Smart Anything Everywhere Initiative

Area 3: Advanced micro-electronics components and Smart System Integration Project: H2020–No 761809

diatomic

Digital Innovation Hubs boosting European Microelectronics Industry

Deliverable 1.3
Data Management Plan (DMP)
(first version)

Author(s):	R. Broechler, B. Ipektsidis, Y. Oikonomidis (INTRASOFT)
Status - Version:	V1.0
Delivery Date (DOW):	28 February 2018
Actual Delivery Date:	27 February 2018
Distribution - Confidentiality:	Public
Code:	DIATOMIC_D1_3 INTRA_FF_20180227

Abstract:

The deliverable at hand describes the data management life cycle for the data to be collected, processed and/or generated by DIATOMIC.

Disclaimer

This document may contain material that is copyright of certain DIATOMIC beneficiaries, and may not be reproduced or copied without permission. All DIATOMIC consortium partners have agreed to the full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

The DIATOMIC Consortium is the following:

Participant number	Participant organisation name	Short name	Country
01	INTRASOFT International S.A.	INTRA	BE
02	F6S NETWORK LIMITED	F6S	UK
03	BioSense	BIOS	SRB
04	Synelixis Solutions	SYN	EL
05	Instituto Pedro Nunes	IPN	PT
06	Fraunhofer IPA	IPA	DE
07	InoSens	INO	SRB
08	Libelium Comunicaciones Distribuidas SL	LIB	ES
09	FastTrack	FASTT	PT

The information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at their sole risk and liability.

Document Revision History

Date	Issue	Author/Editor/Contributor	Summary of main changes
30/01/2018	V0.1	B. Ipektsidis, Y. Oikonomidis	Initial Text
14/02/2018	V0.2	B. Ipektsidis, Y. Oikonomidis	Modified sections, Added section descriptions
21/02/2018	V0.3	R. Hermoso	Added Health Data Set
21/02/2018	V0.4	A. Issa	Added Manufacturing Data Set
22/02/2018	V0.5	Th. Zahariadis	Editing/Reviewing
25/02/2018	V0.6	B. Ipektsidis, Y. Oikonomidis, R. Broechler	Final Editing
27/02/2018	V1.0	B. Ipektsidis, Y. Oikonomidis, R. Broechler	Final version

Table of Contents

List of Abbreviations

API: Application Programming Interface

DIH: Digital Innovation Hub

GDPR: General Data Protection Regulation

Executive Summary

The present deliverable constitutes the Data Management Plan (DMP) of the DIATOMIC project, i.e. a short, general outline of the project policy for data management (i.e. collecting and sharing procedures), including the issues of outlining the types of data already generated and/or foreseen to be generated at this stage of the project; outlining the procedures that will be followed to assess the generated/collected data with respect to their sensitiveness; outlining the data acquisition plan for the duration of the project and finally outlining the measures that are foreseen for the adequate management of the data from the ethical and security points of view.

The described policy herein reflects the current state of consortium agreements regarding data management and is consistent with those referring to exploitation and protection of results and can also be considered as a checklist for the future.

This Deliverable is expected to be updated during the project. Those updates are planned for Months 18 and 36 to reflect the updates in the data management framework of DIATOMIC. The next version of the Data Management Plan will include updated data access, sharing, and licensing details, as well as updated information regarding the datasets described in the present document.

1. Introduction

DIATOMIC's vision is to establish a sustainable ecosystem, which will facilitate digital innovation in the health, agrifood and manufacturing sectors, all of which are under-digitized and of prime importance for society and the economy. The majority of SMEs and midcaps in these sectors are characterised by a low Digital Innovation Index (we name these enterprises "non-tech" hereafter).

At the heart of the DIATOMIC ecosystem, three interconnected sector-specific Digital Innovation Hubs (DIH) pulsate to accelerate digitization in-beat with sector-specific needs, and to enable delivery of AME/SSI based applications to a critical mass of customers. DIHs first assist non-tech companies to find and couple with counterparts from across the EU with complementary technological competencies. Then, acting as a one-stop-shop, they offer the teams both (i) strong technological support to accelerate design, development, prototyping and manufacturing; and (ii) business support to develop solutions with a robust product/market fit, and chart the path to market growth and investment.

The present deliverable constitutes the Data Management Plan (DMP) of the DIATOMIC project, i.e. a short, general outline of the project policy for data management (i.e. collecting and sharing procedures), including the following issues:

1) Outline the types of data already generated and/or foreseen to be generated at this stage of the project, including the context and procedures of this generation, as well as the degree of privacy and confidentiality of the data.

2) Outline the procedures that will be followed to assess the generated/collected data with respect to their sensitiveness.

3) Outline the data acquisition plan for the duration of the project.

4) Outline the measures that are foreseen for the adequate management of the data from the ethical and security points of view. In other words the DMP is a document outlining how research data will be handled during the DIATOMIC project, and after it is completed.

The overall purpose of the DMP is to support the data management life cycle for all data that will be collected, processed or generated by the project.

The described policy herein reflects the current state of consortium agreements regarding data management and is consistent with those referring to exploitation and protection of results.

