UncorrelaTEd

Solid-liquid thermoelectric systems with uncorrelated properties

Deliverable 6.9

Data Management plan

H2020-EU.1.2.1. - FET Open

FETOPEN-01-2018-2019-2020 - FET-Open Challenging Current Thinking

Grant Management 863222

Type of Action: RIA

Start Date: 01 Jan 2020

Duration: 48 months

Project partners

LOGO	Partner full name	Acronym
	Universitat Jaume I	UJI
	Institut de Recerca en Energia de Catalunya	IREC
	Kungliga Tekniska Högskolan	KTH
	University of Warwick	UW
	Solvionic	SOLV
	Intenanomat	INM

Deliverable Name: Dissemination plan

Led by: UJI

Partners: All

Index

Figure index

Table index

1   Introduction

The novel technology of UncorrelaTEd comes from the combination of thermoelectrics (TEs) and electrochemistry. It will set the scientific foundations of electrochemically activated high-efficiency TE energy conversion by correlating for the first-time electrochemical effects with TE performance. The conclusions of the project will improve our currently feeble understanding of the mechanisms governing energy conversion in these new systems. Therefore, among others, UncorrelaTEd will generate:

- new hybrid solid-liquid devices with unprecedented TE efficiency in the heat-to-electricity energy conversion,
- new knowledge by studying phenomena and manipulating matter at the nanoscale, and particularly at solid-liquid interphases,
- new materials with tuneable nanoscale composition and porosity,
- and a simulator to model the hybrid devices and simulate their performances, bridging TE material and electronic device concepts.

As a consequence of the development of the project and its related research activities, a set of theoretical and experimental data will be generated. The data management plan (DMP) provides details of what data the project will produce, how it will be exploited and made accessible for verification and re-use, and how it will be cured and preserved.

This document, entitled "Deliverable 6.9 Data Management Plan", is published because of the "Task 6.4: Creation of the data management plan (UJI)" of the "WP6 Management and dissemination". This deliverable is publicly available on the project's website and Cordis.

2   Objectives

Future and Emerging Technologies (FET) projects starting from 2017 must participate in the Open Research Data Pilot (ORD Pilot) on open access to research data. The ORD pilot aims to improve and maximise open access and re-use of research data generated by H2020 projects. It considers the need to balance openness and protection of scientific information, commercialisation and Intellectual Property Rights, privacy concerns, security as well as data management and preservation questions. The development (and keeping up-to-date) of the DMP is a central pillar of the open research data ORD pilot. The other components related to the research data:

- Deposit it in a public repository.
- Ensure third parties can freely access, mine, exploit, reproduce, and disseminate it.
- Provide related information and identify (or provide) the tools needed to use the raw data to validate the project research.

Notably, the ORD pilot applies to:

- The data (and metadata) required to validate results in scientific publications.

- Other cured and/or raw data (and metadata) specified in this document, following the Grant Agreement conditions.

The DMP identifies key actions and strategies for proper data management. It ensures that research data are of a high quality, safe, sustainable, and (when possible) accessible and re-usable, following the findable, accessible, interoperable, and re-usable (FAIR) principles. Planning for data management is a process, and hence, the DMP will be updated when necessary as the project moves forward. The DMP includes information on:

- the handling of research data during and after the end of the project
- what data will be collected, processed, and/or generated
- which methodology and standards will be applied
- whether data will be shared/made open access and
- how data will be cured and preserved (including after the end of the project).

The present document has been prepared following the H2020 FAIR DMP Template.

3 Data Summary

What is the purpose of the data collection/generation and its relation to the objectives of the project?

UncorrelaTEd is structured in six working packages, according to the specific objectives proposed in the project. Data will be generated by all Working Packages (WPs) except the sixth. Two types of data are identified and shown in Table 1, according to each WP and partner responsible for the data.

Table 1. Experimental and simulation data generated.

WP	WP1	WP2	WP3	WP4	WP5
Type of data	Experimental	Experimental	Experimental	Experimental	Simulation
UJI	x	x	x	x	x
IREC	x		x		
KTH		x			
UW					x
SOLV	x	x	x	x	
INT				x	

WP1 studies in detail the electrochemical phenomena influencing the PF that resulted in breaking the adverse S-σ interdependence by the presence of an electrolyte. Then, WP2, WP3, and WP4 are focused on hybrid TE systems with bismuth telluride alloys, oxide, and polymer materials, respectively. Finally, WP5 performs computational simulations to support the electrolyte optimisation of the systems explored in the project and provide theoretical

predictions. All these five WPs will generate data, and it is clearly identified in which WPs will be experimental (WP1 to 4) and based on simulations (WP5).

