity.
- Regular multi-tiered backup of data on three levels, and with offsite backups. These backups also include data check summing for data integrity.
- Regular, secondary archival of data for long-term storage, including parity files (PAR2 format) for verifying and detecting/repairing any errors which may arise (such as bitrot). These secondary archives will also be part of the multi-tiered archive system mentioned above.

Security of data will be assured during transfer, using TLS 1.3-compliant encryption, with perfect forward secrecy. Regular audits of the web server are performed in order to ensure that it is resistant to vulnerabilities (such as POODLE, BEAST, GOLDENDOODLE, DROWN, etc.) of the encryption layer.

Will the data be safely stored in trusted repositories for long-term preservation and curation?

Yes: Data will be stored in trusted repositories for preservation and curation. The first choice will be Zenodo, but other, secondary repositories will be evaluated during the project.

9. ETHICS

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?

No: No processing of personal data is foreseen in the project.

Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?

Not applicable: No personal data will be processed within the scope of the TEMPEST project.



10. OTHER ISSUES

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

No

11. EUROPEAN COMMISSION (HORIZON): HORIZON EUROPE DMP + - GDPR RECORD

11.1. GDPR record

Have you registered personal data processing activities for this project?

Not applicable: No personal data will be processed within the scope of the TEMPEST project.

12. EUROPEAN COMMISSION (HORIZON): HORIZON EUROPE DMP + - DPIA

12.1. DPIA

Have you performed a DPIA for the personal data processing activities for this project?

Not applicable: No personal data will be processed within the scope of the TEMPEST project.

13. CONCLUSION

This first version of the DMP for the TEMPEST project outlines the philosophy, framework, and methodology for the management of data, especially the data for publishing under Open Access rules, created within the TEMPEST project. A detailed and coherent Data Management Plan is key to ensuring that not only is data effectively managed and protected, but that properly data classification is implemented in order to ensure efficient and effective open access to research data.

This first iteration of the TEMPEST Data Management plan lays the groundwork for an effective data management strategy, and will be regularly updated during the project. The data collection and classification process, which is now underway, has already been useful in establishing a clear method for separating IP-sensitive information from that available for open access, and will further assist in the process of preparing data packets for publication on the TEMPEST project repository on Zenodo.

