HORIZON 2020	**HORIZON 2020** **Self management of healthand disease:** **citizen engagement and mHealth**

Project:

myAirCoach - Analysis, modelling and sensing of both physiological and environmental factors for the customized and predictive self-management of Asthma"

(myAirCoach, Grant Agreement No. 643607)

Deliverable number and title:

D7.6 Data Management Plan	
Lead beneficiary:	CERTH/ITI
WP. no, title and activity type	WP7–Dissemination and Exploitation
Contributing Task (s)	T7.3 IPR Management T7.4 Standardization and Concertation Actions
Dissemination level	PU: [Public]
Delivery date	December 2015
Status	Final Version
File name and size	“MyAirCoach-WP7-D7.6-Data_Management_Plan.pdf”, 1.39 MB

Authors List

Leading Authors(*Editor*)		
Name / Surname	*Beneficiary Name (Short Name)*	*Contact email*
Dimitrios Tzovaras	CERTH	Dimitrios.Tzovaras@iti.gr
Konstantinos Votis	CERTH	kvotis@iti.gr
Dimitrios Kikidis	CERTH	dkikidis@iti.gr
Paliokas Ioannis	CERTH	ipaliokas@iti.gr
Lalas Antonios	CERTH	lalas@iti.gr
Kamadanis Nikolaos	CERTH	kamadan@iti.gr
PavlopoulosVissarion	CERTH	apavlo@iti.gr
Polichronidou Eleutheria	CERTH	epolyc@iti.gr
PapadopoulouEvdoxia	CERTH	evdoxia@iti.gr
Sotiris Kamaris	CERTH	skamaris@iti.gr

Executive Summary

The current deliverable is directly connected with the work performed under the Task 7.4 “Standardization and concertation actions” and serves as the initial plan for the collection, organization, storing and sharing of the knowledge and data created within the project.The described data management plan is based on several inputs, namely : a) the MyAirCoach Description of Action (DOA) document, b) guidelines of the European Commission for the data management of H2020 research projects, c) the outcomes of the plenary project meetings and d) the input from several informal discussions among the project consortium members.

The data management requirements and standardization guidelines specified in this document are expected to form a reference manual to be used throughout the project. In this way, MyAirCoach is aiming to develop a stable, reliable and easy to use platform which will form an open repository for asthma research and extend beyond the framework of the current project’s timeline.

Finally, it is important to underline that the current deliverable will be a living document which will be continuously adapted depending on the needs of the project research and development objectives, and based on the direct input from members of the consortium and actual system users.

Table of contents

List of figures

List of Tables

List of abbreviations and acronyms

(*in alphabetic order*)

ATS	American Thoracic Society
CT	Computed Tomography
DICOM	Digital Imaging and Communication in Medicine
DMP	Data Management Plan
DOI	Digital Object Identifier
EU	European Union
HRCT	High Resolution Computed Tomography
HRDSP	Harvard Research Data Security Policy
IT	Information Technology
MDI	Metered-Dose Inhaler
MRI	Magnetic Resonance Imaging
NHS	National Health Service
NICE	National Institute for Health and Care
NO	Nitric Oxide
OpenAIRE	Open Access Infrastructure for Research in Europe
OWL	Ontology Web Language
PAKE	Password Authenticated Key Exchange
PET	Positron Emission Tomography
SRP	Secure Remote Password
TB	Terabyte
UK	United Kingdom
US	United States
VUMS	Virtual User Modelling and Simulation Standardization
XML	EXtensible Markup Language
XSD	XML Schema Definition

1 Introduction

The MyAirCoach project is aiming to support the research in the field of personalized self-management of health and more specifically develop an innovative solution for the effective and efficient management of asthma. In this direction and based on the project's description of work a number of datasets are going to be collected and utilized for the support of both the development and research tasks of the project. Therefore, it is considered of fundamental importance to define the framework for the collection, organization and sharing of such information in order to increase their long term usability within the project partners but more importantly by the entire research community.

Firstly, the current deliverable is aiming to provide concise summaries of the types of datasets that are expected to be used during the project. These datasets will form the basis for the design, development and testing of the MyAirCoach system and in addition will be used for the academic research activities foreseen by the consortium.

In the second part of the document important issues of the MyAirCoach Data Management Plan (DMP) are discussed in order to outline the specific requirements and guidelines that should be followed throughout the project's timeline.

The proposed plan was designed to allow the efficient dissemination of results and the stimulation of research without jeopardizing any ethical requirements of the project or decreasing the commercial value of the overall MyAirCoach solution.

More specifically the MyAirCoach data management plan is aiming to:

1. Outline the responsibilities for data protection and sharing within an ethical and legal framework.
2. Do not interfere with the protection of the intellectual property created by the project.
3. Support open access to the project's research outcomes and scientific publications
4. Support the openness of data related to both the publications and the development processes of the project
5. Define a documentation framework for the annotation of the collected knowledge towards increased discoverability and validation
6. Allow the creation of an online platform that will support all the above functionalities

Finally, the first version of the online MyAirCoach open portal is presented with special focus on the access to open data by both registered and external users. As the development tasks of the project will be evolved this platform will be enhanced with additional functionalities regarding the data management capabilities but also with additional datasets and links with data from other external sources.

2 MyAirCoach Principles of Data Management

2.1 Data Management Requirements

This section describes the requirements and principles that will form the basis upon which the MyAirCoach data management plan has been defined. More specifically the current deliverable has been based on the guidelines of the EU Commission regarding the openness of the data generated from a project that has been funded by the H2020 framework[1]. According to these guidelines the scientifically-oriented data that are going to be generated by the MyAirCoach project will be formed so that they can be easily **discoverable**, **accessible**, **assessable** and **intelligible**, **usable** beyond the original purpose of their collection and usage but also **interoperable** to appropriate quality standards.

