Prodigio

Developing Early Warning Systems for improved microalgae PROduction and Anaerobic DIGgestIOn

Grant Agreement Number: 101007006

D7.3 DATA MANAGEMENT PLAN

MACARENA SANZ[1], ÁLVARO TAMAYO[1]

[1]ID-Consortium (IDC), Spain

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

Grant Agreement Number: 101007006
Project acronym: PRODIGIO
Project title: Developing early warning systems for improved microalgae PROduction and anaerobic DIGestIOn

Deliverable number:	7.3
Deliverable responsible:	IDC
Work-package:	7 Project Management
Editor:	Macarena Sanz / Álvaro Tamayo
Dissemination Level	

Author(s) – in alphabetical order		
Name	**Organisation**	**E-mail**
Macarena Sanz	IDC	msanz@idconsortium.es
Álvaro Tamayo	IDC	atamayo@idconsortium.es

Document Revision History			
Version	**Date**	**Modifications Introduced**	
		Modification Reason	**Modified by**
V.1	23/06/2021	NEW	Álvaro Tamayo
V.2			
V.3			
V.4			

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

Table of Contents

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020
RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

EXECUTIVE SUMMARY

This document outlines the data management strategies that will be implemented throughout the PRODIGIO research data lifecycle. In particular, it describes (i) the type, format and volume of the generated data, (ii) the metadata and documentation provided to make it findable, interoperable and reusable, (iii) the long-term preservation plan, (iv) how data will be shared and licensed for re-use, (v) the resources that need to be allocated to data management, (vi) data storage and back up policies during the active phase of the project, and (vii) the handling of personal data.

As stipulated in the Guidelines on FAIR Data Management in Horizon 2020, this DMP will be updated when important changes to the project occur and, at least, as part of the periodic reviews and at the end of the project.

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020
RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

1. THE PRODIGIO CONSORTIUM

The PRODIGIO consortium (***Table 1***), which is composed of university researchers, research centers and a consultant SME, is working together to achieve the project's common objective while stimulating an environment where each consortium partner shares and exchanges experiences to achieve the goals set-forth.

Table 1. *The PRODIGIO Consortium*

Logo	Name	Country
CSIC	Spanish National Research Council (CSIC)	SPAIN
AWI Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung	Alfred Wegener Institute (AWI)	GERMANY
ARMINES	Association pour la Recherche et le Développement des Méthodes et Processus Industriels (ARMINES)	FRANCE
IDconsortium Just research it!	IDConsortium SL (IDC)	SPAIN
institute IMDEA energy	Madrid Institute for Advanced Studies (IMDEA)	SPAIN
NMBU Norwegian University of Life Sciences	Norwegian University of Life Sciences (NMBU)	NORWAY
National Taiwan University	National Taiwan University (NTU)	TAIWAN
Universidad de Almería	University of Almeria (UAL)	SPAIN

2. INTRODUCTION

PRODIGIO aims to bring microalgae-based biogas faster to commercialization. To help in this endevour, the project aims to develop a system failure predictiom technology that allows a better monitoring and control of microalgae biomass production and biomass-to-biogas conversion systems.

In the PRODIGIO project data to assess the technological performance of these different methods will be collected. Additionally, the data related to the risk assessment of different techniques used, such as metaOmics extraction of two microalgae species from residues of different water sources, will be generated. The data generated will contribute to evidence-based development of new generation of non-polluting energy sources.

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

Table 2 provides a list of research data categories that will be produced within the PRODIGIO project and basic rules on its management. This document is available to all partners on the project's SharePoint platform and it will be continuously updated as new data is generated during project execution.

These data will not only serve to meet the objectives of the current project but will also be useful for stakeholders including the scientific community, industry and the general public.

Thus, the scientific community will benefit from the development of the strategies to maximize microalgae culture and exploit it capacity to generate biogas. The project will explore novelty techniques on early warning signals identification based on cutting edge time series generation. Finally, the generated data on the exploitation of microalgae as the next generation of non-pollutant energy source as well as the generated communication materials will help scientific community, legislators, and the general public make informed decisions on the regulation and public acceptance of these environmental-friendly resources.

To enhance the usability of the data, open or otherwise widely-used file formats will be the preferred option for data collection (see Table 1). Formats that are open and/or in widespread use stand the best change to be readable in the future; on the contrary, proprietary formats used only by a particular software are likely to becoming obsolete. In those cases in which the laboratory instrument used to perform the measurement outputs the data in an instrument specific proprietary format, a converted version of the output file to an open data format will be shared together with the original file to foster data interoperability.