Experimental data refers to measurements but also to other calculated key parameters used to analyse phenomena, conclusions about optimum electrolytes based on different materials, and procedures to produce these materials. Simulation data instead include results coming from the simulations but also the code for the advanced simulator developed.

Will you re-use any existing data and how? What is the origin of the data?

The data re-used in this project belong to the partners of the project.

The starting point of UncorrelaTEd is based on the promising previous results observed in $Sb:SnO_2$ devices and presented in the article, [Large Power Factor Improvement in a Novel Solid–Liquid Thermoelectric Hybrid Device](#), published by the Coordinator (UJI). These results have motivated the research proposed in UncorrelaTEd, but these experiments will be replicated with the new measurement devices and other materials have been proposed in the comparison.

Materials such as Ionic Liquids will be used, and they will be selected according to their properties, which are data that are available from Solvionic.

The starting point for the development of the simulator is software routines that have been developed by Prof. Neophytou (UW), which will be modified and extended to be applicable to the needs of this project.

What is the expected size of the data?

It is too early at this point to determine the total size of the data generated by the project. Usually, data generated by the software used for measurements can be exported to .txt format. Files range from a few KB to a few MB. In case the data needs to be transferred, and the size is considerably large, data and/or folders will be compressed using the most used software (.zip, .rar, etc.). In some situations, data representation will require other formats. The Consortium has agreed to use the following file formats:

- *Text, documents*: ASCII, Word (.docx), PDF.

- *Presentations*: Microsoft PowerPoint (.pptx).

- *Numerical*: ASCII, MS Excel spreadsheet (.xlsx and comma-delimited .csv), MATLAB (.m).

- *Images*: TIFF or JPEG.

- *Audio*: MP3 or WAV.

- *Video*: MP4 (or another preferred).

- *Software*: Python, Gaussian 09, Dalton, GaussView 5.0, Mathcad, Mathematica, etc.

- *Specific format*: CIF (for crystallography files), etc.

- *Instrument-specific*: LabVIEW Data Format, Nova, ZView, EC-Lab®, MITS Pro, etc.

The simulator will be constructed in MATLAB, and all the files will be created as .m. The data produced by the code will be saved in the MATLAB data files with .mat extension.

To whom might it be useful ('data utility')?

The deposition of research data in repositories will allow different types of third parties to access, mine, exploit, reproduce, and disseminate it. As an example of interested third parties, some of them have been identified: researchers involved in the field of electrochemistry, materials, and energy; and TE modules manufacturers, but also others such as the internet of things (IoT) or automobile industries, among others. Potential beneficiaries from data generated will be able to use data since the data generation process in UncorrelaTEd will comply with the guidelines of the European Commission on Open Access to Research Data in H2020 projects for data management (Figure 1).

The simulator will be useful to theorists and experimentalists that will access electronic and thermal transport in hybrid thermoelectric materials.
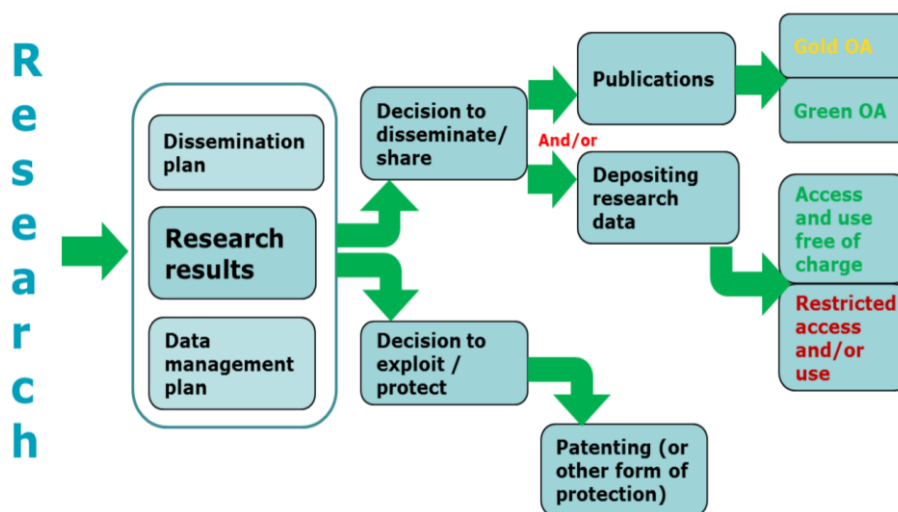

Figure *1. Open access to scientific publication and research data*[1]

4 FAIR data

This document helps make UncorrelaTEd research data findable, accessible, interoperable, and re-usable (FAIR), which is the key conduit leading to knowledge discovery and innovation, and subsequent data and knowledge integration and re-use. Sets of results/data generated by the project tasks will be transferred to the Zenodo repository. Zenodo is an open repository from

---

[1] European Commission, Directorate-General for Research & Innovation (2017). H2020 Programme: Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 (version 3.2). Retrieved from: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf (last access 8 June 2020)

the OpenAIRE H2020 project and CERN. Zenodo repository respects FAIR principles, and additional details from those mentioned in this text can be found here.