Furthermore and due to the health oriented nature of the project two additional but equally important attributes will be taken into account, the **data security** and the **preservation of the participants' privacy**. In this direction, all the the collected medical and sensitive data of patients will be be protected from any unauthorized access but also they will be carefully anonymized in order to be shared through the proposed open data management platform of the project.

In any case the publication of data should always conform with the ethical guidelines of the MyAirCoach project as they were already defined in D8.5 "Ethics, Safety and mHealth Barriers Manual" deliverable.

2.2 EU Commission Guidelines for data management

The EU Commission has published some guidelines for appropriate data management plans in Horizon 2020 projects. This guide is structured as a series of appropriate questions that should be ideally clarified for all datasets produced in any h2020 project. The following Table 2 presents the defined questions' set per aspect along with a comment validating the conformance of the MyAirCoach project to these questions.

Table 1: EU Commission Data Management Plan Guidelines and Assurance of MyAirCoach Conformance

Aspect	Question
Discoverable	DMP question: are the data and associated software produced and/or used in the project discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier)?
Accessible	DMP question: are the data and associated software produced and/or used in the project accessible and in what modalities, scope, licenses (e.g. licensing framework for research and education, embargo periods, commercial exploitation, etc.)?
Assessable and	DMP question: are the data and associated software

intelligible	produced and/or used in the project assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review (e.g. are the minimal datasets handled together with scientific papers for the purpose of peer review, are data is provided in a way that judgments can be made about their reliability and the competence of those who created them)?
Usable beyond the original purpose for which it was collected	DMP question: are the data and associated software produced and/or used in the project useable by third parties even long time after the collection of the data (e.g. is the data safely stored in certified repositories for long term preservation and curation; is it stored together with the minimum software, metadata and documentation to make it useful; is the data useful for the wider public needs and usable for the likely purposes of non-specialists)?
Interoperable to specific quality standards	DMP question: are the data and associated software produced and/or used in the project interoperable allowing data exchange between researchers, institutions, organizations, countries, etc. (e.g. adhering to standards for data annotation, data exchange, compliant with available software applications, and allowing re-combinations with different datasets from different origins)?

2.3 Principles of medical information security

In order to adapt the requirements for openness of data without jeopardizing the rights of the participating patients the principles for the security of medical information (provided by the British Medical Association[2]) were adopted, as defined below:

Principle	Description
Access control	Each identifiable clinical record shall be marked with an access control list naming the people or groups of people who may read it and append data to it. The system shall prevent anyone not on the access control list from accessing the record in any way.
Record opening	A clinician may open a record with herself and the patient on the access control list. Where a patient has been referred, he/she may open a record with herself, the patient and the referring clinician(s) on the access control list.
Control	One of the clinicians on the access control list must be marked as being responsible. Only she may alter the access control list, and she may only add other health care professionals to it.

Consent and notification	The responsible clinician must notify the patient of the names on his record's access control list when it is opened, of all subsequent additions, and whenever responsibility is transferred. His consent must also be obtained, except in emergency or in the case of statutory exemptions.
Persistence	No one shall have the ability to delete clinical information until the appropriate time period has expired.
Attribution	All accesses to clinical records shall be marked on the record with the subject's name, as well as the date and time. An audit trail must also be kept of all deletions.
Information flow	Information derived from record A may be appended to record B if and only if B's access control list is contained in A's.
Aggregation control	There shall be effective measures to prevent the aggregation of personal health information. In particular, patients must receive special notification if any person whom it is proposed to add to their access control list al--- ready has access to personal health information on a large number of people.
Trusted Computing Base	Computer systems that handle personal health information shall have a subsystem that enforces the above principles in an effective way. Its effectiveness shall be subject to evaluation by independent experts.

2.4 Actors

An important step towards the accurate and relevant definition of the data management plan is the identification of all related actors that may be involved in the formation and usage of the project's online open access repository. The following Table 2 presents the actors of the MyAirCoach online platform for accessing and uploading datasets. Each category has its own distinctive characteristics that should be taken into consideration. The basic actors are patients and health care professionals who are the ones directly involved in the management and control of the asthma disease. Researchers dealing with aspects of asthma are also included along with external users who will include commercial entities such as health oriented technology providers and entrepreneurs.

Table 2: MyAirCoach open access actors

Actors	Description
Patients	People who have asthma and are subjects of clinicians' care
Patient families	People in the close environment of patients who are given, by the patients, the right to access their medical record
Health care	Doctors, nurses, trainers, administrative personnel having direct

professional	contact with and responsibility for patients
Researchers	Research institutes, individual researchers, and in general people investigating aspects of asthma
External	Third party users of MyAirCoach data for technology development purposes, such as commercial entities and entrepreneurs

2.5 Self-Audit Process

The Caldicott Report[3] will serve as a guideline for the self-audit processes of the datasets produced within MyAirCoach. The Caldicott report sets out a number of general principles that health and social care organizations should use when reviewing their use of client information. The report makes several recommendations, one of which is the appointment of Caldicott guardians, i.e. members of staff with a responsibility to ensure patient data is kept secure. It is now a requirement for every NHS organization to have a Caldicott guardian.

Within myAirCoach project the ethical advisory board but also the corresponding Patient forum will be in charge of the execution of the defined data management plan and will act as a "Caldicott guardian" supervising the compliance with legal and ethical issues in terms of information security, data protection and ethical issues. Except the datasets produced by the project, the users of the myAirCoach system will be able to upload their own datasets. Thus, the existence of an auditing mechanism is deemed very critical in order to avoid the publication of non-validated clinical data or data collected from campaigns that do not comply with the ethical manual of the MyAirCoach project.

