Project Delivarble 1.2: Plan for Data Management

Deliverable information	
Code	D1.2
Due date	Project month 6/31 January 2023
Delivery date	Project month 6/31 January2023
Work package number and name	1 Project coordination and management
Work package leader	TFNS
Dissemination level	PU
First draft author (institution)	Nataša Đurišić-Mladenović (TFNS)
Reviewers	Nikola Maravić Biljana Pajin Zita Šereš
Pre-approval check for ethics	Dr. Dubravka Vejnović, Independent Ethics Advisor
Accepted by	Project Steering Team
Date of approval	10 January 2023
Version	1.0

Document revision history		
Issue date	Version	Comments

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EU executive agency. Neither the European Union nor the granting authority can be held responsible for them.

Abbreviations

CECs	Contaminants of Emerging Concern
CSIC	Spanish National Research Council, Institute of Environmental Assessment and Water Research
DMP	Data management plan
GA	Grant agreement
GC-MS	Gas chromatograph with mass spectrometer
FAIR	Findable, accessible, interoperable, reusable
FTIR	Fourier-transform infrared spectroscopy
PU	Public dissemination level
RMAU	Research management and administration unit
SEN	Sensitive dissemination level
STSE	Short term scientific exchange
TFNS	Faculty of Technology Novi Sad
UHPLC-HRMS	Ultra high performance liquid chromatograph with high resolution mass spectrometer
UHPLC-MS/MS	Ultra high-performance liquid chromatograph with triple quadruple mass spectrometer
UNL	NOVA University Lisbon, NOVA School of Science and Technology
WP	Work package

Table of content

TwiNSol-CECs, 101059867
WP1 Project coordination and management
D1.2 Plan for Data Management

GA101059867

TWINNING FOR EXCELLENCE IN
ENVIRONMENTAL RESOURCES PROTECTION

Funded by the
European Union

Executive Summary

This document is entitled "Plan for Data Management" in accordance with the project description in GA, and it is fully equivalent to the document usually referred as Data Management Plan or shortly DMP, and all these terms are used equally interchangeably throughout the document.

The Plan describes the data management life cycle for all data to be collected, processed and/or generated by TwiNSol-CECs:

• data handling during and after the project,

• types and formats of data generated/collected,

• methodologies and standards applied,

• shared and open access publications,

• data curated and preserved.

The structure of the document is created in accordance to the Horizon Europe Data Management Plan Template, version 1.0 from May 05, 2021, and it contains a Data Summary for the TwiNSol-CECs project (Section 1), explains the adopted FAIR data principles (Section 2) and other research outputs (Section 3), it considers the allocation of resources (Section 4), data security (Section 5), ethical aspects (Section 6) and other issues (Section 7). As DMP is a fully open access document, automatically posted online, in order to make it as a guide for the readers to the requested elements of DMP, the original questions given in the Template are kept.

First version of DMP of the TwiNSol-CECs is submitted as D1.2 deliverable in the 6th project month. DMP is an editable document, which will be updated through the project's life cycle as new data and other information occur.

1. Data Summary

As Contaminants of Emerging Concern (CECs), which are in the very core of TwiNSol-CECs research, and their presence in the environment are not controlled nor regulated, the collected experimental data on the CECs environmental presence and the efficiency of their removal by tested methods will be evaluated by comparison with data published elsewhere. Methods used for determination of CECs in various samples are going to be validated and compared with similar methods reported to be used for the CECs analysis. The efficiency of tested methods for CECs removal from water will be also compared with available relevant data to assess their applicability. If the risk assessment based on the research data would be performed, the existing, literature-based data and models (if any), needed for such assessment, will be used in compilation with the obtained data. The wide-range screening of CECs by UHPLC-HRMS will be based on using commercial software "*Thermo Scientific™ Compound Discoverer™*" and its libraries and data bases for small molecule identification.

What types and formats of data will the project generate or re-use?
TwiNSol-CECs will include:
- **generated research data** (e.g. numeric analytical results, MS spectra (scans), chromatograms, FTIR spectra, 2D-fluorescence, statistics, textual data on experimental procedures and working conditions, spreadsheets for sample coding and description, numeric data on methods' performance indicators, etc.),
- **re-use of literature-based research data** collected from peer-reviewed scientific publications,
- **use of libraries/data bases of commercial software** "*Thermo Scientific™ Compound Discoverer™*" for a CECs identification,
- **collection of personal data (not those belonging to special categories and respecting "data minimization principle") of researchers submitting the applications for participation in the project knowledge-transfer events** (trainings, summer schools, short-term scientific exchange (STSE), workshops, final conference),
- **generate instruction, mechanism, and/or acts** related to the strategic research plan and functioning of research management and administration unit at TFNS, and
- **photos and/or other media files** for documenting the project events.
The computer file formats, serving the various needs and common practices in research, dissemination, and communication, will be generated (primarily doc, xls, pdf, tiff, jpeg, etc.).