4.1   Making data findable, including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata, identifiable and findable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

Data uploaded to Zenodo is linked to OpenAIRE and the EC portal, which guarantees its accessibility to all public. Moreover, data will be identifiable and findable using Zenodo because it assigns a unique Digital Object Identifier (DOI) to all publicly available uploads. When possible, the open-access publication will be uploaded together with the data used in it.

A community named "UncorrelaTEd H2020 FET Open Project" has been created in Zenodo, for which all datasets will be linked (Figure 2). The website to access is https://zenodo.org/communities/uncorrelated/, and it will be included on the site of the project.
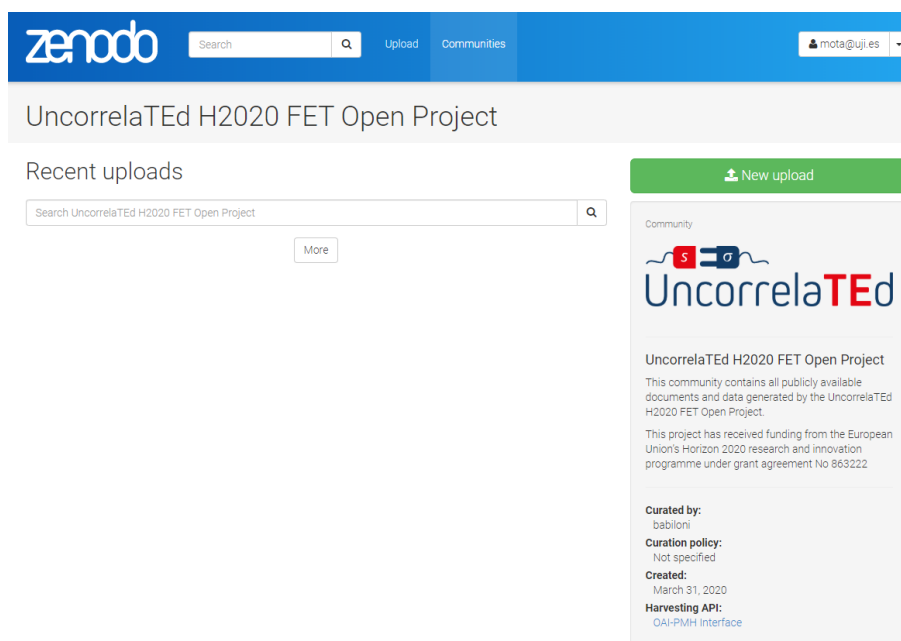
Figure 2. The UncorrelaTEd community in Zenodo.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

README files will contain metadata, including methodology details, equipment type, and uncertainty, among others. For computer simulations, the README file will contain supplementary information about the codes. Metadata will include an acknowledgement to the European Commission, as stated in the dissemination plan; the name of the action, acronym and grant number; the publication date, and length of embargo period if applicable, and a persistent identifier.

What naming conventions do you follow? Will search keywords be provided that optimise possibilities for re-use?

Naming conventions are not initially established for the data and metadata, but complete information (including keywords for optimising re-use possibilities) about them will be included in the repository. English will be used for all data and metadata generated.

Do you provide clear version numbers?

Precise version numbers will be used for internal management of the data and data included in repositories. In principle, only the final version of data and metadata documents is required to be included in repositories. However, if some file already existing in a repository is updated (for any reason), the modifications performed to the last version of a document will be specified in a README file.

When a new dataset is uploaded to Zenodo, this will be announced in the News & Events section of the UncorrelaTEd webpage, using a particular category (Open data). In addition to the Zenodo repository, which will always be prioritised, data can be uploaded to scientific platforms such as ResearchGate (Figure 3). The UncorrelaTEd project has already been included in this platform, as explained in the previous deliverables.
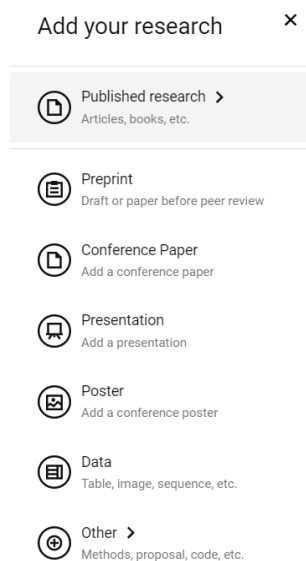
Figure 3. Items that can be added to the ResearchGate website.