This deliverable can also be considered as a checklist for the future. *It is a living document that is expected to mature during the project lifetime and will be updated accordingly.*

Four different datasets have been identified, at this early stage of the project. Table 1 gives an overview of all the data that are expected to be collected or generated by the DIATOMIC project.

Table 1: Overview of the research data of the DIATOMIC project

Dataset ID	Dataset Name
DIATOMIC_Experiments	DIATOMIC data related to the PUSH and PULL experiments Including DIATOMIC_Agrifood, DIATOMIC_Health and DIATOMIC_Manufacturing
DIATOMIC_DIH	DIATOMIC data related with the Digital Innovation Hubs information
DIATOMIC_Open Calls	DIATOMIC data related to the Open Calls
DIATOMIC_Publications	DIATOMIC Public Deliverables & Scientific publications

The four datasets listed in Table 1 are described in the next sections. In more details, in section 2 we report about the datasets of the three Push-experiments that will be conducted in the context of DIATOMIC project and their characteristics. Those characteristics include naming references, descriptions, metadata definitions, key attributes, access, sharing, and licensing policies, and last but not least the archiving and preservation guidelines that will be followed throughout DIATOMIC. These data will be further extended during the Pull-experiments, where data from the projects accepted from the open call will be also published.

Section 3 presents information relate to the Digital Innovation Hubs. It will provide information on various DIHs, the areas that they target, their excellence and the services that may offer to 3rd parties.

Section 4 provides information on the open calls. It will include anonymized and aggregated information on the proposals that we have received, their country/region of origin, their distribution in the DIATOMIC selected application areas, the evaluation process details (e.g. time per proposal evaluation, time per consensus meeting), and the percentage of the selected projects per category DIATOMIC selected application area.

Section 5 presents DIATOMIC's policy with regard to datasets in future publications and reports whereas the remaining sections provide insights about other identified datasets, archiving and preservation policies, and ethics related to the data.

This Deliverable is expected to be updated during the project. Those updates are planned for Months 18 and 36 to reflect the updates in the data management framework of DIATOMIC. The next version of the Data Management Plan will include updated data access, sharing, and licensing details, as well as updated information regarding the datasets described in the present document.

2. DIATOMIC PUSH/PULL Experiments Data

This section reports on the organization of the datasets of the three PUSH-experiments that will be conducted in the context of DIATOMIC project and their characteristics, along with the details on the date to be generated, collected and shared during the PULL-experiments.

In more details the DIATOMIC Experiments data include the following datasets:

2.1. Application Experiment: Agrifood

The dataset includes all data related to the Agrifood Application Experiment of DIATOMIC.

2.1.1. Dataset nomenclature

The dataset Reference is "DIATOMIC_Agrifood" and the dataset names are:

- DIATOMIC Agrifood Micro-climate Data
- DIATOMIC Agrifood Soil and Crop Data
- DIATOMIC Agrifood Irrigation Data
- DIATOMIC Agrifood Disease and crop health

2.1.2. Dataset description and key attributes

The DIATOMIC Agrifood Experiment dataset will consist of data related to:

- Micro-climate& Soil/Crop Data
- Irrigation conditions
- Disease and crop health conditions

The dataset will be stored and processed by Synelixis SynField cloud platform so that they can be calibrated and additional information will be generated, such as crops growing degree days and evapotranspiration.

The dataset will also include crop images and videos captured from flying drops, in case we get relevant permissions for flying drone. In the later case, whatever sensitive information (e.g. person faces in the images/video) will be blurred in order to keep anonymity and privacy according to the DIATOMIC Ethical guidelines.

The purpose of the collected datasets is to support end users, farmers, agriculture scientists, agriculture related applications developers and developers that need micro-climate, soil, irrigation and diseases data.

Key attributes, characteristics, and other information related to DIATOMIC Agrifood Experiment dataset:

Data set #2.1: DIATOMIC Agrifood Micro-climate Data

Description of the data that will be generated or collected	Micro-climate data (air temperature, air humidity, wind direction, wind speed, rain volume, rain intension). The data will be associated with time information and geospatial/location information provided by GPS.
Data origin (in case it is collected)	The micro-climate data will be collected from the SynField nodes that will be installed within the DIATOMIC PUSH experiments. Additional data (weather prediction) from 3rd party sources will be also collected.
Data nature	The data will be mainly in a numerical format
Data scale	The data are considered to be in the range of 300K Bytes per day per node.
To whom could the data be useful	Target audience include people active in the areas of weather applications/weather prediction, precision agriculture, agriculture services.
Information on the existence (or not) of similar data	An initial set of data (some thousands of records) are already available from the FIWARE A16 projects Fractals and FI-ADOPT, along with the KATANA project
Information on the possibilities for integration and reuse of the data	The data could be reused for comparison purposes or being subject to different analyses in projects concerning the potential use of agriculture and food technologies