Table2. *List of data to be produced during the course of the PRODIGIO project.*

Origin	Description	Format	Responsible	Confidentiality
Outdoor reactors sensors data ***Tasks 1.1 & 1.2 (WP1)***	*Time series data* Online routine data from sensors for light, temperature, pH, dissolved oxygen, air injection and CO_2 injection. Three independent time series, one for each experimental treatment (wastewater, fertilizers, seawater).	.xls	UAL	Public through deliverables as per the GA
Indoor reactors sensors data ***Tasks 1.1 & 1.2 (WP1)***	*Time series data* Online routine data from sensors for light, temperature, pH, dissolved oxygen, air injection and CO_2 injection. Three independent time series, one for each experimental treatment (wastewater, fertilizers, seawater).	.xls	UAL	Public through deliverables as per the GA
Outdoor reactors	*Time series data*	.xls	UAL	Public through

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

biological data ***Tasks 1.1 & 1.2 (WP1)***	Offline routine data from microalgae-bacteria culture systems (biomass concentration, absorbance, Fv/Fm). Three independent time series, one for each experimental treatment (wastewater, fertilizers, seawater).			deliverables as per the GA
Indoor reactors biological data ***Tasks 1.1 & 1.2 (WP1)***	*Time series data* Offline routine data from microalgae-bacteria culture systems (biomass concentration, absorbance, Fv/Fm). Three independent time series, one for each experimental treatment (wastewater, fertilizers, seawater).	.xls	UAL	Public through deliverables as per the GA
Outdoor reactors chemical data ***Tasks 1.1 & 1.2 (WP1)***	*Time series data* Offline routine data of chemical measurements (N, P and COD concentrations) in the influent and effluent. Three independent time series, one for each experimental treatment (wastewater, fertilizers, seawater).	.xls	UAL	Public through deliverables as per the GA
Indoor reactors chemical data ***Tasks 1.1 & 1.2 (WP1)***	*Time series data* Offline routine data of chemical measurements (N, P and COD concentrations) in the influent and effluent. Three independent time series, one for each experimental treatment (wastewater, fertilizers, seawater).	.xls	UAL	Public through deliverables as per the GA
Outdoor reactors sample list for external analyses ***Tasks 1.1 & 1.2 (WP1)***	*Metadata* Samples collection from microalgae-bacteria culture systems for genomics, CDOM/FDOM analyses, chemical fingerprinting, microscopy, and cytometry. Three independent sample lists, one for each experimental treatment (wastewater, fertilizers, seawater).	.xls	UAL	Public through deliverables as per the GA
Indoor reactors sample list for external analyses ***Tasks 1.1 & 1.2 (WP1)***	*Metadata* Samples collection from microalgae-bacteria culture systems for genomics, CDOM/FDOM analyses, chemical fingerprinting, microscopy, and	.xls	UAL	Public through deliverables as per the GA

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

	cytometry. Three independent sample lists, one for each experimental treatment (wastewater, fertilizers, seawater).			
Anaerobic reactors experiments ***Tasks 2.1, 2.2 & 2.3 (WP2)***	*Time series data* Data table of routine measurements from anaerobic reactors during perturbation experiments 1, 2 & 3.	.xls	IMDEA-E	Public through deliverables as per the GA
Anaerobic digestion process inhibition ***Tasks 2.1, 2.2 & 2.3 (WP2)***	*Time series data analyses* Table with parameter values that determine a low, moderate or high anaerobic digestion process inhibition.	.xls .odp	IMDEA-E	Public through deliverables as per the GA
Microalgae population analysis ***Tasks 2.1, 2.2 & 2.3 (WP2)***	*Time series data analyses* Table to compare parameter values that determine best strategies to mitigate/prevent process failure.	.xls .odp	IMDEA-E	Public through deliverables as per the GA
Chemical fingerprint (microalgae photo-bioreactors) ***Task 1.4 (WP1)***	*Time series data* Measurements of Non-Purgeable Organic Carbon and Total Dissolved Nitrogen in PBRs.	.xls	AWI	Public through deliverables as per the GA
Chemical fingerprint (anaerobic reactors) ***Task 2.5 (WP2)***	*Time series data* Measurements of Non-Purgeable Organic Carbon and Total Dissolved Nitrogen in ARs.	.xls	AWI	Public through deliverables as per the GA
Chemical fingerprint (microalgae photo-bioreactors) ***Task 1.4 (WP1)***	*Time series data* High resolution mass spectrometry of the dissolved fractions.	.xls	AWI	Public through deliverables as per the GA
Chemical fingerprint (anaerobic reactors) ***Task 2.5 (WP2)***	*Time series data* High resolution mass spectrometry of the dissolved fractions.	.xls	AWI	Public through deliverables as per the GA
Centralized dataset from microalgae production systems ***Tasks 1.1 & 1.2 (WP1)***	*Metadata* Data table listing metadata including information on the experimental treatments (treatment id), sampling date/time, sample id, creator, type of sample, sample volume collected/filtered (if applicable), the purpose of the sample (type of	.xls	CSIC	Public through deliverables as per the GA

