D6.2 Data Management Plan

Project Information

Project name:	**A Multiplexed plasmo-photonic Biosensing platfoRm fOr rapid and intelligent Sepsis dIAgnosis at the point of care**
Project acronym:	AMBROSIA
Project start date:	1/1/2023
Project duration:	48 months
Contract number:	101093166
Project coordinator:	Nikolaos Pleros / AUTH
Topic:	HORIZON-CL4-2022-DIGITAL-EMERGING-01-03: Advanced multi-sensing systems

Document Information

Deliverable number and title:	D6.2 Data Management Plan
Document type:	R-Document, report
Contractual date of delivery:	30/08/2023
Calendar date of delivery:	04/10/2023
Work package number:	WP6 Dissemination and Exploitation
Lead partner:	AUTH
Dissemination level:	PU
Document status:	Version 1

Author List:

Organization	Author
AUTH	Markou Stella, Nikos Pleros, Evangelia Chatzianagnostou, Dimosthenis Spasopoulos
ICN2	M.-Carmen Estevez, Jessica Llop
UB	Jean-Claude Weeber, Aurore Andrieux, Juan-Miguel Arocas Garcia, Kirill Vasilev
XCEL	Alex Farell, Alin Fecioru
VHIR	Juan José González López, Juan Carlos Ruiz
ARGO	Michal Szaj
BIOPIX-T	Alexandros Pantazis
MLIQ	Santiago Cocaña, David Olea, Ikerne Etxebarria
UOI	Dimitris Bellas, Eleftheria Lampadariou, Lefteris Lidorikis
SMART	Guilherme Sombrio
SOTON	Frederic Gardes, Xingshi Yu, Ioannis Zeimpekis, Thalia Dominguez Bucio, Ipsita Chakraborty, Isaac Johnson
LGT	José Carreira

TABLE OF CONTENTS

1 Executive Summary

This document provides the initial Data Management Plan (DMP) of the AMBROSIA project, as defined during the first eight months of the project. In this initial phase AMBROSIA has identified a number of data sets which will be generated within the project and will be shared with the research community via knowledge repositories available in the consortium. With this activity, AMBROSIA aims to improve the exchange and dissemination of research results and, possibly, to enable and promote a wider validation of the project results and to encourage a fair comparison and evaluation of different solutions in the AMBROSIA technical areas.

The AMBROSIA project abides to the European Commission's (EC) vision that information already paid for through EC funding should not be paid for again each time it is accessed or used by the research community and that it should benefit European companies and citizens to the full. This means making publicly-funded scientific information available online, at no extra cost, to European researchers, innovative industries and citizens, while ensuring long-term preservation. This will have to be done, without, of course, compromising any Intellectual Property Rights (IPR) and commercial plans of the participants, and ensuring patient's data protection and confidentiality.

2 Introduction

2.1 Purpose of this document

D6.2 deals with the research data produced, collected and preserved during the project, as well as with the way that those data will be handled after the end of the project. It should be noted that the DMP reported in this document is just an initial version, which provides an overview of the data sets initially identified by AMBROSIA partners.

2.2 Document structure

The present deliverable is split into the following major parts:

- Data Management Plan
- Types of data generated within the project
- Data to be shared and not
- Management plan for the different categories of data

2.3 Audience

The content of the document is public.

3 Data management plan

This section describes the consortium plans for the management of different sets of research data which will be produced, collected, or just used for internal processing and validation in the AMBROSIA project. Internal data refers to data produced within the project including simulation or experimental test-bed measurements as well as outcomes of AMBROSIA evaluation and demonstration. For those datasets, the project will define a suitable data management strategy, identifying proper procedures for their documentation, sharing and maintenance.

Section 4 provides the list of relevant data sets which will be considered in the AMBROSIA project, while the following sections provide the details for each data set. It should be noted that this report is released at the initial stage of the project (M08), when prototypes and testbeds, as well as a specific evaluation plan, are still under definition. For this reason, herein just an initial description of the expected data sets and their management plan is provided, while more details and potentially further data sets will be dealt on a later stage of the project and will be reported in the upcoming project periodic reports.