What is the purpose of the data generation or re-use and its relation to the objectives of the project?
Research data within TwiNSol-CECs are generated within **Work Package 4 - TwiNSol-CECs research projects: wide-range CECs' surveillance in the Serbian environmental resources and comparative study of selected innovative removal treatments**, under the leadership of **CSIC**. This WP consists of three research sub-projects:
- wide range CECs' screening in different samples from Serbia,
- exploring integrated approaches in membrane-based removal of CECs from water, and
- exploring innovative biomaterials in removal of CECs from water
The final purpose of these sub-projects is, respectively:
- to obtain the first data on wide-range screening of CECs in different samples from Serbia, serving as a basis for further spatial and temporal comparison towards possible prioritization,

Funded by the
European Union

- to generate new knowledge on possible application of membrane processes for selected CECs removal from water,
- to explore applicability of biomaterials from local biomass sources for selected CECs removal from water.

Thus, research data generated within these sub-projects will be used to fulfill the research objectives of TwiNSol-CECs; they will be compared, validated and/or harmonized with those relevant literature-based data, and as such they will be published for further dissemination, knowledge transfer and re-use.

As knowledge-transfer activities are important part of TwiNSol-CECs, particularly in WP3 (summer schools, trainings, STSEs) and WP6 (three workshops, final conference), applicants expressing the interest to participate in these events will be asked to provide limited personal data (not those belonging to special categories and respecting "data minimization principle") relevant and necessary for contacts with the applicants and for granting the participation with respect to capacity of an event venue; apart of professional CVs, collected data will contain names, title, affiliation, business postal and email addresses, phone number, and motivation letters. The contact details of data protection officer of each institution participating in TwiNSol-CECs will be provided to the participants.

The collected selected personal data (not those belonging to special categories), such as email addresses, will be included in the mailing list for receiving the project news and invitations for the project events if this is approved in correspondence related to the previous data submission.

The participants in the project events will be informed that photographs or electronic images will be taken during the events and that they may grant the permission and give the consent to the TwiNSol-CECs team for the use of such material for presentation under any legal use. Primarily these documents will be used for presentations, reports, newsletter, and other project communication activities.

Within WP2 and WP5 - packages dedicated to the enhancement of strategic research capabilities of TFNS, it is expected to create the Road Map for stepping up the TFNS excellence and Research Management and Administration Unit (RMAU) at TFNS, respectively, hence, the relevant documents (guidelines, act, etc.) are expected to be generated.

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

It might be assumed that the full size of data within the project will be up to 100 GB, which should be considered as the upper limit with considerable uncertainties.

What is the origin/provenance of the data, either generated or re-used?

Generated research data are experimental data obtained by analytical instruments such as UHPLC-HRMS, UHPLC-MS/MS, GC-MS, FTIR, or 2D-fluorescence, and lab equipment such as membrane, represented as numeric values, spreadsheets with mixed (numeric and textual) data, textual descriptions, images, etc. Generated research data will be subjected to the statistical evaluation and interpretation. Re-used data are those literature-based data relevant for comparison to assess the experimental data.

Personal data (not those belonging to special categories and respecting "data minimization principle") are those collected for trainees and participants in the project events such as summers schools, workshops, trainings, final conference, STSE. The collected data will be limited to only those relevant and necessary for contacts with the applicants and for granting the participation based on the relevant professional background; afterwards, the collected email addresses will be included in the mailing list for receiving the project news and invitations for the project events if this is approved in correspondence related to the previous data submission.

TwiNSol-CECs, 101059867
WP1 Project coordination and management
D1.2 Plan for Data Management

GA101059867

TWINNING FOR EXCELLENCE IN
ENVIRONMENTAL RESOURCES PROTECTION

Funded by the
European Union

The TwinNSol-CECs research data are useful for other researchers or experts who wish to replicate the results and the methodologies performed, and to stakeholders with interest to regulate and control CECs presence in the environment, as well as to innovate the CECs removal techniques.

The project communication documents publicly available are useful for other teams wishing to learn more about the Horizon Europe Twinning funded activities. These documents serve also to public awareness rising on the various CECs issues and the EU funding.

Documents that follow the creation of the TFNS research strategy and RMAU (as two specific objectives of the project in accordance with GA) are of institutional importance, having impact on the whole TFNS.