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

	measurements/data produced), name of data analyst.			
Metabarcoding of algal microbiomes ***Task 1.3 (WP1)***	*Time series data* • fastq files: one per sample (files with DNA sequences)[1]. • OTU tables for the 16S and 18S rRNA gene indicating numbers of sequences per OTU in each sample[2].	.fastq .tsv .xls .odp	CSIC	Public through deliverables as per the GA
Metagenomics of algal microbiomes ***Task 1.3 (WP1)***	*Time series data* • fastq files: one per sample (files with DNA sequences)[1]. • gene catalog[2]. • gene abundance tables[2].	.fastq .fasta .tsv .xls .odp	CSIC	Public through deliverables as per the GA
Metatranscriptomics of algal microbiomes ***Task 1.3 (WP1)***	*Time series data* • fastq files: one per sample (files with DNA sequences)[1]. • gene catalog[2]. • gene expression tables[2].	.fastq .fasta .tsv .xls .odp	CSIC	Public through deliverables as per the GA
FDOM/CDOM ***Task 1.4 (WP1)***	*Time series data* Data tables including the matrices generated by the analyses of spectrofluorimetry of the dissolved organic matter in filtrates of algal PBR samples.	.xls	CSIC	Public through deliverables as per the GA
Flow cytometry (virus and bacteria) ***Task 1.1 & 1.2 (WP1)***	*Time series data* Data tables including the abundance of virus and bacteria in algal PBR samples.	.xls	CSIC	Public through deliverables as per the GA
Microscopy (DAPI) ***Task 1.1 & 1.2 (WP1)***	*Time series data* Data tables including the abundance of virus, bacteria and protists in algal PBR samples.	.xls	CSIC	Public through deliverables as per the GA
Data centralization – PBRs metadata ***Task 3.1 (WP3)***	*Centralized data* Data tables of chemical fingerprinting (from AWI), MetaOMICs, and PBRs routine measurements (from UAL and CSIC). The aim is to centralize information for making it organized, rapidly accessible, and	.xls .odp	CSIC	Public through deliverables as per the GA

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

	ready to analyse. This PBR metadata will be the main data source from which to generate the base of knowledge for the development of a bioreactor systems failure prediction technology. These data tables refer to deliverable 3.1 Data integration – PBRs metadata and gather information generated in WP1.			
Data integration – gene patterns, diversity, and productivity in PBRs ***Task 3.1 (WP3)***	*Time series data* Data tables, diagrams and explanatory notes reporting microbial interactions, food webs / substrate flow, genes, and gene expression patterns, and links between functional diversity, functional redundancy, and bioreactor productivity and stability.	.tsv .xls .odp	CSIC	Public through deliverables as per the GA
DARWIN model simulations ***Task 3.2 (WP3)***	*Time series data* Outputs of the MIT-DARWIN ecosystem model simulations for variables such as biomass, productivity, biodiversity, stability.	.xls .m .odp	CSIC	Public through deliverables as per the GA
Centralized dataset (anaerobic digestion) ***Tasks 2.1, 2.2 & 2.3 (WP2)***	*Metadata* Data table listing metadata including information on the experimental treatments (treatment id), sampling date/time, sample id, creator, type of sample, sample volume collected/filtered (if applicable), the purpose of the sample (type of measurements/data produced), name of data analyst.	.xls	IMDEA / NMBU	Public through deliverables as per the GA
Metagenomics of communities in ADs ***Task 2.4 (WP2)***	*Time series data* • fastq files: one per sample (files with DNA sequences)[1]. • gene catalog[2]. • gene abundance tables[2].	.fastq .fasta .tsv .xls .odp	NMBU	Public through deliverables as per the GA
Metaproteomics of communities in ADs ***Task 2.4 (WP2)***	*Time series data -* • Raw mass spectrometry files, one per sample • Protein catalog • Protein abundance tables	.mzxml .tsv .xls	NMBU	Public through deliverables as per the GA