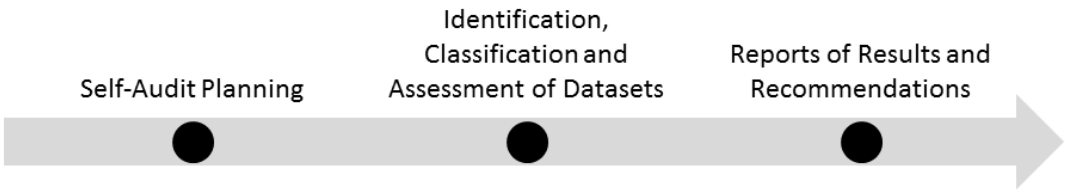

Figure : Self-Audit Process for MyAirCoach Datasets

The steps of the Self-Audit process that will be implemented are summarized below:

- Self-Audit Planning
  - Plan and Set-up Self-Audit
  - Collect Relevant Documents
- Identification, Classification and Assessment of Datasets
  - Analyze Documents
  - Identify Data Sets
  - Classify Data Sets
  - Assess Data Sets
- Report of Results and Recommendations
  - Collate and analyze information from the audit
  - Report on the compliance with the Data Management Plan
  - Identify weaknesses and decide on corrective actions

2.6 Risk Assessment

Data management is directly connected with issues of privacy and as such it should be aiming to the efficient and early identification of risks and their timely solution through appropriate strategies. Initially, the data objects need to be categorized based on the identifying and sensitive information that they contains in order to selected the appropriate mitigation strategies. MyAirCoach will be using the Harvard Research Data Security Policy(HRDSP) scale[4] for the characterization of the risks associated with the privacy of participants.

After categorizing the data objects, the risks related to each category should be determined. The risk analyses and mitigation strategies will be considered separately for every dataset so that the finally publishable data are categorized to Level 1.

Table 3: Categorization of datasets in regards to privacy

HRDSP	Description	MyAirCoach publication rights
Level 1	De-identified research information about people and other non-confidential research information	Can be published on the open access platform
Level 2	Benign information about individually identifiable people	Can be shared within the consortium
Level 3	Sensitive information about individually identifiable people	Can be shared within the consortium
Level 4	Very sensitive information about individually identifiable people	Can be used by the responsible clinical partner only
Level 5	Extremely sensitive information about individually identifiable people	Can be used by the responsible clinical partner only

2.7 Context Categorization of Data

The research data that will be collected or generated during the project lifecycle can be categorized in four groups regarding their context and the informational weight.The Table 4 presents a summary of the categories identified for the categorization of the data collected within the MyAirCoach project.

Table 4: Context categorization of myAirCoach Data

Category	Description	Examples
Raw Collected Data	Obtained data that has not been subjected to any quality assurance or	Measurements collected from sensors/devices (e.g. smart bracelets, sensor

	control	enhanced MyAirCoach inhaler)
Verified/Validated Collected Data	These are the raw data that has been evaluated for completeness, correctness, and conformance/compliance of a specific data set against the standard operating procedure (verified), as well as reviewed for specific analytic quality (validated)	Annotated sensor measurements, Images from patients' tomographies, documents from test campaigns, asthma action plans etc.
Analyzed Collected Data	Validated data are then analyzed, through statistical operations, based on a specific target or application scenario	PatientModels, assessments of inhaler usage, patients' nutritional assessments etc.
Generated Data	The data needed to validate the results presented in scientific publications (pseudo-code, libraries, system design, , etc)	Mutli-parametric indicators of asthma control, algorithmic approaches for the detection of inhaler actuations, workflow for the deployment of User Centered Design in mHealth applications.

3 MyAirCoach Data Management Plan

The current chapter is aiming to provide a detailed description of all the foreseen MyAirCoach datasets through the use of the template of DMP established by the European Commission for Horizon 2020[1]. The definition of all the related aspects of dataset categories (Table 5) indicates the importance long term preservation of data and the requirement widest possible sharing of the knowledge created by EU projects.

Table 5: H2020 Template for Data Management Plan[1]

Aspect	Description
Data set reference and name	Identifier for the data set to be produced
Data set description	Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.
Standards and metadata	Reference to existing suitable standards of the

	discipline. If these do not exist, an outline on how and what metadata will be created.
Data sharing	Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.). In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).
Archiving and preservation (including storage and backup)	Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

In order to indicate the position of the datasets within the MyAirCoach and describe their envisioned use toward the project objectives a number of fields were introduced to the above template as indicated in Table 6

Table 6: MyAirCoach additional aspects of Data Management

Aspect	Description
Relation to the objectives of MyAirCoach	This aspect is introduced in order to provide a summary on how the specific dataset is going to be used within the project and how it is expected to contribute for the successful delivery of the project objectives.
Related Work Packages	List of all the related tasks and work packages of the project's description of work that are related to the specific type of data
Ethical issues and requirements	Description of any ethical requirements and suggestions for mitigation strategies in the case of identified risks.

In order to facilitate the easy use of the datasets through different platforms and operation systems a naming scheme has been proposed for all uploaded files. More specifically the following convention has been selected for the purposes of MyAirCoach and for the files uploaded on the online open access repository.

"[Dataset prefix]_[ID]_[Date]_[Author].[ext]"	
Dataset prefix	is the prefix of the specific type of dataset as identified in Table 7
ID	is the identification number as it is assigned by the online system
Date	is date of upload on the online system following the format: YYMMDD
Author	is the authors username
ext	is the file extension pertaining to the format used.
	The selected names should not include spaces or symbols with the only exception of the underscore

Table 7 summarises the prefixes for the foreseen categories of MyAirCoach datasets alongside a short description of the nature of the specified datasets.

