LABORATÓRIO NACIONAL
DE ENGENHARIA CIVIL

DATA MANAGEMENT PLAN FOR THE MOSAIC.PT PROJECT

REPORT 342/**2020 – DHA/GTI**

LABORATÓRIO NACIONAL
DE ENGENHARIA CIVIL

DATA MANAGEMENT PLAN FOR THE MOSAIC.PT PROJECT

Fundação para a Ciência e a Tecnologia

Lisbon • October 2020

R&D HYDRAULICS AND ENVIRONMENT

REPORT 342/**2020 – DHA/GTI**

Título

DATA MANAGEMENT PLAN FOR THE MOSAIC.PT PROJECT

Authors

HYDRAULICS AND ENVIRONMENT DEPARTAMENT

Anabela Pacheco de Oliveira
Senior Researcher, Information Technology in Water and Environment Research Group

Paula Maria de Santos Freire
Assistant Researcher, Estuaries and Coastal Zones Unit

Collaborations

SCIENTIFIC INSTRUMENTATION CENTRE

Joaquim Neto
IT Expert, Head of the Computer Systems Unit

AV DO BRASIL 101 • 1700-066 LISBOA

e-mail: lnec@lnec.pt

www.lnec.pt

Report 342/2020

File no. 0602/111/2107001

DATA MANAGEMENT PLAN FOR THE MOSAIC.PT PROJECT

Abstract

This document presents the Data Management Plan (DMP) for the MOSAIC.pt project. MOSAIC.pt has chosen to participate in the Open Research Data pilot from FCT. Following the recommendations provided by the European Commission (EU, 2016), in the scope of making data FAIR, this Data Management Plan (DMP) provides the approach to the following topics:

- 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;
- how data will be curated and preserved (including after the end of the project).

Keywords: Data Management Plan

PLANO DE GESTÃO DE DADOS PARA O PROJETO MOSAIC.PT

Resumo

Este documento apresenta o Plano de Gestão de Dados do projeto MOSAIC.pt. Este projeto participa no piloto de dados abertos em investigação da FCT: seguindo as recomendações da União Europeia (EU, 2016) para tornar os dados FAIR, este Plano de Gestão de Dados (DMP) aborda os seguintes tópicos:

- o tratamento dos dados de investigação do projeto durante e após o mesmo;
- a identificação de que dados serão coletados, processados e/ou gerados;
- que metodologias serão usadas na sua recolha e em que standards estão inseridos;
- que dados serão abertos e que dados serão confidenciais;
- como serão armazenados os dados e mantidos no futuro.

Palavras-chave: Plano de gestão de dados

Table of contents

List of tables

1 | Introduction

This document is developed as part of the MOSAIC.pt - *Multi-source flood risk analysis for safe coastal communities and sustainable development* project, funded by Fundação para a Ciência e a Tecnologia (FCT) with the reference PTDC/CTA-AMB/28909/2017. The Data Management Plan (DMP) constitutes the Activity 8 Milestone 8.2 Intermediate Data Management Plan. The project MOSAIC.pt has chosen to participate in the Open Research Data pilot of FCT.

The project MOSAIC.pt includes ten Activities from which seven of them (Table 1.1) will result in relevant information to be shared with the research and end-user communities. The purpose of the DMP is to support the data management life cycle for all data that will be collected, processed or generated by the MOSAIC.pt project. Following the recommendations provided by the FCT and the European Commission, in the scope of making data FAIR (findable, accessible, interoperable and reusable), this DMP provides an initial approach to the following topics:

- 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
- how data will be curated and preserved (including after the end of the project)

More developed versions of the Plan may be released at later stages, if updates on the data management occur.

Table 1.1 – Summary of data-producing work packages (WP)

WP Name	Partner in charge of this WP
WP 1 - Critical typologies selection through preliminary flood evaluation	LNEC
WP 2 - Numerical modeling of coastal inundation	LNEC
WP 3 - Vulnerability dimension assessment	CES
WP 4 - Development multi-source monitoring and predictive methodology	LNEC
WP 5 - Risk framework design	LNEC
WP 6 - Institutional capability for emergency response	CES
WP 7 - Safe communities involvement and response	CES

The MOSAIC.pt DMP was prepared using the Digital Curation Centre (DCC) DMP online tool (https://dmponline.dcc.ac.uk/), which provides a DMP template that match the demands and suggestions of the Guidelines on FAIR Data Management in Horizon 2020 (EU, 2016).