The datasets that will be published within AMBROSIA's DMP will belong solely to the partners that generated the associated Intellectual Property Rights (IPR), according to the provisions of the Consortium Agreement (CA). The respective partners are responsible for postponing or restricting data sharing to allow sufficient time for publishing the results in peer-reviewed journals or for seeking patents. To facilitate handling of datasets especially where multiple partners are involved and to expedite data dissemination, the process is overseen by the IPR board.

Before a dataset is uploaded to the designated database, the partner responsible for uploading and preserving the data will seek consent for data preservation and sharing from all partners involved. The Data Management Plan will be consistent with the IPR policy (i.e. IP, confidentiality, and publication provisions) defined in the CA at the beginning of the project and updated as any changes occur in the consortium policies. The CA clarifies how issues related to data protection are handled to help partners maximize the return on the investment. Rules for the (i) ownership of knowledge, (ii) the protection of knowledge and (iii) the access rights are clearly set in the agreement.

For the data that can be made publicly available, partners will use an open access repository (e.g. *Zenodo*: https://zenodo.org/, the *Institutional Repository of Scientific Publications of the Aristotle University of Thessaloniki*: https://ikee.lib.auth.gr/?ln=en and the *University of Southampton Institutional Research Repository*: https://eprints.soton.ac.uk/), connected to the tools proposed by the EC (e.g., open AIRE) to grant access to the publications and to bibliographic metadata in a standard format including information requested by the EC. The Zenodo community is already acting as a one-stop-shop for data generated by EU projects and, thus, makes it easier for the data to be discovered by interested parties. Links to some of the data, especially in terms of publications, posters, videos etc. will also be directly available on the project website: http://ambrosia-h2022.eu/. A private section within the project website will be used by the partners for data update and exchange, and to build a body of data to be used in the post-project exploitation phase. The data on the project website will be available for a period of 5 years after the project's end.

The AMBROSIA project also involves personal and clinical data from humans (patients and healthy donors) something that arise ethical issues. For this reason and in compliance with current legislation[1],

[1] Recommendation CM/Rec(2016)6 of the Committee of Ministers to member States on research on biological materials of human origin. Biomedical Research (Law 14/2007, of July 3rd, Biomedical Research) and the Royal Decree 1716/2011 (of November 18th).

the AMBROSIA project has been approved by the Ethics Committee of Clinical Research of Hospital Universitari Vall d'Hebron (reference number PR(AG)396-2023). The collected data will be stored at ICS-Vall d'Hebron Hospital server for at least 15 years. Approval for the Sepsis Bank of HUVH Biobank was obtained in 2016 with the Ethics Committee of Clinical Research reference number PR(AG)11/2016.

4 Types of data generated within the project

Type of data generated and collected by the project regard material property characterization, experimental procedures, simulation results, fabrication process flows, prototype specifications, tests and measurement results and donors' personal and clinical data.

More specifically, the project is expected to generate the following types of data:

(a) Data related to the design, development and characterization of the AMBROSIA integrated plasmo-photonic sensor (WP3 - T3.1, T3.4)
(b) Data related to the design, development and characterization of the neuromorphic photonic (NP) circuit (WP2 – T2.3, WP3 - T3.2, T3.4)
(c) Data related to the design and development of the InP actives and their integration on the SiN platform through micro-transfer printing (μTP) (WP3 - T3.3, T3.4)
(d) Data related to the sensor's surface functionalization/characterization (WP4 – T4.1)
(e) Data related to the design and development of the microfluidic system (WP4 – T4.2)
(f) Data related to the design and development of the electro-optical read-out and diagnostic system (WP4 – T4.3, WP2 – T2.3)
(g) Data related to the design and development of the portable POC unit (WP5 -T5.1)
(h) Data related to the experimental validation of the AMBROSIA technology in sepsis diagnosis in laboratory environment with protein and bacteria dilutions in standard buffer. (WP5 -T5.2)
(i) Data related to the experimental validation of the AMBROSIA technology in sepsis diagnosis with real clinical samples. (WP5 – T5.3)
(j) Patient's and healthy donor's clinical data. (WP5 – T5.3)
(k) Business plan (WP8 - T8.2)

For each of the data sets that are foreseen to be produced internally in the project, the responsible partner has defined a suitable data management strategy which is reported in the following sections. Results and technical data will be shared given that they will not impose implications in the project dissemination and IP exploitation activities. Data from human clinical samples will be obtained with informed consent through the established protocols previously approved by the Clinical Research and Ethics Committee of the Hospital. All data will be protected in accordance with the EU Data protection Directive 95/46/EC "on the protection of individuals with regard to the processing of personal data and on the free movement of such data". Specifically, all samples will be anonymized and no personally identifiable data will be included in any dataset.