Data set #2.2: DIATOMIC Agrifood Soil and Crop related Data

Description of the data that will be generated or collected	Soil and Crop related data (leaf wetness, soil type, soil temperature, soil humidity, soil conductivity). Moreover, this data set will calculate the field evapotranspiration. The data will be associated with time information and geospatial/location information provided by GPS.
Data origin (in case it is collected)	The Soil and Crop related data will be collected from the SynField nodes that will be installed within the DIATOMIC PUSH experiments. Moreover, the evapotranspiration will be generated by the SynField cloud server.
Data nature	The data will be mainly in a numerical format
Data scale	The data are considered to be in the range of 180K Bytes per day per node.
To whom could the data be useful	Target audience include people active in the areas of precision agriculture, agriculture services
Information on the existence (or not) of similar data	An initial set of data (some thousands of records) are already available from the KATANA project
Information on the possibilities for integration and reuse of the data	The data could be reused for comparison purposes or being subject to different analyses in projects concerning the potential use of agriculture and food technologies

Data set #2.3: DIATOMIC Agrifood Irrigation Data

Description of the data that will be generated or collected	Irrigation data (crop, irrigation frequency, irrigation time, irrigation water pipes pressure, volume of irrigation water consumed). Moreover, this data set will calculate the crop growing degree days. The data will be associated with time information and geospatial/location information provided by GPS.
Data origin (in case it is collected)	The irrigation data will be collected from the SynField nodes that will be installed within the DIATOMIC PUSH experiments. Additional data such as the crop growing degree days will be generated by the SynField cloud server.
Data nature	The data will be mainly in a numerical format
Data scale	The data are considered to be in the range of 250K-350K Bytes per day per node.
To whom could the data be useful	Target audience include people active in the areas of precision agriculture, agriculture scientists and water management services.
Information on the existence (or not) of similar data	According to our knowledge there are no similar data provided as open data.
Information on the possibilities for integration and reuse of the data	The data could be reused for comparison purposes or being subject to different analyses in projects concerning the potential use of agriculture and food technologies

Data set #2.4: DIATOMIC Agrifood Disease and crop health

Description of the data that will be generated or collected	Disease and crop health will include the crop response in multispectral cameras (i.e. the colour), along with images and video from flying drones over the crop. Moreover, additional information such as the presence of a specific disease will be included manually. The data will be associated with time information and geospatial/location information provided by GPS.
Data origin (in case it is collected)	The Disease and plantation health data will be collected from the Plant-o-meter nodes and potentially drones that will be used within the DIATOMIC PUSH experiments. Additional data such as the crop growing degree days will be generated by the SynField cloud server. Additional processing for protecting privacy will be offered.
Data nature	The data will be mainly in a numerical format for the data collected from the plant-o-meter and video format in case of drones.
Data scale	The data are considered to be in the range of 100K-150K Bytes per plant-o-meter measuring and 150-200M MB per video.
To whom could the data be useful	Target audience include people active in the areas of precision agriculture, agriculture scientists, agriculture and pesticide management services.
Information on the existence (or not) of similar data	According to our knowledge there are no similar data provided as open data.
Information on the possibilities for integration and reuse of the data	The data could be reused for comparison purposes or being subject to different analyses in projects concerning the potential use of agriculture and food technologies

2.1.3. Standards and metadata

Typically all data come in a formal numerical or text format, conforming to specific rules depending on its intended use (e.g. comma separated fields). Documentation files are simple text (.txt) files.

Images and other multimedia files included in the profiles come in proper containers (e.g. JPEG file format for still images and MP4 for video sequences).

Metadata are associated to the various files in a direct or indirect manner. Direct association to metadata is supported from specific file types (e.g., compression ratio, image size, etc., in PNG files) whereas indirect association to metadata stems from external tools that are being used on data, such as versioning or management tools, and create metadata regarding modifications, commit information, etc.

2.1.4. Data access, sharing, and licensing

Decisions regarding data access and licensing strategies will be made in due time, once the datasets are available. Future versions of the present document will report all relevant details.

To facilitate a good level of collaboration between the consortium's partners, data repositories will be available at Synelixis SynField cloud platform.

DIATOMIC will use mainly two EU's public repositories, OpenAIRE[1] and Zenodo[2], for providing access to the DIATOMIC Agrifood Experiment dataset. However, the exact access policy has not been defined at this stage, as issues related to data privacy, confidentiality and anonymity have to be taken into consideration first. Access to this dataset will also be provided through DIATOMIC's website, either in the form of data queries or in the form of direct web links to the data repositories. The decision for the exact form has yet to be made.

In case DIATOMIC Agrifood Experiment dataset contains personal data (e.g., names, addresses, etc), it will be carefully examined by the consortium whether this data will be shared after a subject's consent or after data anonymization, or not at all.

__Note__: DIATOMIC consortium declares that it reserves the right to exclude any information belonging to the datasets described in the present deliverable (or later versions of it) from the dataset they belong to and/or from Open Access in general in case it interferes with ethical rules, or contains personal data, or raises IPR, commercial, privacy-related or security issues.

2.1.5. Dataset from PUSH and PULL experiments

The dataset described in this section mainly covers the DIATOMIC Agrifood PUSH experiments.

Similar data are expected to be provided also from the PULL experiments. Within the criteria for the evaluation of the open call proposals, the provision of relevant data as open data will also be considered. In this way, the PULL experiments will be motivated to provide open data, which will be published as extension of DIATOMIC PUSH experiments given that they fulfil the necessary ethical and privacy conditions.