Table 7: Naming Prefixes of Dataset Categories

No	Naming Prefix	Description
01	*DS_InhalerUsage*	Datasets related to inhaler usage measurements including both the time and technique of use
02	*DS_Physiology*	Datasets of physiology assessments including both sensor measurements and doctor diagnosis and comments.
03	*DS_Lifestyle*	Datasets related to the lifestyle of asthma patients with special focus on activity levels
04	*DS_Nutritional*	Datasets containing information regarding nutritional aspects of asthma patients
05	*DS_ExhaledNO*	Datasets of Exhaled Mitric Oxide Measurements of asthma patients and healthy subjects.
06	*DS_Environmental*	Datasets of Environmental Measurements.
07	*DS_Tomography*	Datasets of Patient Tomography of the Lungs.
08	*DS_PatientModels*	Datasets containing indicative patient models to be used for the multi-parametric description of asthma.
09	*DS_EducationAndTraining*	Datasets of Educational and Training Content describing the disease of asthma and the proper use of different types of inhalers.

10	*DS_ActionPlans*	Dataset of asthma action plans and medication strategies prescribed by doctors.
11	*DS_UserRequirements*	Datasets containing outcomes and information related to the assessment of user requirements and feedback sessions within the UCD process
12	*DS_TestCampaigns*	Other Datasets collected during the Test Campaigns of the project categorized with regards to the collection site (first and second test campaigns).

3.1 Datasets of Inhaler Usage Measurements

Name	Dataset of Inhaler Usage Measurements
Naming Prefix	DS_InhalerUsage
Summary	The current type of dataset will include measurements and data collected in regards to the use of inhaler by patients. More specifically, it is expected to include sound and acceleration measurements from sensors attached on the inhaler device.
Positioning within the MyAirCoach project	
Relation to the project objective	MyAirCoach is aiming to develop novel algorithmic approaches for the automatic detection of inhaler actuations and the analysis of the technique of use. It is therefore considered of fundamental importance to produce a dataset from testing sessions which will be used not only for the training of machine learning approaches but also the validation of results.
Related Work Packages	**WP3** Smart sensor based inhaler prototype and WBAN **WP4** Computational models, intelligent information processing and DSS module **WP6** Evaluation
Description of Dataset Category	
Origin of Data	Raw data will be collected by sensing elements attached on the inhaler devices. The annotation of collected data for the detection of actuation events and the characterization of inhaler technique will be done by experienced researchers.
Nature and scale of data	The data of this category will be in the form of time series describing measured parameters during the actual use of

	an inhaler. CSV (Comma Separated Values) is the advised file format in this category since it allows the easy use of the data both through programming languages and spreadsheet software packages (e.g. Open Office Calc, Microsoft Excel). In this case timestamps for every measurement or the sampling rate should be defined. For the specific case of sound measurements commonly used formats of sound representation can be also considered with WAV being the advised option. The annotation files are advised to be stored in the CSV format corresponding to the actual time series of data or in XML format for the identification of positioning of start and stop of events and user actions (e.g. breath-in, inhaler actuation)
Use by researchers and healthcare professionals	The datasets in this category can support research in the field of biomedical signal processing and serve as a basis for the comparative validation of different algorithmic approaches. Furthermore, the current type of datasets can be used for the testing of the accuracy of possible commercial products that rely on the same sensing capabilities. Finally, the annotation of the data as it relates to the technique of inhaler use by patients can be used as indicators of common errors of patients while using their inhaler.
Indicative existing similar dataset	There have not been identified any online available datasets in this category and for any method of sensing.
Indicative scientific publications	Unfortunately a very small number of publications are available in this field of studies and they are mainly focusing on the understanding of Dry Powder Inhalers (DPIs)[5,6], with only one identified exception of a scientific article monitoring the use of Metered Dose Inhaler (MDI)[7]
Standards and Metadata	
	The dataset will be accompanied by detailed documentation of its contents along with metadata describing the demographics of the samples from which the data were generated and detailed description of the data collection process. Indicative metadata include: (a) description of the experimental setup and procedure that led to the generation of the dataset, (b) documentation of the

	variables recorded in the dataset. The metadata will be in a format that may be easily parsed with open source software.
Data Sharing	
Access type	In accordance with the ethical and legal requirements regarding data obtained from human participants, the dataset will be initially available to the Consortium Members and only after its careful anonymization. Personal information regarding the participants will be kept strictly private. As the project progresses and the collected data are used for the research and development processes of the project they will become available at the projects open data platform after the approval by the ethics committee of the MyAirCoach project. The inclusion of a subject's data in the public part of this dataset will be done on the basis of appropriate informed consent to data publication
Access procedure	In the first stages of the dataset sharing, and as soon it reaches an anonymized formed, it will be shared among the consortium through the wiki page of the project. For the second stage of dataset publication, the anonymized data will be published through the open data platform of the project in order to be used by registered users and subsequently by any interested party aiming to use them for research and development.
Embargo periods (if any)	No preset embargo periods. Selection of the appropriate time of publication based on the research and development timeline of the project, the protection of intellectual property and the proper safeguarding of the privacy of participants
Technical mechanisms for dissemination	The public part of the datasets in this category will be accessible through the projects open data platform.
Necessary S/W and other tools for enabling re-use	No specific type of software required. Required characteristics include reading capabilities of CSV and WAV files
Repository where data will be stored (institutional, etc., if already existing and identified)	The dataset will be accommodated at the wiki page of the MyAirCoach project, as well as at an Open Data Platform of the final system.
Archiving and preservation (including storage and backup)	
For how long should the	The public part of the dataset will be preserved online for

data be preserved?	as long as there are regular downloads within the online platform of the MyAirCoach system. After that, it would be made accessible by request in order to reduce any issues regarding the overall performance of the system. The private part of the dataset will be preserved by responsible MyAirCoach partner at least until the end of the project.
Approximated end volume of data	Unknown
Indicative associated costs for data archiving and preservation	Probably two dedicated hard disk drives will be allocated for the dataset; one for the public part and one for the private. There are no costs associated with its preservation of the data.
Indicative plan for covering the above costs	Any cost will be covered by the project.
Ethical issues and requirements	
	The collected data should be carefully anonymized for the preservation of the privacy of participants. Sounds measurements should be carefully reviewed and delete any sections were participants speak and reveal important aspects of their way of life or identify them.