5 Data to be shared and not

Some of the above data will be shared with the scientific community, following the principles of open data sharing. In particular:

- Data from category **(a)** might be shared with the scientific community, following research publications and presentations in conferences by AUTH, UOI, UB, LGT and the involved partners. Since this involves

core activities of the project with significant exploitation potential, particular attention will be paid to IP protection issues, prior any publication. Data may also include know-how of the consortium which also cannot be openly shared due to IP issues.

- Data from category **(b)** might be shared with the scientific community, following research publications and presentations in conferences by SOTON, AUTH and the involved partners. Since this involves core activities of the project with significant exploitation potential, particular attention will be paid to IP protection issues, prior any publication. Data may also include know-how of the consortium which also cannot be openly shared due to IP issues.
- Data from category **(c)** might be shared with the scientific community, following research publications and presentations in conferences by UOI, SMART, XCEL, LGT and the involved partners. Since this involves core activities of the project with significant exploitation potential, particular attention will be paid to IP protection issues, prior any publication. Data may also include know-how of the consortium which also cannot be openly shared due to IP issues.
- Data from category **(d)** might be shared with the scientific community, following research publications and presentations in conferences by ICN2 and the involved partners.
- Data from category **(e)** might be shared with the scientific community, following research publications and presentations in conferences by MLIQ and the involved partners.
- Data from category **(f)** are not expected to be openly sharable. They refer to the software development for the control and communication of all subsystems of the final PoC unit. However, the final application for the user interface will be available to the public at least for some basic functions.
- Data from category **(g)** are not expected to be openly sharable. They refer to the development of the final PoC unit, which includes intellectual property and potential commercial products of the partners.
- Data from category **(h)** might be shared with the scientific community, following research publications and presentations in conferences by ICN2, AUTH, VHIR and the involved partners.
- Data from category **(i)** might be shared with the scientific community, following research publications by ICN2, VHIR, AUTH, and the involved partners.
- Data from category **(j)** will not be shared with the community since refer human's personal and clinical data.
- Data from category **(k)** will not be shared since it is confidential information.

6 Management plan for the different categories of data

In the sections that follow, details on the data to be shared with the community as well as patient's and healthy donor's clinical data are presented. Focus is put on:

a) Dataset description
b) Standards and metadata, where applicable
c) Information about tools and instruments
d) Accessibility

6.1 Data related to the design, development and characterization of the AMBROSIA integrated plasmo-photonic sensor

The development of the AMBROSIA integrated plasmo-photonic sensor generates data as an output of Tasks 3.1 and 3.4, mainly during: (i) simulation, (ii) mask design, (iii) characterization during fabrication and (iv) characterization during sensor measurements. The simulations are based on internally developed

MATLAB codes or Lumerical projects and generate datasets exported as numbers into any post processing software (Origin, MATLAB etc.). The mask design is based on Synopsis OptoDesigner software and generates mask files of the industry standard type GDSII. The characterization during fabrication may involve optical microscopy, SEM, AFM, ellipsometry, reflectometry, profilometry, electrical measurements and optical measurements. The characterization during sensor measurements will mainly be performed by optical or opto-electrical measurements.

The following types of data will be produced:

1. Confidential data for internal use only.
2. Possible data used in scientific and technical publications.

Information about tools and instruments:

Documentation will be available either in the Microsoft Word format, Microsoft PowerPoint format or as PDF file. Measured data will be available as Excel tables, Origin files, text files with headers or files originating from the instrumentation mentioned above.

Standards and metadata:

To guarantee a high academic standard, results will be documented considering the typical nomenclature of the particular field of research. Measured data are either self-explanatory or will be prepared or described by an additional document. Metadata will be made available for each shared document or dataset.

Accessibility:

Once a dataset is made available on the project repository, it will freely be available to the community, in the way described in Sections 3 and 5.