---

[1]https://www.openaire.eu/

[2]http://zenodo.org/

2.2. Application Experiment: Healthcare and fitness

The dataset includes all data related to the Health Application Experiment of DIATOMIC.

2.2.1. Dataset nomenclature

The dataset Reference is "DIATOMIC_Health" and the dataset name is "DIATOMIC Health Experiment".

2.2.2. Dataset description and key attributes

The DIATOMIC Health Experiment dataset will consist of data collected by various sensors which will monitor health related parameters from selected Libelium employees within their facilities.

The dataset will be stored in Libelium Cloud Platform, where raw data and graphs will be displayed. Raw data will be sent to Virtual Fortknox where it will be stored and modified in order to fit HL7 standard. After this, information will be redirected to eVida in order to make it available for different applications. The dataset will also include a link to an image of the patient if provided.

This dataset's objectives are: to assess whether actions need to be taken in case the patient's state is problematic, to improve health and well-being outcomes, to promote healthy and active ageing.

Key attributes, characteristics, and other information related to DIATOMIC Health Experiment dataset:

Data set #2.5: DIATOMIC Healthcare and fitness related Data

Description of the data that will be generated or collected	Healthcare and fitness (Oxygen saturation, heart rate, weight, body fat, bone mass, muscle mass, visceral fat, water percentage, calories, diastolic pressure, systolic pressure).
Data origin (in case it is collected)	The different health parameters will be collected from some employees working in Libelium, sensors used will be weight scale, blood pressure sensor and oximeter.
Data nature	Data will be mainly in numeric format. There will be additional information to identify sensors and users.
Data scale	The data sent will be in range of 4-5 KB per measurement performed if all three sensors are used to take the measure.
To whom could the data be useful	Target users for these datasets can be eHealth application developers, medical researchers, doctors, data scientists that need to model healthcare data
Information on the existence (or not) of similar data	According to our knowledge there are no similar data provided as open data.
Information on the possibilities for integration and reuse of the data	The data could be reused for comparison purposes or being subject to different analyses in projects concerning the potential use of eHealth technologies

2.2.3. Standards and metadata

Data comes initially in json format, but will be transformed onto HL7 standard format on Virtual Fortknox platform in order to fit the international standards for transfer of clinical and administrative data between software applications.

Metadata are associated to the various files in a direct or indirect manner. Direct association to metadata is supported from specific file types (e.g., compression ratio, image size, etc., in PNG files) whereas

indirect association to metadata stems from external tools that are being used on data, such as versioning or management tools, and create metadata regarding modifications, commit information, etc.

Examples of metadata that can be collected in this dataset are date and time of measurements, user name, user surname, member ID, time of the last measurement performed, height, weight, birthday, department where the patient belongs to and an image url from the user profile in PNG format. All these included in case the user provides the mentioned information.

2.2.4. Data access, sharing, and licensing

Data access, sharing, and licensing policies for this dataset are similar to DIATOMIC Agrifood Experiment dataset corresponding policies described in 2.1.4. Data repositories will be hosted on Libelium Cloud Platform, whereas EU's OpenAIRE and Zenodo platforms and DIATOMIC's website will be used for providing public access.

Note: DIATOMIC consortium declares that it reserves the right to exclude any information belonging to the datasets described in the present deliverable (or later versions of it) from the dataset they belong to and/or from Open Access in general in case it interferes with ethical rules, or contains personal data, or raises IPR, commercial, privacy-related or security issues.

2.2.5. Dataset from PUSH and PULL experiments

The dataset described in this section mainly covers the DIATOMIC Health PUSH experiments.

Similar data are expected to be provided also from the PULL experiments. Within the criteria for the evaluation of the open call proposals, the provision of relevant data as open data will also be considered. In this way, the PULL experiments will be motivated to provide open data, which will be published as extension of DIATOMIC PUSH experiments given that they fulfil the necessary ethical and privacy conditions.

2.3. Application Experiment: Manufacturing

The dataset includes all data related to the Manufacturing Application Experiment of DIATOMIC.

2.3.1. Dataset nomenclature

The dataset Reference is "DIATOMIC_Manufacturing" and the dataset name is "DIATOMIC Manufacturing Experiment".

2.3.2. Dataset description and key attributes

The DIATOMIC Manufacturing Experiment dataset will consist of manufacturing orders, including a CAD-file and meta information regarding the order. Default data sets will be generated by the researchers. Further sets may be added by customers. The dataset will be processed by custom built software and a slicer "Slic3r" (under the GNU Affero General Public License, version 3). These software components will be hosted in the Virtual Fort Knox Cloud Platform.

The dataset will also include CAD files for incoming manufacturing orders and according meta data regarding the orders.

This dataset's objective is to act as input for the manufacturing experiment by triggering manufacturing orders with the associated data.