3.2 Datasets of Physiology Assessments

Name	Dataset of Physiology Assessments
Naming Prefix	*DS_Physiology*
Summary	The current type of dataset will include different types of physiological measurements collected within the project, such as the measurements of smart health bracelets, etc.. Furthermore, this category will also include the physiological assessments done in the healthcare environment by trained practitioners (especially all assessment done in the project test and evaluation campaigns)
Positioning within the MyAirCoach project	
Relation to the project objective	MyAirCoach is aiming to propose a novel modelling approach for the personalized and the overall understanding of asthma. It is therefore, of crucial importance to collect an adequate amount of data in

	order to define a modelling framework that will effectively cover the most important aspects of the disease. Furthermore, the MyAirCoach project is aiming to develop decision support tools and risk prediction functionalities that will be based on the physiological condition of the patient. In this regards, the collected data will be used for the training and the validation of the algorithmic approaches that will allow such functionalities.
Related Work Packages	**WP2** Test Campaigns, measurements and clinical analysis **WP4** Computational models, intelligent information processing and DSS module **WP6** Evaluation
Description of Dataset Category	
Origin of Data	Patients' physiology assessments can be assessed either manually by the corresponding doctors based on medical examinations or automatically by the myAirCoach system based on the analysis of health data extracted by utilized sensors.
Nature and scale of data	Data will be represented based on the openEHR framework, using the available archetypes when possible or developing new types of archetypes when it is required.
Use by researchers and healthcare professionals	The datasets in this category can support research in the field of medical decision support and can form the basis for the comparative validation of different algorithmic approaches. Furthermore the aggregated data can be used for the validation or comparison of commercial medical decision support tools Finally, the current type of datasets can be used for the development of alternative modelling approaches of asthma disease of be used for the extension of the project outcomes to other respiratory medical issues.
Indicative existing similar dataset	There have not been identified any online available datasets in this category and for any method of sensing.
Indicative scientific publications	Although a variety of scientific publications are available on the study of physiological parameters in regards to asthma, a unified approach for the use of the diverse information of electronic medical records as envisioned

	by the MyAirCoach project has not been identified.
Standards and Metadata	
	The dataset will be accompanied by detailed documentation of its contents along with metadata describing the demographics of the samples from which the data were generated and detailed description of the data collection process. Indicative metadata include: (a) description of the experimental setup and procedure that led to the generation of the dataset, (b) documentation of the variables recorded in the dataset. The metadata will be in a format that may be easily parsed with open source software.
Existing suitable standards	The openEHR open standard specification for health informatics describing the management, storage, retrieval and exchange of health data in electronic health records (EHRs)[8]. OpenEHR is currently identified as the main data representation framework to be followed by MyAirCoach system The HL7 framework (and related standards) for the exchange, integration, sharing, and retrieval of electronic health information[9]
Data Sharing	
Access type	In accordance with the ethical and legal requirements regarding data obtained from human participants, the dataset will be initially available to the Consortium Members and only after its careful anonymization. Personal information regarding the participants will be kept strictly private. As the project progresses and the collected data are used for the research and development processes of the project they will become available at the projects open data platform after the approval by the ethics committee of the MyAirCoach project. The inclusion of a subject's data in the public part of this dataset will be done on the basis of appropriate informed consent to data publication.
Access procedure	In the first stages of the dataset sharing, and as soon it reaches an anonymized formed, it will be shared among the consortium through the wiki page of the project. For the second stage of dataset publication, the

	anonymized data will be published through the open data platform of the project in order to be used by registered users and subsequently by any interested party aiming to use them for research and development.
Embargo periods (if any)	No preset embargo periods. Selection of the appropriate time of publication based on the research and development timeline of the project, the protection of intellectual property and the proper safeguarding of the privacy of participants
Technical mechanisms for dissemination	The public part of the datasets in this category will be accessible through the projects open data platform.
Necessary S/W and other tools for enabling re-use	The data will be only accessible through the use of software components and productsthat support openEHR[8]
Repository where data will be stored (institutional, etc., if already existing and identified)	The dataset will be accommodated at the wiki page of the MyAirCoach project, as well as at an Open Data Platform of the final system.
Archiving and preservation (including storage and backup)	
For how long should the data be preserved?	The public part of the dataset will be preserved online for as long as there are regular downloads within the online platform of the MyAirCoach system. After that, it would be made accessible by request in order to reduce any issues regarding the overall performance of the system. The private part of the dataset will be preserved by responsible MyAirCoach partner at least until the end of the project.
Approximated end volume of data	Unknown
Indicative associated costs for data archiving and preservation	Probably two dedicated hard disk drives will be allocated for the dataset; one for the public part and one for the private. There are no costs associated with its preservation of the data.
Indicative plan for covering the above costs	Small one-time costs covered within the MyAirCoach project.
Ethical issues and requirements	
	The collected data should be carefully anonymized for the preservation of the privacy of participants. All doctors' comments accompanying the assessments should be carefully reviewed and delete any sections that

	can be used to identify the respective patient.