6.2 Data related to the design, development and characterization of the neuromorphic photonic (NP) circuit

The development of the NP circuit generates data as an output of Tasks 3.2 and 3.4, mainly during (i) circuit designs, (ii) mask design, (iii) characterization during fabrication and (iv) characterization during optical testing. Circuit designs are based on internally developed MATLAB/Python codes and generate datasets exported as numbers into the same or in any other data processing software such as Origin, MATLAB and Python. The design is drawn on Synopsis OptoDesigner or Nazca software and generates GDSII mask files. The characterization during fabrication may involve optical microscopy, SEM, ellipsometry, profilometry, four-point and Van der Pauw electrical characterization, and optical measurements. The characterization during optical testing will be performed by optical measurements.

The following types of data will be produced:

1. Confidential data for internal use only.
2. Possible data used in scientific and technical publications.

Information about tools and instruments:

Documentation will be available either in the Microsoft Word or PowerPoint format or as PDF file. Measured data will be available as Excel tables, Origin files, text files with headers or files originating from the instrumentation mentioned above.

Standards and metadata:

To guarantee a high academic standard, results will be documented considering the typical nomenclature of the particular field of research. Measured data are either self-explanatory or will be

prepared or described by an additional document. Metadata will be made available for each shared document or dataset.

Accessibility:

Once it has been decided that a dataset can be made available on the project repository, it will freely be available to the community, in the way described in Sections 3, 4 and 5.

6.3 Data related to the design and development of the InP actives and their integration on the SiN platform through micro-transfer printing (μTP)

The development of the InP actives and their integration on the SiN platform generates data as an output of Tasks 3.3 and 3.4, mainly during (i) simulation and design of the laser and photodiode, (ii) simulation of the InP-to-SiN interface (iii) mask design, (iii) characterization during fabrication and integration and (iv) characterization during optical testing. Simulations are based on either Lumerical's commercial software utilizing the 3D-FDTD, FDE and EME solvers or on internally developed codes in Python, Fortran and C++. The datasets generated from the simulations are exported as numbers and can be analyzed and/or post-processed into the same or any other processing software i.e. Microsoft Excel, Origin, Matlab and Python. The design of the lithographic mask layout is performed on Synopsys OptoDesigner software, generating the GDSII mask files which are used for the fabrication of the devices. The characterization during fabrication may involve optical microscopy, SEM, TEM AFM, ellipsometry, reflectance spectrometry, fluorescence spectroscopy, profilometry and optical measurements. Characterization during integration may involve yield and placement accuracy assessment using the built-in machine vision-enabled optical pattern recognition capability of the MTP tool. The characterization during sensor measurements will mainly be performed by optical and opto-electrical measurements.

The following types of data will be produced:

1. Confidential data for internal use only.
2. Possible data used in scientific and technical publications.

Information about tools and instruments:

Documentation will be available either in the Microsoft Word or PowerPoint format or as PDF file. Measured data will be available as Excel tables, Origin files, text files with headers or files originating from the instrumentation mentioned above.

Standards and metadata:

To guarantee a high academic standard, results will be documented considering the typical nomenclature of the particular field of research. Measured data are either self-explanatory or will be prepared or described by an additional document. Metadata will be made available for each shared document or dataset.

Accessibility:

Once it has been decided that a dataset can be made available on the project repository, it will freely be available to the community, in the way described in Sections 3 and 5.

6.4 Data related to the sensor's surface functionalization/characterization

Data from the characterization of the functionalized sensor surface will be generated as an output of Task 4.1 and will include the description of the experimental procedure, as well as measurements needed for the characterization of the functionalized surface, such as spectroscopic (XPS and IR) and optical analysis.

This includes the characterization of biofunctionalization processes and label-free detection, to be performed directly on chip by optical or opto-electrical measurements.

The following types of data will be produced:

1. Confidential data for internal use only.
2. Possible data used in scientific and technical publications.

Information about tools and instruments:
Documentation will be available either in the Microsoft Word format or as PDF file. Measured data will be available as Excel tables, Origin files or text files originating from the instrumentation used.

Standards and metadata:
To guarantee a high academic standard, results will be documented considering the typical nomenclature of the particular field of research. Measured data are either self-explanatory or will be prepared or described by an additional document. Metadata will be made available for each document or dataset.