2. FAIR data

2.1. Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?
Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.
Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?
Will metadata be offered in such a way that it can be harvested and indexed?

TwiNSol-CECs will deposit the research data in an open online research data repository. TwiNSol-CECs has selected the ZENODO repository as its data archive follows the FAIR data principles. ZENODO allows researchers to deposit publications (articles, presentations, posters) and data (e.g., datasets, images), providing tools to linking them to these through persistent identifiers and data citations. Data are described with rich metadata. ZENODO's metadata is compliant with DataCite's Metadata Schema minimum and recommended terms. Metadata clearly and explicitly include the identifier of the data it describes; the DOI is a top-level and a mandatory field in the metadata of each record. Metadata of each record is indexed and searchable directly in Zenodo's search engine immediately after publishing. Metadata of each record is sent to DataCite servers during DOI registration and indexed there. Keywords are provided in the metadata to optimize the possibility for discovery and re-use.

Additionally, UNS Institutional Repository (if operable in the moment of first project publications) will be used for deposition of the project publications published with acknowledgement to TwiNSol-CECs. All researchers have personal pages on scientific social networking sites (e.g., LinkedIn, ResearchGate, Academia.edu) offering interested peer free access to non-sensitive data.

The project website (www.twinsol-cecs.com) will be also used as a repository of the project scientific articles, presentations, abstracts.

2.2. Making data accessible

Repository: Will the data be deposited in a trusted repository? Have you explored appropriate arrangements with the identified repository where your data will be deposited?
Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

TwiNSol-CECs has selected the ZENODO repository as its data archive follows the FAIR data principles. ZENODO is set up to facilitate the finding, accessing, re-using, and interoperating of data sets. It allows researchers to deposit publications (articles, presentations, posters) and data (e.g., datasets, images), providing tools to linking them to these through persistent identifiers and data

TwiNSol-CECs, 101059867
WP1 Project coordination and management
D1.2 Plan for Data Management

GA101059867

TwiNSOl-CECs

TWINNING FOR EXCELLENCE IN
ENVIRONMENTAL RESOURCES PROTECTION

Funded by the
European Union

citations. ZENODO registers DOIs (via DataCite) for all uploads/deposited records. Objects with pre-existing DOIs may be uploaded and the external DOI displayed; ZENODO allows including information on alternate persistent identifiers, as well as linking to related persistent identifiers.

The project website will contain the list of the project publications, including the research publications linked with the ZENODO repository, as well as the reports and deliverables declared as public in GA.

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible. Will the data be accessible through a free and standardized access protocol? If there are restrictions on use, how will access be provided to the data, both during and after the end of the project? How will the identity of the person accessing the data be ascertained? Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

Final research data will be made openly available, and they will be accessible through a free and standardized access protocol. Research data needed to validate the results in the scientific publications will be deposited in a data repository at the same time as a publication (Meta)data are retrievable by their identifier using a standardized communications protocol.

If an embargo is applied to give time to publish, it will be long as required by the publisher.

Currently, no beneficiary has the intention to close their data. If the situation changes, the DMP will be updated accordingly.

Restricted access will include the collected personal data of the applicants for the project events (as mentioned above, these personal data are not those belonging to special categories); this kind of data will be available only to the project event's organizers, i.e., selected members of the project team, for the purpose of the events organization.

Only the following project documents (deliverables) are declared in GA as being at sensitive level for dissemination, containing some internal data for restricted sharing (within the consortium, among the project team members, and with the EU funding body): D1.1-Project handbook, D1.5-Mid-term report on the WP1 meetings, D1.6-Final report on the WP1 meetings, D2.1-Report on development of Road Map for stepping up the TFNS excellence, D2.2-Report on the Road Map implementation, D3.1- Mid-term report on performed meetings with EU experts on new research arrangement, D3.2-Final report on performed meetings with EU experts on new research arrangement. These sensitive documents will be uploaded at the Funding and Tender Portal with designation SEN and they will be available to the consortium members at the internal page of the project website.

Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data? How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available? Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

ZENODO supports the DataCite Metadata Schema v4. The following additional article level fields are supported: journal title/volume/issue/pages, conference title/acronym/dates/pjlace/website, book

TwiNSol-CECs, 101059867
WP1 Project coordination and management
D1.2 Plan for Data Management

TWINNING FOR EXCELLENCE IN
ENVIRONMENTAL RESOURCES PROTECTION

GA101059867

Funded by the
European Union

publisher/place/ISBN/title/pages, alternate persistent identifiers. Content may be uploaded free of charge by those without ready access to an organised data centre.

All metadata in ZENODO may be freely used under the CC0 waiver.

Data files and metadata are backed up nightly and replicated into multiple copies in the online system.