3.3 Datasets of Lifestyle Assessment

Name	Dataset of lifestyle Assessments
Naming Prefix	*DS_Lifestyle*
Summary	The current type of dataset will include different types of assessments and data related to the lifestyle and behavior activity levels of patients as they will be collected within the project during the measurement campaigns and also through the sensing devices of used by the project (i.e. smart health bracelets or smartphones)
Positioning within the MyAirCoach project	
Relation to the project objective	MyAirCoach will try to contribute to the effects of the lifestyle of patients and especially their activity levels on the asthma condition and outline significant correlations that may help doctors to better help their patients and medical researchers to understand the condition of asthma though a mutli-parametric view.
Related Work Packages	**WP2** Test Campaigns, measurements and clinical analysis **WP4** Computational models, intelligent information processing and DSS module **WP6** Evaluation
Description of Dataset Category	
Origin of Data	Patients activity levels can be produced either manually by the corresponding doctors based on specialized questionnaires or automatically by the myAirCoach system based on the analysis of health data extracted by utilized sensors.
Nature and scale of data	The current type of dataset willinclude responses to questionnaires as they will be used in the measurement campaigns or though the final version of the MyAirCoach system. In addition the current category may include measurements of activity as they will be assessed by the sensing devices of the project namely: smart health bracelets and smartphone sensors
Use by researchers and healthcare professionals	The current dataset will help medical researchers to identify correlation between the activity level of patients

	and the risk of asthma exacerbations. Furthermore, the collected data can be used for the validation and comparison of algorithmic approaches studying the activity levels of people through the use of acceleration measurements of modern smart devices.
Indicative existing similar dataset	There have not been identified any online available datasets in this category and for any method of sensing.
Indicative scientific publications	There have not been identified any online available datasets in this category and for any method of sensing.
Standards and Metadata	
	The dataset will be accompanied with detailed documentation of its contents and of all the parameters and selected procedures during the deployment of the campaigns or the characteristics of the sensors used for their assessment through sensing devices.
Existing suitable standards	No existing standards identified
Data Sharing	
Access type	In accordance with the ethical and legal requirements regarding data obtained from human participants, the dataset will be initially available to the Consortium Members and only after its careful anonymization. Personal information regarding the participants will be kept strictly private. As the project progresses and the collected data are used for the research and development processes of the project they will become available at the projects open data platform after the approval by the ethics committee of the MyAirCoach project. The inclusion of a subject's data in the public part of this dataset will be done on the basis of appropriate informed consent to data publication.
Access procedure	In the first stages of the dataset sharing, and as soon it reaches an anonymized formed, it will be shared among the consortium through the wiki page of the project. For the second stage of dataset publication, the anonymized data will be published through the open data platform of the project in order to be used by registered users and subsequently by any interested party aiming to use them for research and development.
Embargo periods (if any)	No preset embargo periods. Selection of the appropriate time of publication based on

	the research and development timeline of the project, the protection of intellectual property and the proper safeguarding of the privacy of participants
Technical mechanisms for dissemination	The public part of the datasets in this category will be accessible through the projects open data platform.
Necessary S/W and other tools for enabling re-use	The data will be only accessible through the use of software components and products that support openEHR[8]
Repository where data will be stored (institutional, etc., if already existing and identified)	The dataset will be accommodated at the wiki page of the MyAirCoach project, as well as at an Open Data Platform of the final system.
Archiving and preservation (including storage and backup)	
For how long should the data be preserved?	The public part of the dataset will be preserved online for as long as there are regular downloads within the online platform of the MyAirCoach system. After that, it would be made accessible by request in order to reduce any issues regarding the overall performance of the system. The private part of the dataset will be preserved by responsible MyAirCoach partner at least until the end of the project.
Approximated end volume of data	Unknown
Indicative associated costs for data archiving and preservation	Probably two dedicated hard disk drives will be allocated for the dataset; one for the public part and one for the private. There are no costs associated with its preservation of the data.
Indicative plan for covering the above costs	Small one-time costs covered within the MyAirCoach project.
Ethical issues and requirements	
	The collected data should be carefully anonymized for the preservation of the privacy of participants. All doctors' comments accompanying the assessments should be carefully reviewed and delete any sections that can be used to identify the respective patient.

3.4 Datasets of Nutritional Assessments

Name	Dataset of Nutritional Assessments
Naming Prefix	*DS_Nutritional*
Summary	The current type of dataset will cover different types of assessments related to the nutritional habits of asthma patients as they will be collected within the project (questionnaires, etc.).
Positioning within the MyAirCoach project	
Relation to the project objective	MyAirCoach will try to contribute to the understanding of the nutritional habits of asthma patients in the evolution of their disease and outline significant correlations that may help doctors to better help their patients and medical researchers to understand the condition of asthma though a mutli-parametric view.
Related Work Packages	**WP2** Test Campaigns, measurements and clinical analysis **WP6** Evaluation
Description of Dataset Category	
Origin of Data	Data collected and conclusions drawn from the measurements campaigns of the project.
Nature and scale of data	The current category of datasets will include mainly anonymized responses to questionnaires as they will be used in the measurement campaigns or assessed through the MyAirCoach final system
Use by researchers and healthcare professionals	The datasets of this category are aiming to become a useful component for the study of asthma condition by medical researchers and hopefully be extended by the input of other projects in the field of asthma related research.
Indicative existing similar dataset	There have not been identified any online available datasets in this category and for any method of sensing.
Indicative scientific publications	There have not been identified any online available datasets in this category and for any method of sensing.
Standards and Metadata	
	The dataset will be accompanied with detailed documentation of its contents and of all the parameters and selected procedures during the deployment of the campaigns