2 | Data summary

The main objective of MOSAIC.pt is to develop an innovative flood risk management framework for coastal zones based on the integration of predictive models and real-time monitoring data, and taking into account the different dimensions of the vulnerability. Project approach includes a preliminary evaluation of the most critical coastal typologies in the Portuguese coast based on historical information. This evaluation allows the selection of an observatory or observatories that will be used as testbed for development and validation of the predictive, monitoring and data integration methodologies, and where the different dimensions of vulnerability will be evaluated. Throughout the project, data will be produced and reused to support the development of research and innovation activities focusing on: developing a database of flood occurrences in the Portuguese continental coast and defining the critical coastal typologies (Activity 1); developing flood predictive models for different coastal typologies (Activity 2); assess the vulnerability in the observatory considering its different social and territorial dimensions (Activity 3); developing a multi-source methodology for coastal flood hazard prediction aggregating monitoring information and models prediction (Activity 4); designing a comprehensive flood risk framework to support emergency planning and response (Activity 5); improving the response capacity of the institutions responsible for the emergency management with the design of a comprehensive real-time emergency response system (Activity 6); evaluating the capability of emergency response of the community and development of guidelines of how should be prepared (Activity 7). Table 2.1 summarizes the types of data produced / made available in different work packages of the MOSAIC.pt project. Table 2.2 presents a more detailed description of data and its use (or re-use), formats, standards and metadata, data availability (open, confidential) and the expected resources to store, curate and preserve (size, backup frequency, resources to maintain it after project ends, repository choices for open and private data).

Table 2.1 – Summary of types of data to be used / created in different Activities (ACT)

TYPE OF DATA		ACT
In-situ and remote sensing data	GNSS beach topography, drone point cloud, orthomosaic and digital surface model	1
	Water level and wave parameters times series	2
	Remote sensing image database from Satellites	2,4
	Remote sensing image database from cameras	4
Hindcast/Forecast modeling results	Hydrodynamic model outputs	2,4
	Morphodynamics mode outputs	2,4
	Overtopping formulas output	2,4
	Inundation maps	2,4
	Time series of model results at selected data points	2,4
	Digital terrain model outputs	1
Vulnerability information	Environmental data from satellites and other open sources	3
	Socioeconomic public reports analysis	3
	In situ vulnerability assessment	3

Página intencionalmente deixada em branco

Table 2.2 – Description of data to be used / created in different work packages, including standards and metadata, availability, storing resources for open and confidential data

Partner	WP	Type of data/ data description	Methodology of data production	Data documentation and how that will be made available	Data standards	Which data sets will be classified as Open Access	Which data sets will be classified as confidential	In case of datasets that are not shared, reasons for that	To whom data classified as Open Access could be useful	How will open data be shared / by whom and where	How will confidential data be archived / by whom and where	How is backup and versioning realised?	When data will be produced	When data will be placed in Open Access	How will the data be disseminated	How will data be available after the end of the project
LNEC	ACT 1 (task A1.3)	Beach topographic surveys	Field data acquisition and data analysis	Technical document available in the project's webpage	Not applicable	All data is open access	None	Not applicable	Researchers, end-users, service providers	Can be shared to all through MOSAIC.pt repository at ZENODO	Not applicable	Process done using ZENODO's own procedures	Data will be produced during ACT1 and ACT4	At the end of the project	Technical document available in the project's webpage	Through publication and available through ZENODO during the period that this repository offers for H2020 projects
	ACT 2 (task A2.1)	Series of topo-hydrographic data of the surf-beach-dune zone during maritime storm events	XBeach morphodynamic numerical model	Technical document available in the project's webpage	Not applicable	All data is open access	None	Not applicable	Researchers, end-users, service providers	Can be shared to all through MOSAIC.pt repository at ZENODO	Not applicable	Process done using ZENODO's own procedures	During the development of the Task	At the end of the project	Technical document available in the project's webpage	Through publication and available through ZENODO during the period that this repository offers for H2020 projects
	ACT 2 (task A2.1)	Model outputs from SCHISM-WWM	SCHISM-WWM numerical model	Publications	netcdf	All data is open access	None	Not applicable	Researchers, end-users, service providers	Can be shared to all through MOSAIC.pt repository at ZENODO	Not applicable	Process done using ZENODO's own procedures	During the development of the Task	At the end of the project	Publications	Through publication and available through ZENODO during the period that this repository offers for H2020 projects
	ACT 4 (task A4.2)	Environmental data extracted from satellites and Airborne images	Remote sensing and field data acquisition	Technical document available in the project's webpage	Not applicable	All data is open access	None	Not applicable	Researchers, end-users, service providers	Can be shared to all through MOSAIC.pt repository at ZENODO	Not applicable	Process done using ZENODO's own procedures	Data will be produced during ACT4	At the end of the project	Technical document available in the project's webpage	Through publication and available through ZENODO during the period that this repository offers for H2020 projects