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

Data centralization – ARs metadata ***Task 4.1 (WP3)***	*Centralized data* Data tables of chemical fingerprinting (from AWI), MetaOMICs, and ARs routine measurements (from IMDEA and NMBU). The aim is to centralize information for making it organized, rapidly accessible, and ready to analyse. This AR metadata will be the main data source from which to generate the base of knowledge for the development of a systems failure prediction technology. These data tables refer to deliverable 4.1 Data integration – ARs metadata and gather information generated in WP2.	.xls .odp	IMDEA / NMBU	Public through deliverables as per the GA
Data integration – gene patterns, diversity, and productivity in Ars ***Task 4.1 (WP4)***	*Time series data analyses* Data tables, diagrams and explanatory notes reporting microbial interactions, food webs / substrate flow, genes, and gene expression patterns, and links between functional diversity, functional redundancy, and bioreactor productivity and stability.	.tsv .xls .odp	IMDEA / NMBU	Public through deliverables as per the GA
ADM1 simulation model ***Task 4.2 (WP4)***	*Time series data* Outputs of the ADM1 simulations for AD inhibition scenarios.		NMBU	Public through deliverables as per the GA
Life Cycle Inventory of the PRODIGIO biogas production chain and the early-warning detection systems ***(WP5)***	Mass and material flows of the different elements building the life cycle of the production of biogas from microalgae implemented in the PRODIGIO project. The mass and inventory flows of the different early-warning systems will also be listed. This inventory is specific to the systems present in the PRODIGIO project.	.xls	ARMINES	Public through deliverables as per the GA
Life Cycle Impacts of the PRODIGIO biogas production chain	Results of the LCA of the different biogas production pathways specific to the PRODIGIO project.	.xls	ARMINES	Public through deliverables as per the GA

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

(WP5)				
Parameterized model of the biogas production pathways ***(WP5)***	Parameterized life cycle assessment model describing, in a general way, biogas production pathways from microalgae.	.jpyn	ARMINES	Public through deliverables as per the GA
List of the potential positive and negative social impacts of the PRODIGIO biogas production chain ***(WP5)***	List describing the expected positive and negative social impacts of the PRODIGIO biogas production chain.	.xls	ARMINES	Public through deliverables as per the GA
Life cycle costs of the PRODIGIO biogas production chain ***(WP5)***	List of all costs associated with the different elements present in the life cycle of the PRODIGIO biogas production chain.	.xls	ARMINES	Public through deliverables as per the GA
Method for the life cycle sustainability assessment of biogas production pathways ***(WP5)***	General methodology to compute an overall life cycle sustainability assessment of biogas production pathways.	.doc	ARMINES	Public through deliverables as per the GA
Empirical dynamic modelling ***Task 3.3 (WP3)***	*Time series data analyses* Time series of chemical and biological data will be assembled and analysed using the Empirical Dynamic Modelling (EDM) framework. EDM will calculate interaction matrix estimates (Jacobians) from time series data of algal microbiomes, their topologies and estimates of the strength of interactions. EDM outputs will be provided as data tables along with explanatory reports.	.xls .odp	NTU	Public through deliverables as per the GA
Empirical dynamic modelling ***Task 3.3 (WP3)***	*Time series data analyses* List of Early Warning Signals and estimates of warning thresholds and warning times for algal photobioreactor, and their computation.	.odp	NTU	Public through deliverables as per the GA
Empirical dynamic modelling	*Time series data analyses* Time series of chemical and	.xls .odp	NTU	Public through deliverables as

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

Task 4.3 (WP4)	biological data will be assembled and analysed using the Empirical Dynamic Modelling (EDM) framework. EDM will calculate interaction matrix estimates (Jacobians) from time series data of anaerobic microbiomes, their topologies and estimates of the strength of interactions. EDM outputs will be provided as data tables along with explanatory reports.			per the GA
Empirical dynamic modelling ***Task 4.3 (WP4)***	*Time series data analyses* List of Early Warning Signals and estimates of warning thresholds and warning times for anaerobic reactors and their computation.	.odp	NTU	Public through deliverables as per the GA

[1]fastq files will be deposited at the European Nucleotide Archive.
[2]Tables and other processed data will be deposited at figshare.