Existing suitable standards	No existing standards identified
Data Sharing	
Access type	In accordance with the ethical and legal requirements regarding data obtained from human participants, the dataset will be initially available to the Consortium Members and only after its careful anonymization. Personal information regarding the participants will be kept strictly private. As the project progresses and the collected data are used for the research and development processes of the project they will become available at the projects open data platform after the approval by the ethics committee of the MyAirCoach project. The inclusion of a subject's data in the public part of this dataset will be done on the basis of appropriate informed consent to data publication.
Access procedure	In the first stages of the dataset sharing, and as soon it reaches an anonymized formed, it will be shared among the consortium through the wiki page of the project. For the second stage of dataset publication, the anonymized data will be published through the open data platform of the project in order to be used by registered users and subsequently by any interested party aiming to use them for research and development.
Embargo periods (if any)	No preset embargo periods. Selection of the appropriate time of publication based on the research and development timeline of the project, the protection of intellectual property and the proper safeguarding of the privacy of participants
Technical mechanisms for dissemination	The public part of the datasets in this category will be accessible through the projects open data platform.
Necessary S/W and other tools for enabling re-use	The data will be only accessible through the use of software components and products that support openEHR[8]
Repository where data will be stored (institutional, etc., if already existing and identified)	The dataset will be accommodated at the wiki page of the MyAirCoach project, as well as at an Open Data Platform of the final system.
Archiving and preservation (including storage and backup)	
For how long should the data be preserved?	The public part of the dataset will be preserved online for as long as there are regular downloads within the online platform of the MyAirCoach system. After that, it would

	be made accessible by request in order to reduce any issues regarding the overall performance of the system. The private part of the dataset will be preserved by responsible MyAirCoach partner at least until the end of the project.
Approximated end volume of data	Unknown
Indicative associated costs for data archiving and preservation	Probably two dedicated hard disk drives will be allocated for the dataset; one for the public part and one for the private. There are no costs associated with its preservation of the data.
Indicative plan for covering the above costs	Small one-time costs covered within the MyAirCoach project.
Ethical issues and requirements	
	The collected data should be carefully anonymized for the preservation of the privacy of participants. All doctors' comments accompanying the assessments should be carefully reviewed and delete any sections that can be used to identify the respective patient.

3.5 Datasets of Exhaled Nitric Oxide Measurements

Name	Dataset of Exhaled Nitric Oxide Measurements
Naming Prefix	*DS_ExhaledNO*
Summary	The current type of dataset will include measurements and data collected in regards to the concentration of Nitric Oxide (NO) in the exhaled breath of patients. In the framework of the MyAirCoach project exhaled NO will be measured by the NIOX device developed by AEROCRINE.
Positioning within the MyAirCoach project	
Relation to the project objective	Measurement of fractional nitric oxide (NO) concentration in exhaled breath (FeNO) is a quantitative, non-invasive, simple, and safe method of measuring airway inflammation that provides a complementary tool to other ways of assessing airways disease, including asthma[10]. There are various devices used for measuring the amount of FeNO in the breath. The National Institute for Health and Care (NICE) has assessed 3 devices including NIOX

	device of AEROCRINE[11] The MyAirCoach project is aiming to analyze the FeNO measurements of patients for the better understanding of their asthma condition, the personalization of medication approaches and the prediction of dangerous exacerbation incidents.
Related Work Packages	**WP2** Test campaigns, measurements, clinical analysis **WP3** Smart sensor based inhaler prototype and WBAN **WP4** Computational models, intelligent information processing and DSS module **WP6** Evaluation
Description of Dataset Category	
Origin of Data	Raw data will be collected by NIOX devices of AEROCRINE
Nature and scale of data	The data of this category will be in the form of time series describing measured parameters during the exhalation of patients CSV (Comma Separated Values) is the advised file format in this category since it allows the easy use of the data both through programming languages and spreadsheet software packages (e.g. Open Office Calc, Microsoft Excel). In this case timestamps for every measurement or the sampling rate should be defined.
Use by researchers and healthcare professionals	The datasets in this category can support research in the field of biomedical signal processing and serve as a basis for the comparative validation of different algorithmic approaches for the analysis of FeNo measurements Furthermore, and if the collected data cover an adequate number of patients with accurately assessed levels of asthma control, the analysis of FeNO measurements can reveal important asthma indicators.
Indicative existing similar dataset	National Health and Nutrition Examination Survey[12]
Indicative scientific publications	Exhaled Nitric Oxide For The Diagnosis Of Asthma In Adults And Children: A Systematic Review[13] Exhaled nitric oxide levels to guide treatment for adults with asthma[14] Exhaled nitric oxide levels to guide treatment for children with asthma[15]
Standards and Metadata	

	The dataset will be accompanied by detailed documentation of its contents along with metadata describing the demographics of the samples from which the data were generated and detailed description of the data collection process. Indicative metadata include: (a) description of the experimental setup and procedure that led to the generation of the dataset, (b) documentation of the variables recorded in the dataset. The metadata will be in a format that may be easily parsed with open source software.
Data Sharing	
Access type	In accordance with the ethical and legal requirements regarding data obtained from human participants, the dataset will be initially available to the Consortium Members and only after its careful anonymization. Personal information regarding the participants will be kept strictly private. As the project progresses and the collected data are used for the research and development processes of the project they will become available at the projects open data platform after the approval by the ethics committee of the MyAirCoach project. The inclusion of a subject's data in the public part of this dataset will be done on the basis of appropriate informed consent to data publication.
Access procedure	In the first stages of the dataset sharing, and as soon it reaches an anonymized formed, it will be shared among the consortium through the wiki page of the project. For the second stage of dataset publication, the anonymized data will be published through the open data platform of the project in order to be used by registered users and subsequently by any interested party aiming to use them for research and development.
Embargo periods (if any)	No preset embargo periods. Selection of the appropriate time of publication based on the research and development timeline of the project, the protection of intellectual property and the proper safeguarding of the privacy of participants
Technical mechanisms for dissemination	The public part of the datasets in this category will be accessible through the projects open data platform.
Necessary S/W and other	No specific type of software required.