LNEC	ACT 4 (task A4.1)	Environmental data in-situ sensors	field data acquisition	Metadata document available at the project's webpage	csv	All data is open access	None	Not applicable	Researchers, end-users, service providers	Can be shared to all through MOSAIC.pt repository at ZENODO	Not applicable	Process done using ZENODO's own procedures	Data will be produced during ACT4	At the end of the project	Technical document available in the project's webpage	Through publication and available through ZENODO during the period that this repository offers for H2020 projects
	ACT 4 (task A4.1)	Model forecast outputs from SCHISM-WWM and Xbeach	SCHISM-WWM and Xbeach numerical model	Metadata document available at the project's webpage	netcdf	All data is open access	None	Not applicable	Researchers, end-users, service providers	Can be shared to all through MOSAIC.pt repository at ZENODO	Not applicable	Process done using ZENODO's own procedures	During the development of the Task	At the end of the project	Publications	Through publication and available through ZENODO during the period that this repository offers for H2020 projects
LNEC	ACT 4 (task 4.3)	Beach drone survey	Photogrammetric survey with drone and data processing	Technical document available in the project's webpage	Not applicable	All data is open access	None	Not applicable	Researchers, end-users, service providers	Can be shared to all through MOSAIC.pt repository at ZENODO	Not applicable	Process done using ZENODO's own procedures	Data will be produced during ACT4	At the end of the project	Technical document available in the project's webpage	Through publication and available through ZENODO during the period that this repository offers for H2020 projects
LNEC	ACT 2 (task 2.1)	Results from empirical formula, neuronal tools (NN_OVERTOPPNG2) or numerical models (SWAN, DREAMS, XBEACH or others)	Numerical results and data analysis	Technical document available in the project's webpage	Not applicable	All data is open access	None	Not applicable	Researchers, end-users, service providers	Can be shared to all through MOSAIC.pt repository at ZENODO	Not applicable	Process done using ZENODO's own procedures	Data will be produced during ACT2	At the end of the project	Technical document available in the project's webpage	Through publication and available through ZENODO during the period that this repository offers for H2020 projects
CES	ACT 3 (task 3.1)	Digital Terrain Model using the beach drone survey	Numerical processing of elevation data from the survey with drone	Technical document available in the project's webpage	Not applicable	All data is open access	None	Not applicable	Municipal and local decision makers, researchers	Can be shared to all through MOSAIC.pt repository at ZENODO	Not applicable	Process done using ZENODO's own procedures	Data will be produced during ACT3	At the end of the project	Technical document available in the project's webpage	Through publication and available through ZENODO during the period that this repository offers for H2020 projects
	ACT 3 (task 3.1)	Social Vulnerability data built from social and demographic data retrieved from the 2011 Population Census	Statistical analysis of Census Data at the statistical block and sub-block	Technical document available in the project's webpage	Not applicable	All data is open access	None	Not applicable	Municipal and local decision makers, risk practitioners, researchers	Can be shared to all through MOSAIC.pt repository at ZENODO	Not applicable	Process done using ZENODO's own procedures	Data will be produced during ACT3	At the end of the project	Technical document available in the project's webpage	Through publication and available through ZENODO during the period that this repository offers for H2020 projects

3 | FAIR data

3.1 Making data findable, including provisions for metadata

MOSAIC.pt will produce and reuse a variety of data types, from images to time series and georeferenced information in GIS format, covering a broad range of areas (remote sensing, hydrodynamics, morphodynamics, territorial vulnerability). Metadata will be produced for all data generated in the project, using standards when available. Standards such as ISO19115 (http://rd-alliance.github.io/metadata-directory/standards/iso-19115.html) or the OGC Sensor Observation Service (SOS) Interface Standard (http://rd-alliance.github.io/metadata-directory/standards/observations-and-measurements.html) are expected to be adopted. Consistency between metadata for similar data sets will be sought when standards are not available. Elements to be included in the metadata include a clear description of the data, the institution and person of contact responsible for the data creation, its format, creation date and possible modifications, data units and georeferencing (when applicable) and a number of keywords (metatags). The choice of adequate keywords will be included to promote and ease the discoverability of data. These keywords will include a number of fixed, common keywords in MOSAIC.pt's scientific area and several new, free keywords that can help attract researchers from other areas to use and adapt MOSAIC.pt's results to their scientific fields.

For all open data in the project, ZENODO will be used as the project's open data repository. ZENODO provides Digital Object Identifiers for all data sets, thus guaranteeing that all open data in WADI will have persistent and unique identifiers.

For consistency and promotion of data discovery, consistent naming conventions will also be used and agreed among the partners (to be defined later).

Open access publication will also be sought, with direct links to the underlying data sets deposited in ZENODO.

LNEC will be responsible for uploading data and other items in ZENODO, through the project designated data manager. Each partner/group will provide the datasets and publications to be integrated in the ZENODO repository to the data manager, dully informed on its access policy. The datasets and the expected time of availability and access policies are described in Table 1.2. For publications subject to embargo periods (due to the publishers' policies), the data manager will upload them in ZENODO as soon as the embargos are finished.