Moreover, scientific journals in which is planned to submit the scientific articles offer the possibility of including research data to interlink with the published articles and validate research findings. If data is already available in another depository (Zenodo in the case of this project), articles can be directly linked to datasets, even during the submission stage. Furthermore, publications can include a complete methodology section, and extended information can be added in Annexes.

Some of the project deliverables are public, such as those that present the optimum electrolytes to maximise the PF in different materials, or those coming from the simulation Work Package. In these reports, the data included in repositories will be cited using the DOI received.

4.2 Making data openly accessible

Which data produced and/or used in the project will be made openly available as default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

A summary of the data generated in UncorrelaTEd is included in Table 2, according to the principles established in the DMP. In all cases, metadata will be provided along with each set of data. Metadata will include details of the equipment used, such as the make and model of the instrument, the settings used and information on how it was calibrated, etc.

Table 2. Summary of the data generated.

Confidential	Public (after the protection of the results)
Procedure to fabricate films and TE materials with controlled/different porosity.	Phenomena that influence the PF improvements. Optimum electrolytes to maximise the PF. Advanced simulator and best electrolytes identified by the simulations.

Submitting open data to repositories does not change the obligations to protect results, confidentiality, and security, as specified in the Grant Agreement. If it is decided not to grant open access to additional research data during the project, this will be informed in the periodic technical/scientific review meeting documents.

How will the data be made accessible (e.g., by deposition in a repository)?

The data will be made accessible by deposition in the Zenodo repository, which will give open access to data generated by the project.

What methods or software tools are needed to access the data?

No unique methods or software will be needed to access the data, only an internet browser and a text editor or spreadsheet software.

Is documentation about the software needed to access the data included?

Not applicable.

Is it possible to include the relevant software (e.g., in open source code)?

Not applicable.

Where will the data and associated metadata, documentation, and code be deposited? Preference should be given to certified repositories that support open access where possible.

Zenodo platform offers possibilities to include associated metadata and other types of documentation within the code.

Have you explored appropriate arrangements with the identified repository?

The conditions applicable to the selected repositories were explored when the DMP was written.

If there are restrictions on use, how will access be provided?

All data submitted to open repositories have no restrictions on use.

Is there a need for a data access committee?

Partners must inform the other members of the Consortium about intended publications of results 20 days before submission for publication. The same principle will be applied to data, for which partners will also inform the Consortium of the data that is going to be made open access, specifying if it is related to any future scientific communication.

Are there well-described conditions for access (i.e., a machine-readable license)?

Not applicable.

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

As an open-access platform, only general statistics can be retrieved from the access to the different versions submitted of data (views, downloads, data volume, unique views, and unique downloads). Therefore, the identity of the person who accesses the data cannot be ascertained. Besides, Zenodo offers the possibility of consulting the number of citations per element, and therefore, the scientific impact of the dataset can be quantified.

Other means mentioned in the previous section for storing open data have more restrictions than Zenodo. ResearchGate requires free registration on the website, but the identity of the person accessing the data can be ascertained.

Information about the code developed in WP5 by UW will be published in a deliverable. Therefore, it can be submitted to UW's webpage and the public GitHub platform. GitHub is a development platform where programmers can host and review code, manage projects, and build software. GitHub allows adding other files complementary to the program (Figure 4). For instance, readme files with additional information of the program developed (requirements, installation guide, instructions for use with provided data, citations, etc.) will be included.

📁 Code	Add files via upload	9 months ago
📁 Data	Delete sdas	9 months ago
📁 Results	Delete asda	9 months ago
📄 LICENSE	Create LICENSE	9 months ago
📄 README.docx	Add files via upload	9 months ago
📄 README.md	Update README.md	9 months ago

Figure 4. Example of files and folders in a GitHub program.

Data contained in files (and respective folders) will have an appropriately descriptive title/name. UncorrelaTEd Consortium involves research institutions and companies that already have naming conventions established. Therefore, a one-model-fits-all approach would create difficulties. In any case, a set of minimum essential information (WP, date, institution, …) is going to be used to ensure cross-platform coherence.

4.3   Making data interoperable

Are the data produced in the project interoperable, that is 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)?

Data produced in the project will be interoperable, according to details provided in this DMP.

What data and metadata vocabularies, standards, or methodologies will you follow to make your data interoperable?

IUPAC nomenclature will be used as well as International Standards and metric units.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

Not applicable.