3. FAIR DATA

Projects funded by the European Commission should develop a Data Management Plan and deposit data in open access following the FAIR principles, and should therefore be Findables, Accessibles, Interoperables and Reusables, in other words, they can be found, they can be accessed, interoperables and they can be reused to increase the efficiency and transparency of research.

3.1 Making data findable, including provisions for metadata

The provision of adequate metadata (a description of the key attributes and properties of each dataset) is fundamental to enable the finding, understanding and reusability of the data, as well as the validation of research results. Descriptive metadata in particular, aims to provide searchable information that makes data discovery and identification possible. PRODIGIO will adopt the DataCite Metadata Schema, one of the broadest cross-domain standards available, as the basis for dataset description. The minimum set of descriptors established for a PRODIGIO dataset include:

Type: a description of the resource. Recommended best practice: use of a controlled vocabulary such as the DCMI Type.

Identifier: a unique string that identifies a resource. Provided by repository where the dataset is stored. Preferred option: digital object identifier (DOI); also accepted URL, URN, Handle, PURL, ARK.

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

Publication date: date when the data was or will be made publicly available.

Publication Date Format: YYYY-MM-DD.

Title: a name by which a resource is known (free text).

Authors: the main researcher(s) involved in producing the data, or the authors of the publication, in priority order and affiliation. Recommended inclusion of a name identifier. Personal name format: family, given. Affiliation format: free text.

Description: additional information that does not fit in any of the other categories. Example: publication abstract. Format: open.

Version: the version number of the resource. Format: track major_version. minor_version. Examples: 1.0, 2.1

Language: primary language of the resource.

Rights: information about rights held in and over the resource.

Values: openAccess, embargoed Access, restricted Access, closed Access.

Licence: information about the type of licence applying to the dataset

Contributors: institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource. This property must also be used to allow unique and persistent identification of the funder.

Values: European Commission (EU), H2020, Research and Innovation action, PRODIGIO, Grant Agreement Number 101007006.

Subject: subject, keywords, classification code, or key phrase describing the resource (free text).

Additionally, metadata elements and documentation providing specific information about the data collection processes, methodology, data analysis procedures, variable definitions, or relationships between the different files of a dataset will be compiled to ensure data interpretability and reusability. The relevant metadata categories mentioned above will also be applied for data related to stakeholder interactions.

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

3.2 Making data openly accessible

The PRODIGIO project results will be made openly accessible provided that open publication does not interfere with the obligation to protect and exploit the results or the protection of personal data.

Regarding protection of results, to ensure that dissemination of the PRODIGIO research outputs does not jeopardize their exploitation potential, project results will be subject to evaluation prior to any dissemination activity. Results approved for dissemination will be made accessible through a variety of channels including project webpage (https://prodigio-project.eu/) social media, scientific conferences, scientific publications in peer-reviewed journals, and data repositories, among others.

Regarding the protection of personal data, stakeholder views will be either audio recorded or documented in writing as interview or workshop notes. A restricted access policy will be implemented for stakeholder consultation data in order to ensure confidentiality of personal data. These raw data will be handled and analysed by the teams conducting the respective research tasks. Summaries of stakeholder views will be presented in project reports which will be made publicly available on the project website and in open-access repositories. In these reports, stakeholder views will be presented in a pseudonymised way. No reference will be made to individual stakeholder representatives or individual stakeholder organisations.

Being part of the Open Research Data Pilot (ORDP), the PRODIGIO consortium is committed to provide Open Access (free-of-charge access) to all scientific publications and associated research data.

Open Access (OA) to PRODIGIO peer reviewed scientific publications will be mostly granted through "Gold" OA, although "Green" OA will be also be considered if "Gold" OA is not provided by the selected journal. Final versions of articles accepted for publication and their associated metadata (see section 2.1 and below) will be deposited in Zenodo, an interdisciplinary open data repository service created through the European Commission's OpenAIRE project and hosted at CERN, and will be made openly accessible at the time of publication ("Gold" OA) or with a maximum of 6 months embargo (for "Green" OA). Zenodo is compliant with the FAIR principles: it assigns a DOI to each deposited object, supports DOI versioning, is compliant with the DataCite Metadata Schema, is searchable, provides clear and flexible licensing, and provides secure backup.