In the scope of the surveillance system developed in MOSAIC.pt, a succession of data sets will be produced, creating several databases of images at different stages of development and processing, from the raw data from cameras to the processed and quality-certified images included in the end-user application. This sequence can be labelled as several versions of a single dataset or it can be identified and managed as different datasets. Regardless of the approach chosen by the partners for this data, a clear versioning policy will be adopted and linked with detailed metadata and supporting documentation.

3.2 Making data openly accessible

MOSAIC.pt will create or reuse a variety of data sets, which all will be freely available at least at the end of the project, through deposit at Zenodo (unless their size prevents that archiving choice). Deposits in ZENODO will include the data, their metadata and their documentation. For most data sets, access is granted through generalized use software such as ArcGIS or similar.

3.3 Making data interoperable

All data developed in MOSAIC.pt will be fully documented and accompanied with detailed metadata supported by a set of select keywords, to facilitate automatic discovery and integration of data for other purposes. Besides usual metadata fields, technical aspects such as units (complying with SI standards) and spatial and temporal references will be supplied. All data will be provided in generally used extensions, adopting well established formats (csv, shapefiles, image formats,...) whenever possible which will also facilitate its use by other parties.

3.4 Increase data re-use

Open data availability will occur as soon as possible in MOSAIC.pt while respecting the team's publication targets. Typically, open data will be available for publication in ZENODO at the end of respective WP (Table 1.2), and its publication will occur within a few months. The team expects that this fast publication of data created in MOSAIC.pt will promote its reuse by other researchers and end-users, thereby contributing to the dissemination of its methodology and tools.

4 | Allocation of resources

A repository in ZENODO was created (https://zenodo.org/communities/mosaicpt/?page=1&size=20) for the projects´ open data, therefore ensuring data availability, backup and versioning. Long-term preservation will be guaranteed for the lifetime of the ZENODO repository (https://zenodo.org/policies). This is currently the lifetime of the ZENODO's host laboratory CERN, which currently has an experimental program defined at least for the next 20 years. After the end of that period, the data will only be kept at the data owners' servers and repositories.

Publications featuring the data will be produced in the project (specifically by the research partners) and will be made available through open access (using open access journals or journals selected for a short embargo period). This channel will provide a long-term availability of data and data analysis.

The methodology and data products developed in the project are expected to have a large impact in this field, through generalized application to other coastal sites. Moreover, we expect that the scientific and technological developments may serve as a basis in other fields of application, with the data demonstrating and promoting MOSAIC.pt's exceptional quality of service. The long-term usefulness of the data collected during MOSAIC.pt is guaranteed as coastal evolution relays in long term monitoring which will be obtained both for satellites, video cameras and low cost sensors.

The DMP will be updated throughout the project if important changes in data management occurs.

5 | Data security

Open data security will be addressed in MOSAIC.pt taking advantage of ZENODO's services of secure storage, backup and preservation and protected transfer mechanisms. Regarding the confidential data, different approaches will be used by each data-owner institution/group, but common rules apply. Data will be housed on servers under direct management of the institution's personnel to be installed in already provisioned data centres. These data centres are expected to be equipped with various features ranging from secure physical access, air conditioning, generators and fire extinguishing measures. Typically, hardware / electricity failure are addressed with redundant hardware and generators.

Taking into account the size of the data at stake that requires regular backup (be it either for security versus a hardware failure or for archival purposes), a sequence of regular full backups, differential backups and incremental backups on an increasingly frequent basis are envisaged, and following already installed procedures at the institutions. The physical media used to store the data will be maintained in secure locations. Access to these backups is limited to the personnel authorized to use the backup system, and as a general rule, not authorized for external sources.

All data transfers should be encrypted to render all stolen/lost data useless. Encryption methods are to be specified later.

6 | Conclusions

This document presents the Data Management Plan (DMP) for the MOSAIC.pt project. Its intent is to support the data management life cycle for all data to be collected, processed or generated by this project, aiming at making these data FAIR (Findable, Accessibility, Interoperability, and Reuse).

Lisbon, LNEC, October 2020

APPROVED

Head of the Estuaries and Coastal Zone Unit

Luís Portela

Head of the Hydraulics and Environment Department

Helena Alegre

AUTHORS

Anabela Oliveira

Senior Researcher

Paula Freire

Assistant Researcher

References

European Commission, 2016 – **H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020**. Version 3.0, July 2016. (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

www.lnec.pt/

AV DO BRASIL 101 • 1700-066 LISBOA • PORTUGAL
tel. (+351) 21 844 30 00 • fax (+351) 21 844 30 11
lnec@lnec.pt **www.lnec.pt**