



Project: **SEAWave**

Updated Data Management Plan, DMP

Work Package: WP11

Deliverable: D11.5

Deliverable No.: D43

Abstract

The Data Management Plan (DMP) of SEAWave was described thoroughly in Deliverable 11.2 (D40). After one year of its application and the trial use of Zenodo public repository this report includes a review of this usage period, updates for the data collection procedure, and harmonization actions within CLUE-H cluster, until November 30, 2023.

Project Details

Project name	SEAWave
Grant number	101057622
Start Date	01 Jun 2022
Duration	36 months
Scientific coordinator	Prof. T. Samaras, Aristotle University of Thessaloniki (AUTH)

Deliverable Details

Deliverable related number	D11.5
Deliverable No.	D43
Deliverable name	Updated Data Management Plan, DMP
Work Package number	WP11
Work Package name	Project and Consortium Management
Editors	Caterina Merla, ENEA - Italy
Distribution	Public
Version	1
Draft/final	Final
Keywords	Data Management, Data security, Open Research Data, FAIR Data

Contents

1	Updates, objectives and guiding principles	4
2	Summary on data type and structure, exchange and outputs	5
3	Record repartition.....	6
4	Challenge and future actions	7
5	Convergence and harmonization of DMPs in the CLUE-H cluster	8

1 Updates, objectives and guiding principles

SEAWave DMP deliverable D11.2 (submitted on 30/11/2022 and updated with the current deliverable on 30/11/2023) describes the data management life cycle of the project. As part of making research data findable, accessible, interoperable and reusable (FAIR), the submitted DMP included precise information on:

- the handling of research data during and after the end of the project
- what collected, processed and/or generated
- methodology and standards applied
- policy for data shared/made open access
- policy for data curation and preservation (also during and after the end of the project).

The above points were all addressed in detail in the submitted DMP and this document can be found online at the project website where it is accessible to the public. Further updates to this DMP are provided in this report.

The DMP objectives outlined in D11.2 deliverable are to:

- gather evidence of 5G signal monitoring and exposure assessment in different environments (urban and industrial) as well as its safety concerning carcinogenicity highlighted performing human, *in vivo* and *in vitro* experiments;
- make the records for such methods and experimental evidence consultable by project participants as well as, when possible, to the wider research and general public community;
- make the experimental procedures for the human, *in vivo*, *in vitro* testing along with the protocols for biological testing as widely accessible as possible;
- enforce a metadata recording standard which is common to all SEAWave participants and can be potentially adopted and/or improved by the wider community of the CLUE-H (a cluster of four projects involved in 5G studies for exposure and safety assessments).
- provide evidence that the SEAWave participants are committed to Open Access publication and dissemination practices.

The selected tool for meeting the outlined objectives is the OpenAire supported **public repository Zenodo** (<https://zenodo.org>) and the flexible licensing capability, which is therein embedded to

gather all SEAWave records in areas with different levels of open access compliance. Although a minority of records were deposited under embargoed and restricted access conditions, SEAWave's commitment remains to transfer as much data as possible to the open area at the earliest convenience.

The D11.2 report outlines the utilization of the Zenodo platform as the central account for the SEAWave project, under the management of the SEAWave data manager. Additionally, it details the development and implementation of a metadata recording program, which was subsequently uploaded as a record through a Zenodo Community named SEAWave. In this system, individual collaborators maintain separate Zenodo accounts, retaining ownership of their respective records. A dedicated SEAWave Zenodo community has been established, complete with a shared uploading link

(https://zenodo.org/communities/seawave_data_managing?q=&l=list&p=1&s=10&sort=newest).

Consortium members have been duly informed and notified about this community, enabling them to contribute records through the provided common uploading link. The community collection, curation and harvesting are ensured and periodically verified by the SEAWave data manager.