In addition to the scientific publication, OA will also be provided to the research data required to validate the published results. Although Zenodo allows the deposit of data as well as publications, the use of discipline-specific repositories is often a more convenient option since (i) they have been developed to cover the subject specific needs and (ii) being widely used by the community, facilitate integration with other datasets

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020
RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

All data deposited in the repository will also have a record in Zenodo for the associated publication with a link to the externally deposited data files. Additionally, Zenodo will be the repository of choice for those data types for which a disciplinary repository is not available. The deposited dataset will include all the information needed to interpret and re-use the data following reporting standards when available (see section 3.6). These will include publication file, raw and processed data files (in open or widely used formats), detailed protocols with information on instruments and settings used, a codebook for the variables used, and a readme file describing the files that compose the dataset and the relation between them.

As already mentioned, open or widely used file formats that can be accessed with open software (or software that is in widespread use) will be the preferred option for data collection. When the use of proprietary formats is necessary, the name and version of the software used to generate the file will be indicated in a readme.txt file included in the dataset.

All data deposited in a repository will be made openly accessible under no access restrictions other than the embargo period for "Green" OA publications mentioned above.

3.3 Making data interoperable

Promoting data exchange and integration to its full potential requires the use of standardised data formats, metadata elements, and ontologies that ensure the reusability of the underlying data. As discussed in section 1, open or otherwise widely used file formats will be used to collect and share the data derived from PRODIGIO research activities, thus facilitating data retrieval and analysis by other users.

With regard to metadata, likewise discipline-specific repositories, discipline specific metadata schemes broadly accepted by the scientific community should be the preferred alternative since they have been developed to cover subject specific needs. Accordingly, disciplinary repositories often show compliance with such specific metadata standards in combination with (recommended) controlled vocabularies.

3.4 Increase data re-use

Re-use is one of the pillars of FAIR data. Data re-use increases the impact and visibility of research, maximises transparency and accountability, promotes the improvement and validation of research methods, stimulates innovation through new data uses, and saves resources avoiding unnecessary replications. For data to be re-usable it should be in an open or widely-used file format, well described with rich metadata that meet domain-relevant community standards, and released under a clear data usage licence. The way PRODIGIO will approach the first two points has already been discussed in section 1 (file formats) and sections 2.1 and 2.3 (metadata). Regarding licensing, as a default standard PRODIGIO will share scientific publications and the associated research data under a Creative Commons Attribution Licence CC-BY whenever possible. CC-BY does not impose any restriction on

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020
RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006

access and reuse of the data; it allows users to copy, distribute, transmit, adapt and make commercial use of the data with the sole condition that the creator is appropriately credited. Most data repositories as well as most open access and hybrid publishers support the use of CC-BY licence.

Data quality check is the responsibility of the partners involved in the generation of the dataset and will be supported by a peer-review process at publication. Should errors be detected in already published data, these will be corrected and adequately documented in a new version of the dataset.

3.5 Allocation of resources

Adequate data management is an integral part of good research practice and as such it concerns every person involved in the research process. All PRODIGIO partners have agreed to the general guidelines set up in this DMP and it is the responsibility of the group leaders to ensure that they are known and implemented by all members of their research group. For each dataset, the partner that generates the data is accountable for registering and storing all data and metadata according to the guidelines of this DMP, applying adequate back up policies, and sharing all public data through the selected open access repository. The project coordinator is in addition responsible for the maintenance of the project share-point hosting service for the sharing and storing of PRODIGIO main documents during the active phase of the project.

Resources for data storage and back up during the active phase of the project will be provided by the respective partner's institutions (costs included in standard indirect costs). No direct costs for data sharing and long term preservation are anticipated given that all the considered data repositories are free of charge.

3.6 Data security

All PRODIGIO partners have adequate storage capabilities and back up policies at their respective institutions that guarantee the safe storage of the generated research data during the active phase of the project. All project documents (grant and consortium agreements, deliverables, meeting minutes, project reports and presentations, scientific manuscripts) are stored on a dedicated SharePoint platform hosted and administered by the project partner UAL.

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020
RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT #101007006