Accessibility:
Once a dataset is made available on the project repository, it will freely be available to the community, in the way described in Section 3 and 5.

6.5 Data related to the design and development of the microfluidic system

Data from the design and development of the microfluidic system will be generated as an output of Task 4.2.

The following types of data will be produced:

3. Confidential data for internal use only.
4. Possible data used in scientific and technical publications.

Information about tools and instruments:
Documentation will be available either in the Microsoft Word format or as PDF file.

Standards and metadata:
To guarantee a high academic standard, results will be documented considering the typical nomenclature of the particular field of research. Metadata will be made available for each document or dataset.

Accessibility:
Once a dataset is made available on the project repository, it will freely be available to the community, in the way described in Section 3 and 5.

6.6 Data related to the design and development of the electro-optical read-out and diagnostic system

The data regarding the development of the electro-optical readout and diagnostic system will be generated as an output of T4.3 where the target is to develop the necessary software that will be responsible for a series of functionalities with the ultimate goal to extract and visualize the diagnostic results. Firstly, it will process the electrical data coming from the sensor chip including deployment of noise cancellation algorithms in order to (i) extract and visualize the necessary sensograms and (ii) to drive the laser modulators of the photonic neuromorphic chip. Additionally, based on AI algorithms and the protocol for sepsis severity stratification, it will drive the PCM-based weighting elements of the photonic neuromorphic circuit, and receive and visualize the diagnostic results of the photonic neuromorphic chip.

The following types of data will be produced:

1. Possible data used in scientific and technical publications.
2. Possible data used in patent applications.

We will not share technical details on the implementation of the diagnostic system and the PCB design. Only some basic functions of the actual software with regard to the user interface of the PoC unit will be available to the public and only closed models (3D views, not editable) of PCBs will be used for presentation purposes.

Information about tools and instruments:
Documentation will be available either in the Microsoft Word format or as PDF file. Measured data will be available as csv files or similar file format originating from the instrument.

Standards and metadata:
To guarantee a high academic standard, results will be documented considering the typical nomenclature of the particular field of research. Measured data are either self-explanatory or will be prepared or described by an additional document. Metadata will be made available for each document or dataset.

Accessibility:
Once a dataset is made available on the project repository, it will freely be available to the community, in the way described in Section 3 and 5.

6.7 Data related to the experimental validation of the ABROSIA technology in sepsis diagnosis in laboratory environment with protein and bacteria dilutions in standard buffer.

Data from the experimental validation of the AMBROSIA technology in a laboratory environment will be generated through the measurements carried out on-chip by optical or optoelectrical measurements, involving the optical instrumentation implemented for such purpose at the laboratory level and the AMBROSIA PoC platform. The data will be processed and generate data related to the analytical characterization of detection assays: (i) the establishment of standard calibration curves for the determination of analytical parameters for protein and bacteria solutions in standard buffer, and (2) the evaluation and optimization of the direct analysis in biological fluids.
The following types of data will be produced:

1. Confidential data for internal use only.
2. Possible data used in scientific and technical publications.

Information about tools and instruments:
Documentation will be available either in the Microsoft Word format or as PDF file. Measured data will be available as Excel tables or files originating from the instrumentation used and data processing like Origin files or GraphPad Prism files.

Standards and metadata:
To guarantee a high academic standard, results will be documented considering the typical nomenclature of the particular field of research. Measured data are either self-explanatory or will be prepared or described by an additional document. Metadata will be made available for each document or dataset.

Accessibility:

Once a dataset is made available on the project repository, it will freely be available to the community, in the way described in Section 3 and 5.

6.8 Data related to the experimental validation of the AMBROSIA technology in sepsis diagnosis with real clinical samples

The experimental validation includes the analysis of the collection of real samples generated from the measurements carried out on-chip by optical or optoelectrical measurements, involving the optical instrumentation implemented for such purpose at laboratory level and the AMBROSIA PoC platform. The data will be processed to generate data and clinical parameters related to clinical validation of the developed assays. Data generated by analysis of blood samples by the AMBROSIA technology developed to measure sepsis biomarker will be made available in deliverable 5.3 "Report on the validation of the POC sepsis diagnostic platform" and patient data that permits his/her identification will not be shared.