2 Summary on data type and structure, exchange and outputs

The data in SEAWave are collected in text, numerical and image formats, following the associated DMP instruction, and they are stored in files whose extension are defined by the equipment/software used for the generation and the collection. Some examples include, but are not limited, to .xlsx (Excel spreadsheet), .ppt (PowerPoint), .mat (Matlab), .txt (text), .s2p (touchstone), .avi (video). Since either individuals or groups of researchers from the consortium partners will generate data, the identification of the partner that created the specific file or data set is also included in project's DMP.

Data exchange within the SEAWave consortium will be managed using the Zenodo repository, as above described, whereas data transfers between SEAWave partners working in the same WP will take place using each partner's institutional repositories, where a virtual area for data sharing will be created and used following the established procedures. In all cases, the final data and procedures will be transferred within the SEAWave Zenodo repository following the file naming which is described in the project's DMP. Within the SEAWave project, the primary objective of data collection and generation is to obtain evidence of the quantification of 5G exposure. This involves the development of new tools and instruments tailored specifically for this purpose. The

overarching goal is to generate fresh scientific data that can be utilized to assess, in a robust manner, the potential health risks associated with 5G-FR2 exposures.

Additionally, data from different environments (ex. urban areas and smart industries) will be collected in order to permit the assessment of exposure in real 5G scenarios to establish appropriate standards and validate the results coming out from the biological investigations *in vitro*, *in vivo* and on human volunteers.

Therefore, the collected data will include both the experimental and theoretical procedures for characterizing exposure, along with the protocols for biological testing. Finally, simulation data and codes, such as those related to the automatic control of measurement instruments or dosimetric assessment, will be also collected.

3 Record repartition

To date, a total of 5 records have been published in the SEAWAVE Zenodo repository, all of which (100%) are available in the open access area. None of the records were placed under embargo thus far, and the expiry date is currently set to match with the end date of the project. The embargoed access was sought in order to allow the project management to consider appropriate intellectual property (IP) protection initiatives, and is scheduled to be periodically reviewed. The depositors will strive to reduce or remove the embargo status as soon as the achieved level of IP protection will be either deemed adequate or no longer necessary. The number of records placed under restricted access (limited to the consortium participants) are up to now equal to zero records. The restricted access area will encompass data that holds immediate relevance for the project participants, especially when potential IP issues are yet to be evaluated. A small fraction of records with restricted access was also uploaded for self-training and familiarization with the Zenodo ecosystem.

Among the currently published material 5 records are conference papers (including published papers and/or slides and/or posters presented during the conference), 0 passed or planned press releases/coverage examples; 0 workshop presentations, 0 experimental protocol procedures, 0 records include internal project meeting slides; 0 record include computer programs and measurements.

The reported distribution and prevalence of open access records indicate that data management has been pursued in full compliance with the DMP.

In order to promote the dissemination of the material published in open access and increase its impact by facilitating more efficient indexing in research database as Google Scholar or Research Gate, a dedicated dissemination webpage has been implemented on the SEAWave website

(<https://seawave-project.eu/seawave-dissemination/>) including a list of published papers and conference presentations, each with a direct link to the associated material (papers, slides, posters, etc.) on either the corresponding Open Access journal website or Zenodo.

4 Challenge and future actions

The main challenge to the effective implementation of a DMP resides in human nature and time constraints: researchers are generally more focused on meeting the demanding technical goals of the project than on optimal data collection. Furthermore, the subtleties in proper metadata creation are often hard to perceive at the time data is being generated. The accuracy of the records, and the consistency of the recording practices, both over time and for different depositors, also appear as formidably challenging aspects. This is exacerbated by the fact that modifying published records on Zenodo is not a straightforward process. Finally, it is difficult to strike the right balance between a centralized approach where a unique depositor uploads and verifies the accuracy of most records (which results in huge workloads unevenly distributed), and a distributed approach where multiple collaborators take responsibility for record submission (which results in poor homogeneity).