Key attributes, characteristics, and other information related to DIATOMIC Manufacturing Experiment dataset:

Data set #2.6: DIATOMIC Manufacturing Order related Data

Description of the data that will be generated or collected	Order Data (Design information, quantity, meta data such as material, colour and maximum dimensions)
Data origin (in case it is collected)	Default data sets will be provided by the drivers of the manufacturing application experiment. Further Data might be added by customers or partners for application experiments found via the open calls
Data nature	The data will be mainly in a numerical format
Data scale	The dataset's size mainly depends on the size of the associated CAD files. These may become as large as 30 MB or more.
To whom could the data be useful	Target audience includes researchers and developers in the area of 3D printing
Information on the existence (or not) of similar data	3D-service platforms may have similar data. However such data is not publicly available to our knowledge. Similar printable products as CAD-models can be found in 3D printing repositories (e.g. grabcad.com).
Information on the possibilities for integration and reuse of the data	The data could be reused as example data for other 3D-printing related experiments.

2.3.3. Standards and metadata

Typically all data come in a formal numerical or text format, conforming to specific rules depending on its intended use (e.g. comma separated fields). Documentation files are simple text (.txt) files.

Images and other multimedia files included in the profiles come in proper containers (e.g. JPEG file format for still images and MP4 for video sequences).

Metadata are associated to the various files in a direct or indirect manner. Direct association to metadata is supported from specific file types (e.g., compression ratio, image size, etc., in PNG files) whereas indirect association to metadata stems from external tools that are being used on data, such as versioning or management tools, and create metadata regarding modifications, commit information, etc.

2.3.4. Data access, sharing, and licensing

Data access, sharing, and licensing policies for this dataset are similar to DIATOMIC Agrifood Experiment dataset corresponding policies described in 2.1.4. Data repositories will be hosted on Virtual Fort Knox Cloud Platform, whereas EU's OpenAIRE and Zenodo platforms and DIATOMIC's website will be used for providing public access.

In case of personal data, the same strategy as in 2.1.4 will be followed.

***Note**: DIATOMIC consortium declares that it reserves the right to exclude any information belonging to the datasets described in the present deliverable (or later versions of it) from the dataset they belong to and/or from Open Access in general in case it interferes with ethical rules, or contains personal data, or raises IPR, commercial, privacy-related or security issues.*

2.3.5. Dataset from PUSH and PULL experiments

The dataset described in this section mainly covers the DIATOMIC Manufacturing PUSH experiments.

Similar data are expected to be provided also from the PULL experiments. Within the criteria for the evaluation of the open call proposals, the provision of relevant data as open data will also be considered. In this way, the PULL experiments will be motivated to provide open data, which will be published as extension of DIATOMIC PUSH experiments given that they fulfil the necessary ethical and privacy conditions.

3. DIATOMIC Digital Innovation Hubs Data

This section reports on the organization of the datasets related to the Digital Innovation Hubs (DIH) distributed throughout Europe. It will include information related to their expertise, services that they offer, contact details etc. In more details the DIATOMIC_DIH data include the following datasets:

3.1. DIH Dataset

The dataset Reference is "DIATOMIC_DIH" and contains information related to :

- DIH Expertise Data
- DIH Services Data
- DIH Contact Data

3.1.1. Dataset description and key attributes

The DIATOMIC DIH dataset will be stored and processed in INTRASOFT's Redmine platform so that date can be calibrated and additional information will be added. The dataset may also include images and videos from the DIH facilities.

The purpose of the collected datasets is to support end users, SMEs, researchers and developers that need DIH in order to validate their products or increase their TRL. Key attributes, characteristics, and other information related to DIATOMIC DIH dataset:

Data set #3.1: DIATOMIC DIH Data

Description of the data that will be generated or collected	It will include information on DIH Expertise and technical competence, services and TRL level that they may offer, potentially additional information such as reference projects, reference clients, images and video describing their facilities or achievements and contact details.
Data origin (in case it is collected)	The data will be collected from DIF applications, information that will be collected during the PUSH and PULL experiments along with data that could be collected from the European Commission services or the Internet.
Data nature	The data will be mainly text, numbers, URLs and potential some images and videos
Data scale	Currently there are about 520 European DIHs and we expect to have 10K to 30M per DIH based on the availability of facilities and previous products images and videos.
To whom could the data be useful	Target audience include end users, SMEs, researchers and developers that need DIH in order to validate their products or increase their TRL.
Information on the existence (or not) of similar data	An initial set of data is available at the S3 (Smart Specialisation Platform) http://s3platform.jrc.ec.europa.eu/digital-innovation-hubs-tool
Information on the possibilities for integration and reuse of the data	The data could be reused for comparison purposes or being subject to different analyses in projects concerning the potential utilization of DIH

3.1.2. Standards and metadata

Typically all data come in a formal numerical or text format, conforming to specific rules depending on its intended use (e.g. comma separated fields). Documentation files are simple text (.txt) files.

Images and other multimedia files included in the profiles come in proper containers (e.g. JPEG file format for still images and MP4 for video sequences).

Metadata are associated to the various files in a direct or indirect manner. Direct association to metadata is supported from specific file types (e.g., compression ratio, image size, etc., in PNG files) whereas indirect association to metadata stems from external tools that are being used on data, such as versioning or management tools, and create metadata regarding modifications, commit information, etc.