Information about tools and instruments:
Documentation will be available either in the Microsoft Word format or as PDF file. Measured data will be available as Excel tables or files originating from the instrumentation used and data processing like Origin files or GraphPad Prism files.

Standards and metadata:
To guarantee a high academic standard, results will be documented considering the typical nomenclature of the particular field of research. Measured data are either self-explanatory or will be prepared or described by an additional document. Metadata will be made available for each document or dataset.

Accessibility:

Once it has been decided that a dataset can be made available on the project repository, it will freely be available to the community, in the way described in Section 3 and 5.

6.9 Patient's and healthy donor's clinical data

The experimental validation of the AMBROSIA technology includes the collection of patients' and healthy donors personal and clinical data including sociodemographic characteristics collected from medical records (e.g. age, gender), microbiology results (e.g. pathogen identification), laboratory results including all sepsis biomarkers measurements, quantification of sepsis biomarkers provided by the AMBROSIA instrument, clinical parameters such as source of infection, inflammatory response, main diagnostic, Acute Physiology and Chronic Health Evaluation (APACHE) score II at critical care unit admission, SOFA score, co-morbidities based on Diagnosis-Related Group codings, sepsis-specific medication(s), ICU and hospital length of stay, and vital status during the ICU and hospital stay.

Each patient or donor will sign an informed consent form to allow sample and data collection with research purposes. Data collected from those patients/healthy donors will be collected and managed using REDCap (Research Electronic Data Capture) electronic data capture tools hosted at ICS-Vall d'Hebron University Hospital. REDCap is a secure, web-based application designed to support data capture for research studies. REDCap can be installed in a variety of environments for compliance with such standards as HIPAA, 21 CFR Part 11, FISMA (low, moderate, high), and international standards. REDCap is fully personalized to meet security policies and user needs.

The Vall d'Hebron Institut de Recerca (VHIR), in compliance with the data protection law (Organic Law 15/1999 and Royal Decree 1720/2007), has declared all the files that contain personal data to the "Agencia Española de Protección de Datos" (Spanish Data Protection Agency). In compliance with article 88, Royal Decree 1720/2007, VHIR has a document ("Documento de Seguridad", Security Document) that contains all the technical and organizational measures adopted to fulfil the security standards required by this law.

In the context of the AMBROSIA project, only the personnel of the research group of ICS-Hospital Vall d'Hebron has access to medical records of patients through their access to Hospital database of medical records. The data required for this project will be collected from the Hospital database and registered in a database to which only the Vall d'Hebron Research group have access by using the REDCap software. Each patient will be assigned a code that will be used for the registry of each case in the database. This code will be used to label the blood samples. No personal data that permits identification of the human participants will be collected in the database. The cross-reference between each patient and the code will be annotated in the Hospital database of medical records.

More details are given in D7.1.

Information about tools and instruments:

Data will be collected from the electronic medical records of all patients during their hospital stays at the time of inclusion in the study. As stated previously, the collected data will be sociodemographic characteristics, laboratory results including all sepsis biomarkers measurements, Acute Physiology and Chronic Health Evaluation (APACHE) score II, co-morbidities based on Diagnosis-Related Group codings, sepsis-specific medication(s), ICU and hospital length of stay, and vital status during the ICU and hospital stay. The quantification of the sepsis biomarkers will be obtained at ICS-Vall d'Hebron Microbiology Laboratory using both standard laboratory techniques and an AMBROSIA system, and will be collected and managed by this partner in the data file using REDCap electronic data capture tools hosted at ICS-Vall d'Hebron University Hospital

Standards and metadata:

Not applicable. Data will be only managed at ICS-Vall d'Hebron Hospital.

Accessibility:

Data will be only collected and managed by personnel involved in the AMBROISA project at ICS-Vall d'Hebron Hospital and will be stored at ICS-Vall d'Hebron Hospital server for at least 15 years. No personal data that permits identification of the human participants will be collected in the database. Other participating groups of this project will not have access to data that permits identification of the human participants.

7 Conclusion

This document has provided the first version of the Data Management Plan defined by the AMBROSIA consortium. The DMP has identified a number of data sets which will be most likely generated within the project and, for each of them, has defined the strategy for the documentation, open access sharing and maintenance of the associated data.