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?

Not applicable.

4.4   Increase data re-use (through clarifying licences)

How will the data be licensed to permit the widest re-use possible?

Open-access data will be licensed following Creative Commons 4.0, for which partners give the public the right to use a work protected by copyright. Users will be allowed to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material) but giving appropriate credit, providing a link to the license, and indicate if changes were made to the data.

Third parties can use these data for research purposes as state-of-the-art, to avoid duplication of efforts and as the basis for future investigations and research on the topic. For example, the generated codes for simulations can be used to support the discovery of new high-performance solid materials and electrolytes combinations. Also, new materials preparation methodologies for porous materials could be implemented in companies acting as materials suppliers.

If a publication is going to be prepared using data generated from UncorrelaTEd activities, the partner will first wait for the final acceptance of the manuscript. Then, the data will be uploaded to public repositories at a reasonable time after the decision. For open data not related to publications, no embargo period is expected to be set.

Nevertheless, the commercial use of part of the data generated by the project will be limited if any patent or exploitation agreement is expected to be filled or signed by the Consortium members.

The data will remain re-usable in open repositories after the end of the project. ZENODO repository will host the data generated along with the project life and beyond.

Reputation and trajectory of partners involved in UncorrelaTEd assure the reliability and quality of data generated. Any other expert laboratory will be able to replicate the experiments and validate the quality of the dataset.

5   Allocation of resources

No associate costs are expected from making data FAIR in UncorrelaTEd. Platforms selected to make data FAIR offers their services free of charge. For those articles published in open access, a part of the budget has been reserved, and they can be claimed as eligible costs.

The owner of the data will be the partner(s) who generate them. Moreover, as can be seen in Table 1, in any WP data will be produced in collaboration with two or more partners. Therefore, these partners will be responsible for the data management in conjunction with the Coordinator (if it is not already an owner of the data).

As agreed in the Consortium Agreement, each partner will provide all generated data to be published as Open Data to the Coordinator every six months. Other actions related to data management will be notified to the Coordinator. Moreover, each partner will provide, promptly,

data needed for the initiation, progress and completion of subsequent tasks established in the project.

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

Resources for long term preservation of data are not needed. All data will be included in open or institutional repositories free of charge. Then, data linked to open-access publications only have costs associated to the moment of publication. Also, the data available on the UncorrelaTEd website will be available after the finalisation of the project, since UJI will host the project website.

6   Data security

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

For safe storage and transfer of sensitive data, all partners will store a remote copy of the data, which will be kept in a cloud-based storage system (Intranet). The intranet is based on Google Drive. It is accessible from the project website, for which only the project members/beneficiaries will have access for visualisation. To prevent problems and improve the organisation of the folders, only the Coordinator will have access and will introduce the data and metadata when asked by the other beneficiaries. Moreover, each partner would regularly back up and store the data in an external Hard Disk, which will be used only for research purposes and will be located in the partners' premises.

The computers (desktop and stand-alone), including those used in the laboratories for measurements, must have a strong-password methodology, antivirus, and firewall regularly updated. Mainly, rules for the password, strength, renovation, and forgot password verification for all partners are as follows:

When using a public or non-personal computer, the user must log out after finishing the activity. The password for institutional accounts is personal and non-transferable. The password must have a minimum of 6 characters, and it cannot be based on a dictionary word. Personal computers (managed desktop and stand-alone) should be equipped with virus and firewall protections, which are regularly and automatically updated. All data will be stored and managed in compliance with current regulations and policies.

However, the following partners will follow these additional rules:

Is the data safely stored in certified repositories for long term preservation and curation?

Laboratory notebooks are handwritten or digital files and include in sequentially marked pages experimental descriptions, drawings, calculations, text, plots, or different ideas. Each entry is marked with the date and follows chronological order. The aim is to collect protocols,

methodologies, step-by-step procedures, results, and other related information to laboratory work written down in notebooks must be monthly digitalised and backed up by the owner of these results, as explained before. Confidential documents must be appropriately destroyed by a company if available; otherwise, each partner is responsible for ensuring their appropriate destruction.

It is not envisaged privacy issues concerning the data because there are not personal data involved.

7   Ethical aspects

This project does not handle personal data and, therefore, does not involve ethical issues to be managed.

8   Additional considerations

The DMP will be updated throughout the project, as a minimum in time with the periodic evaluation of the project (M12, M30, and M48). The DMP needs to be updated with significant changes, such as (but not limited to):

- new data,
- changes in consortium policies (e.g., innovation potential, decision to file for a patent),
- and changes in consortium composition and external factors (e.g., new consortium members joining or old members leaving).