Data and metadata will be retained for the lifetime of the repository; this is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.

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

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

Will your data include qualified references[1] to other data (e.g. other data from your project, or datasets from previous research)?

Common data and metadata standards and formats are a key aspect for technological and semantic data operability. Standardization makes the data discoverable. Data and metadata in ZENODO platform use a formal, accessible, broadly applicable language for knowledge representation, as well as vocabularies that follow FAIR principles.

Research data will include qualified references to other data including those from the project.

2.4. Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

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

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

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

Will the provenance of the data be thoroughly documented using the appropriate standards?

Describe all relevant data quality assurance processes.

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

For validation of final research data and facilitating data re-use, data on quality assurance/quality control will be available together with the research data in the same manner publicly available to enable widest re-use possible. Analytical data will be followed with standard validation data (e.g. recovery, precision, linearity, etc.).

---

[1] *A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)*

TwiNSol-CECs, 101059867
WP1 Project coordination and management
D1.2 Plan for Data Management

Funded by the
European Union

GA101059867

TWINNING FOR EXCELLENCE IN
ENVIRONMENTAL RESOURCES PROTECTION

For scientific publications in open access journals, the licensing schema of the publication repository should be followed. In most cases this it will be Budapest Open Access Declarations (CC-BY) schema thus TwiNSol-CECs will strive to adopt it also for collected/generated data when depositing in public repositories.

Publishing in peer-reviewed scientific journals will be used as the main mean of public disclosure of collected/generated research data. All researchers will ensure that the data underlying peer-reviewed publications will be made available for verification purposes after publication which is in accordance to accepted standards. After publication of research data in Open Access journals, the underlying data will be uploaded into suitable public and institutional repositories. All data ought to be disseminated through peer-reviewed publication will be made available at the time of publication or the latest after embargo period if requested.

3. Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g., software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).
Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

Currently, no other research outputs are foreseen in TwiNSol-CECs. If this change, the consortium will agree on the updates and new version of DMP.

4. Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g., direct and indirect costs related to storage, archiving, re-use, security, etc.)?
How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)
Who will be responsible for data management in your project?
How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

ZENODO offers free of charge, long-term (for the next 20 years, at least) open storage for research data.

Open Access scientific papers with the project research results will be covered with the available project budget line.

The project web domain has been purchased for 5 years, enabling accessibility pf the project reports and public deliverables 2 years after the project end, while its sustainability will be considered by TFNS team through indirect costs of follow-up projects.

TFNS project team members responsible for TwiNSol-CECs data management are: Project Coordinator, Project Manager for Strategic Issues, Communication Coordinator, and co-leader of WP4.

5. Data security

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?
Will the data be safely stored in trusted repositories for long term preservation and curation?

Research data, both intermediate and final, will be kept by the co-workers involved in the research at their own computers. The final research data uploaded in open repositories will be also kept at the internal page of the project website, but also at the password-protected computer (server) purchased within the TwiNSol-CECs, which is located at TFNS in the office of three project team members.

Collected personal data will be kept in electronic version as long as they are relevant and after that time they will be deleted. Approach to this kind of data is restricted to the project members with responsibilities in the relevant project event organization; such data are stored on password protected computers.

Sensitive documents (e.g., deliverables declared as SEN) will be kept at the internal page of the project website, where they will be clearly marked as sensitive with note about meaning of this level of dissemination. Recovery of such documents will be possible from the password-protected server computer purchased within TwiNSol-CECs.

6. Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?
These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).
Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?

Apart of research data, selected personal data (not those belonging to special categories) will be collected in relation to the participation in the project knowledge-transfer events (trainings, summer schools, short-term scientific missions, workshops, final conference). The collected data will be limited (respecting "data minimization principle") to only those relevant and necessary for contacts with the applicants and for granting the participation based on the relevant professional background described in the submitted CV during the application; apart from professional CVs, collected data will contain names, title, affiliation, business postal and email addresses, phone number, and motivation letters. As some of these knowledge transfer activities are planned to be organized at CSIC and UNL it can be expected that limited personal data (names, titles, institution, postal and email addresses, professional expertise) will be transferred to the EU partner institutions in compliance to General Data Protection Regulation (GDPR).

Signed lists of participants at the project trainings, workshops, summer schools, and conference, will contain names and institution details. The signed lists represent a means of verification for these events (as project milestones) and as such they will be provided if necessary to the EC. All collected data will be subject to appropriate safeguards and to free and fully informed consent of the persons concerned, and all involved persons will be made aware that they take part in the TwiNSol-CECs project. There will be no further processing of collected personal data.

7. 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 other procedures than the one described in this document will be used for data management.