3.1.3. Data access, sharing, and licensing

Decisions regarding data access and licensing strategies will be made in due time, once the datasets are available. Future versions of the present document will report all relevant details.

To facilitate a good level of collaboration between the consortium's partners, data repositories will be available at INTRASOFT's Redmine platform.

DIATOMIC will use mainly two EU's public repositories, OpenAIRE and Zenodo, for providing access to the DIATOMIC DIH dataset. However, the exact access policy has not been defined at this stage, as issues related to data privacy, confidentiality and anonymity have to be taken into consideration first. Access to this dataset will also be provided through DIATOMIC's website, either in the form of data queries or in the form of direct web links to the data repositories. The decision for the exact form has yet to be made.

In case DIATOMIC DIH dataset contains personal data (e.g., names, addresses, etc), it will be carefully examined by the consortium whether this data can be shared after a subject's consent or after data anonymization, or not at all.

__Note__: DIATOMIC consortium declares that it reserves the right to exclude any information belonging to the datasets described in the present deliverable (or later versions of it) from the dataset they belong to and/or from Open Access in general in case it interferes with ethical rules, or contains personal data, or raises IPR, commercial, privacy-related or security issues.

.

4. DIATOMIC Open Calls Data

This section reports on the organization of the datasets of the DIATOMIC Open Calls data that will be conducted in the context of DIATOMIC project.

In more details the DIATOMIC Open Call data include the following dataset:

4.1. Open Calls Dataset

The dataset Reference is "DIATOMIC _Open Calls" and will consist of data related to:

- Submitted proposals statistics
- Evaluation statistics
- Selected projects statistics

The dataset will be stored and processed by F6S so that they can be calibrated and additional information will be generated. Whatever sensitive information is included (e.g. person names) will be anonymized in order to keep privacy according to the DIATOMIC Ethical guidelines.

The purpose of the collected datasets is to support incubators, SME accelerators, projects with open call procedures, national research secretariats and Crowd funding platforms though out Europe.

Key attributes, characteristics, and other information related to DIATOMIC Open Calls dataset:

Data set #4.1: DIATOMIC Proposals Data

Description of the data that will be generated or collected	Proposals data including the number of proposals that we have been received, their country/region of origin, their focus and distribution in the DIATOMIC selected application areas. Of course, any information related to names and contact details of the proposals participants will be excluded in order to keep the anonymity of the applicants.
Data origin (in case it is collected)	The proposals data will be collected from F6S during the open call submission procedure.
Data nature	The data will be mainly in a text format
Data scale	The data are considered to be in the range of 100KBytes per proposal.
To whom could the data be useful	Target audience include incubators, SME accelerators, projects with open call procedures, national research secretariats and Crowd funding platforms though out Europe.
Information on the existence (or not) of similar data	We are not aware of any such data availability as open data
Information on the possibilities for integration and reuse of the data	The data could be reused for comparison purposes or being subject to different analyses in projects concerning open calls or incubators' platforms

Data set #4.2: DIATOMIC Proposals Evaluation Data

Description of the data that will be generated or collected	This data set will include the evaluation process details (e.g. time per proposal evaluation, time per consensus meeting, rate of success, average requested funding, number of participating companies etc.) It will also include the characteristics of the evaluators that DIATOMIC has selected (i.e. nationality, affiliation types, age category, specialty group, gender, etc.). Of course any information related to names and contact details of the evaluators will be excluded in order to keep the anonymity of the evaluators.
Data origin (in case it is collected)	The proposals data will be collected from F6S during the open call submission procedure.
Data nature	The data will be mainly in a text format
Data scale	The data are considered to be in the range of 100KBytes per proposal and 100Kbytes per evaluator.
To whom could the data be useful	Target audience include incubators, SME accelerators, projects with open call procedures, national research secretariats and Crowd funding platforms though out Europe.
Information on the existence (or not) of similar data	We are not aware of any such data availability as open data
Information on the possibilities for integration and reuse of the data	The data could be reused for comparison purposes or being subject to different analyses in projects concerning open calls or incubators' platforms.

Data set #4.3: DIATOMIC Projects Data

Description of the data that will be generated or collected	This data set will include the characteristics of the accepted projects per application area, the level of funding that will receive and small description of what they will deliver. Of course any information related to names and contact details of the project participants will be excluded in order to keep the anonymity of the evaluators.
Data origin (in case it is collected)	The proposals data will be collected from F6S during the open call submission procedure.
Data nature	The data will be mainly in a text format
Data scale	The data are considered to be in the range of 500KBytes per project.
To whom could the data be useful	Target audience include incubators, SME accelerators, projects with open call procedures, national research secretariats and Crowd funding platforms though out Europe.
Information on the existence (or not) of similar data	We are not aware of any such data availability as open data
Information on the possibilities for integration and reuse of the data	The data could be reused for comparison purposes or being subject to different analyses in projects concerning open calls or incubators' platforms.

4.1.1. Standards and metadata

Typically all data come in a formal text format, conforming to specific rules depending on its intended use (e.g. comma separated fields). Documentation files are simple text (.txt) files.

Images (e.g. logos) in the profiles come in proper containers (e.g. JPEG file format for still images).