The areas of improvement identified for the SEAWave data management effort include:

- increasing the number of published records;
- periodically assessing and improving their accuracy/consistency;
- oversee transfer of restricted and embargoed access record to the open access area where IP consideration permit;
- educate and train the fellow project participants such that the publishing workload is more evenly distributed and they gain insights into better data preservation strategies.

A further area of improvement pertains the very nature of the chosen open access repository: Zenodo is incredibly powerful in terms of gathering vast amounts of heterogenous content. However searching for some specific data can be difficult at times. The adopted record naming convention, with the creation of lengthy metadata strings including essential information on a particular record, may appear somewhat unfriendly to the non-trained user. As such, a more intuitive index of the stored records is being compiled which will allow for more direct research of content, especially when the user has an idea of the record type, and will be periodically uploaded/updated on the Zenodo platform.

5 Convergence and harmonization of DMPs in the CLUE-H cluster

An important task in data managing is the harmonization of procedures and data storage among the four projects funded by EU on 5G grouped into the CLUE-H Cluster. As these four projects have different outcomes in terms of experimental activities, exposure assessment procedures, output data and tasks, a total harmonization of data format and type is at least impossible to achieve. This option would result in a complex and time-consuming effort with a limited practical interest.

Hence, what was established, by CLUE-H Working Group 2 (WG2) work (instituted among all the projects to coordinate data management), is to harmonize only metadata files in order to have a common and homogenized description of the different type or format of data, if possible. Then, these data will be accessible on each single repository in accordance with what reported on the single DMP of each project (D11.2 in SEAWave). The total exchange of data would have resulted in an extremely complex procedure and in some cases this would have been even not possible for ethical issues of privacy and confidentiality of data related, for example, in human trials.

It is important to set the different kind of outputs and defined common metadata. WG2 will develop a common format to collect metadata from the individual projects, in order to enhance FAIR principles across the projects. The metadata will be updated regularly and published on the CLUE-H website (<https://www.emf-health-cluster.eu/>). Metadata will include a short description of the type of data, format, size, accessibility, ownership, and contact persons.

Currently, no unified data storage of individual project's results are planned in the cluster. Data storage in CLUE-H pertains to CLUE-H activities (e.g. policy briefs, meta-data, inventories e.g. of measurement devices and approaches, etc.). The corresponding information will be provided on the CLUE-H website, where it will be regularly updated. In addition, metadata containing file and data set information will be available in the specific repository of each project under the established conditions of each single consortium i.e. full open access, with embargo or confidential, as previously described for SEAWave (D11.2 and update).

The unified data framework in CLUE-H will be developed concerning metadata of the individual projects. For this aim, a harmonized form will be used and the collected results will be made accessible in a searchable way on the CLUE-H website in a timely manner. WG2 group discussed and established the creation and organization of a common metadata template usable for the four projects. This task will be performed under the Common European Research Information Format (CERIF) standard. Different types of data have to be taken into account for the creation of the metadata table as .xlsx (excel spreadsheet), .ppt (power point presentation), .mat or .m (Matlab related file extension), .txt (text file), .s2p (touchstone), .avi (video files) and others.

Metadata will have specific fields to be filled depending on the specific type of file. The different fields required for each type of file are reported in the next Table. Once the metafile has been created the data set identifier has to be generated using the following organization:

Project”_”Date”_”Time”_”Name”_”Type”_”Extension”_”Place”_”Creators”_”Target user”_”Other”.

“Project” is the name of the specific project at which the metadata belongs so for example SEAWave. Date is the date in format “YMMMDD”, “Time” is the time of metadata creation in the format “HHMMSS”. “Name” is a short name of the data eventually with the specification of WP in which these data were generated. “Type” describes the type of data as publication, measured data, simulation data, protocol description or others. “Extension” describes the data file extension. “Place” describes the location where the data were produced. “Creators” defines the individual(s) who generated the data. Finally, “Target user” defines the target audience of the data, if known.

All this information with details for each project was published in DMP of the Cluster released at the end of August 2023 and public available in the cluster web site.