4.1.2. Data access, sharing, and licensing

Decisions regarding data access and licensing strategies will be made in due time, once the datasets are available. Future versions of the present document will report all relevant details.

In case DIATOMIC Open Calls dataset contains personal data (e.g., names, addresses, etc), it will be carefully examined by the consortium whether this data will be shared after a subject's consent or after data anonymization, or not at all.

__Note__: DIATOMIC consortium declares that it reserves the right to exclude any information belonging to the datasets described in the present deliverable (or later versions of it) from the dataset they belong to and/or from Open Access in general in case it interferes with ethical rules, or contains personal data, or raises IPR, commercial, privacy-related or security issues.

5. Publications and reports

The dataset includes all files (source code, final documents) related to the reporting and dissemination activities that will take place in DIATOMIC.

5.1. Data nomenclature

The dataset Reference is "DIATOMIC_Publications" and the dataset name is "DIATOMIC Publications".

5.2. Dataset description

The DIATOMIC Publications dataset consists of material regarding:

- Scientific Publications, including papers, presentations, posters, and other material used in scientific conferences, journals, workshops, etc. In every case, a direct link to the online version of the material will be provided.
- Educational and Promotional purposes, including presentations, lectures, and other material used in an educational institution's premises, as well as material related to promotional activities (flyers, posters, etc).
- Project deliverables, including all reports and related files which form DIATOMIC's deliverables as specified in the project's description of work.

5.3. Standards and metadata

Data come in text format, conforming to the intended use needs and to the requirements of the software application that will be used to render them in a readable and printable format (e.g. Portable Device Format - pdf).

Images and other multimedia files come in proper containers such as JPEG for static images.

Metadata are associated to the various files in a direct or indirect manner. Direct association to metadata is supported from specific file types (e.g., compression ratio, image size, etc., in PNG files) whereas indirect association to metadata stems from external tools that are being used on data, such as versioning or management tools, and create metadata regarding modifications, commit information, etc.

5.4. Data sharing

To facilitate a good level of collaboration between the consortium's partners and also to encourage future work on the project's findings, DIATOMIC Publications source files will be available at INTRASOFT's Redmine platform which will serve as the main hub. DIATOMIC intends to also make use of ResearchGate[3] platform and is also considering on submitting any publications to Open Access journals so as to use their networks to share results. Regarding submissions to closed access journals or conferences, the consortium will attempt to make publicly available any versions of them eligible for open access (e.g. postprint versions). However, only educational material will be made publicly available as it is the consortium's belief that publications and other material (e.g., promotional material) under DIATOMIC's name should be controlled.

---

[3]https://www.researchgate.net/

Moreover, according to the project's description of work, in contrast to public deliverables, deliverables that have been characterized as confidential will not be shared. The table below provides an overview of the two deliverable categories, public and confidential.

Data set #5.1: DIATOMIC Deliverables

	Public Deliverables		Confidential Deliverables
D1.3	Data Management Plan (DMP) (1st version)	D1.1	DIATOMIC Project Management Handbook
D1.4	Data Management Plan (DMP) (2nd version)	D1.2	Periodic report
D1.6	Data Management Plan (DMP) (final version)	D1.5	Final report
D4.1	Third-Party Financing Rules	D2.2	DIATOMIC DIHs Network Expansion Strategy
D4.2	Financial Oversight Report (first version)	D2.3	Quality Assurance Report (1st version)
D4.4	Open calls Report (first version)	D2.5	Quality Assurance Report (2nd version)
D4.5	Open calls Report (final version)	D2.6	Quality Assurance Report (final version)
D4.7	Financial Oversight Report (final version)	D3.1	DIATOMIC PUSH experiments coordination and technology requirements identification
D6.1	Public Outreach Strategy	D3.5	DIATOMIC experiments assessment
D6.2	PULL Call Dissemination Report	D4.6	Experiments evaluation results (1st version)
D6.3	Report on DIATOMIC Ecosystem activities (first version)	D4.8	Experiments evaluation results (final version)
D6.7	Report on DIATOMIC Ecosystem activities (final version)	D5.1	DIATOMIC Design-Development-Market report (1st call)
		D5.2	DIATOMIC Design-Development-Market report (2nd call)
		D6.4	Report on Portfolio Management activities per call (1st call)
		D6.5	DIATOMIC Growth and Sustainability Plan (first version)
		D6.6	DIATOMIC Growth and Sustainability Plan (final version)
		D6.8	Report on Portfolio Management activities per call (2nd call)
		D7.1	GEN - Requirement No.1
		D7.2	H - Requirement No. 2

6. Archiving and preservation

Each owner partner involved in a dataset will be responsible for the archiving and preservation of the corresponding data. After the completion of the project, the data collected will be maintained for a period determined by each partner.

The project coordinator (INTRASOFT) will be responsible for archiving and maintaining the project data files after the completion of the project, for a period covering at least two years. The DIATOMIC web-site will be maintained for at least two years after the completion of the project and all the information hosted in the web-site will be maintained by BIOS. Also, the storage server with all project files (including the Datasets) will host a snapshot of all the files produced for at least two years after the completion of the project. Once the push-experiment datasets of the project are moved to a public repository, INTRASOFT will be responsible for the maintenance of any new data that will be created.

At this stage of the project, the volume of the data that will be collected cannot be safely determined. The same stands for the associated costs which are tightly connected to the data volume. The Consortium thus far considers the costs of:

- Domain name registration and web hosting fees for the DIATOMIC web site, which have already been covered for the duration of the project by SYN.

- The repository provided by INTRASOFT is hosted on private infrastructure and thus, the availability is guaranteed for the project duration. It will be put offline after project completion for maintenance and bandwidth cost reduction reasons, however it will still be available for internal use, conforming to each partner's archiving practices. DIATOMIC website and a final snapshot will be kept in BIOS storage server whereas the source code final snapshot will be hosted on a public code repository (e.g., Zenodo). Data volume is not expected to mandate the use of special systems, therefore storage costs for the final source code repository will be covered by each partner.

Zenodo supports Closed, Open and Embargoed[4] Access. However, only Open Access uploads are displayed on the Zenodo website front-page. Closed Access uploads are still discoverable through search queries, their DOI[5], and any community collections where they are included. Metadata is licensed under Creative Commons "No Rights Reserved" license (CC0)[6], except for email addresses. All metadata is exported via Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)[7] and can be harvested. Access to metadata and data files is provided over standard protocols such as HTTP.

Zenodo accepts data under a variety of licenses, but extra benefits, in terms of visibility and credit, and additional services and upload quotas are offered to data deposited under the most open licenses.

Note: *The full set of Zenodo "Terms of Use" can be found online at* *https://zenodo.org/terms while the full set of* *Zenodo Policies can be found online at* *https://zenodo.org/policies.*

---

[4] Users may deposit content under an embargo status and provide and end date for the embargo. The repository will restrict access to the data until the end of the embargo period; at which time, the content will become publicly available automatically.

[5] A Digital Object Identifier (DOI) is a serial code used to uniquely identify objects. Further information can be found at https://en.wikipedia.org/wiki/Digital_object_identifier.

[6] The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a low-barrier mechanism for repository interoperability. Further information can be found at: https://www.openarchives.org/pmh/.

[7] The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a low-barrier mechanism for repository interoperability. Further information can be found at: https://www.openarchives.org/pmh/.

7. Ethics and confidentiality

Any ethical issue that might arise will be dealt by DIATOMIC with the appropriate care, in a professional way, following very closely established EU regulations and corresponding national laws about privacy, confidentiality, and consent.

All personal and professional data will be addressed by the consortium to ensure that they do not contravene national laws on the protection of this type of data. All systems developed in DIATOMIC will conform to both current legislation and that anticipated in the future. The systems that will be developed in DIATOMIC will consider "Trust and confidence" as a key attribute due to the sensitivity of personal and business confidential information in terms of exploitation. Therefore, extra care will be taken into preserving this data.

Important notice: With respect to Serbian partners (BIOS and InoSens) and activities held in Serbia, all activities related to data protection will be performed in accordance with the Serbian national *Law on personal data protection* (Official Gazette of RS no.97/2008, 104/2009, 68/2012 and 107/2012), which in Article 10 emphasizes that consent (written or oral) to data processing is deemed to be valid if given by a person who has received prior information from the collector of the data. Article 15 of the same Law provides the details on what this prior information has to include (e.g. the identity of the interviewer, purpose of data collection/processing, how data will be used, who will use the data, is data provided on voluntary base, etc.).

All the above have been sufficiently detailed in the DoA (Section 5) and will be taken into consideration in executing our data management plan.

8. Conclusions

In this deliverable, we have presented the data that will be generated in the context of DIATOMIC including data from all three push application experiments. Our data management plan is built upon analysing the generated data with respect to their confidentiality level and employing a different sharing approach depending on this level. Data considered confidential or rated as not shareable according to ethical considerations will not be shared. Finally, the DMP strictly builds on ensuring the necessary informed consents, as well as respecting the sphere of privacy of each participant. This document will be further developed, as we gain more information about the specific requirements of our subjects and the usage scenarios implemented by the DIATOMIC system.

9. References

[1] DIATOMIC Consortium, "DIATOMIC Description of Work," Approved by EC on 27 April 2017

[2] DIATOMIC Consortium Deliverable D7.1 "GEN – Ethics Requirement No. 1", September 2017

10. Annex I: ZENODO Policies

As Zenodo will be among the major data publishing platforms for DIATOMIC, this appendix provides information related to ZENODO Archiving and preservation Policies.

10.1. Retention period

Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.

10.2. Functional preservation

ZENODO makes no promises of usability and understandability of deposited objects over time.

10.3. File preservation

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

10.4. Fixity and authenticity

All data files are stored along with a MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant.

10.5. Succession plans

In case of closure of the repository, best efforts will be made to integrate all content into suitable alternative institutional and/or subject based repositories.

10.6. Cost

ZENODO is free for the long tail of Science. In order to offer services to the more resource hungry research, we will introduce a ceiling to the free slice and offer paid for slices above, according to the business model developed within the sustainability plan.