tools for enabling re-use	Required characteristics include reading capabilities of CSV
Repository where data will be stored (institutional, etc., if already existing and identified)	The dataset will be accommodated at the wiki page of the MyAirCoach project, as well as at an Open Data Platform of the final system.
Archiving and preservation (including storage and backup)	
For how long should the data be preserved?	The public part of the dataset will be preserved online for as long as there are regular downloads within the online platform of the MyAirCoach system. After that, it would be made accessible by request in order to reduce any issues regarding the overall performance of the system. The private part of the dataset will be preserved by responsible MyAirCoach partner at least until the end of the project.
Approximated end volume of data	Unknown
Indicative associated costs for data archiving and preservation	Probably two dedicated hard disk drives will be allocated for the dataset; one for the public part and one for the private. There are no costs associated with its preservation of the data.
Indicative plan for covering the above costs	Small one-time costs covered within the MyAirCoach project.
Ethical issues and requirements	
	The collected data should be carefully anonymized for the preservation of the privacy of participants.

3.6 Datasets of Environmental Measurements

Name	Datasets of Environmental Measurements
Naming Prefix	*DS_Environmental*
Summary	The current type of datasets will cover the assessment of environment parameters such as air temperature and humidity and also levels of pollution and concentration of common asthma irritants when possible.
Positioning within the MyAirCoach project	
Relation to the project objective	Asthma is a multi-parametric condition that is being affected significantly by the conditions in the environment of patients. In order to corer this usually neglected view of asthma

	disease, MyAirCoach project is aiming to use the collected measurements from the environment of patients in order to outline important indicators of asthma attacks and contribute to the avoidance of such harmful incidents by warning the patients and suggesting mitigation actions.
Related Work Packages	**WP2** Test campaigns, measurements, clinical analysis **WP3** Smart sensor based inhaler prototype and WBAN **WP4** Computational models, intelligent information processing and DSS module **WP6** Evaluation
Description of Dataset Category	
Origin of Data	Raw data will be collected online resources of environmental conditions and sensing components of the MyAirCoach project.
Nature and scale of data	The data of this category will be in the form of time series describing the conditions in the patients environment, or in a specific location. CSV (Comma Separated Values) is the advised file format in this category since it allows the easy use of the data both through programming languages and spreadsheet software packages (e.g. Open Office Calc, Microsoft Excel). In this case timestamps for every measurement or the sampling rate should be defined.
Use by researchers and healthcare professionals	The datasets in this category can support research in the field of biomedical signal processing as they hold the promise to correlate clinical indicators of asthma attacks with environmental parameters.
Indicative existing similar dataset	London Air Quality Network – King's College London[16] Air Quality – The City of London[17] Air quality information and campaigns – Manchester City Council[18] GreatAir Manchester – The air quality website for the Greater Manchester region[19] Weather data for research and projects – University of Reading[20] Historical monthly open data for UK meteorological stations – Met Office[21] UK Humidity open datasets[22]
Indicative scientific	Effect Of Atmospheric Conditions On Asthma Control And

publications	Gene Expression In The Airway[23] Synoptic weather types and aeroallergens modify the effect of air pollution on hospitalizations for asthma hospitalizations in Canadian cities[24]
Standards and Metadata	
	The dataset will be accompanied by detailed documentation of its contents along with metadata describing the demographics of the samples from which the data were generated and detailed description of the data collection process. Indicative metadata include: (a) description of the experimental setup and procedure that led to the generation of the dataset, (b) documentation of the variables recorded in the dataset. The metadata will be in a format that may be easily parsed with open source software.
Data Sharing	
Access type	In accordance with the ethical and legal requirements regarding data obtained from human participants, the dataset will be initially available to the Consortium Members and only after its careful anonymization. Personal information regarding the participants will be kept strictly private. As the project progresses and the collected data are used for the research and development processes of the project they will become available at the projects open data platform after the approval by the ethics committee of the MyAirCoach project. The inclusion of a subject's data in the public part of this dataset will be done on the basis of appropriate informed consent to data publication.
Access procedure	In the first stages of the dataset sharing, and as soon it reaches an anonymized formed, it will be shared among the consortium through the wiki page of the project. For the second stage of dataset publication, the anonymized data will be published through the open data platform of the project in order to be used by registered users and subsequently by any interested party aiming to use them for research and development.
Embargo periods (if any)	No preset embargo periods. Selection of the appropriate time of publication based on the research and development timeline of the project, the protection of intellectual property and the proper

	safeguarding of the privacy of participants
Technical mechanisms for dissemination	The public part of the datasets in this category will be accessible through the projects open data platform.
Necessary S/W and other tools for enabling re-use	No specific type of software required. Required characteristics include reading capabilities of CSV
Repository where data will be stored (institutional, etc., if already existing and identified)	The dataset will be accommodated at the wiki page of the MyAirCoach project, as well as at an Open Data Platform of the final system.
Archiving and preservation (including storage and backup)	
For how long should the data be preserved?	The public part of the dataset will be preserved online for as long as there are regular downloads within the online platform of the MyAirCoach system. After that, it would be made accessible by request in order to reduce any issues regarding the overall performance of the system. The private part of the dataset will be preserved by responsible MyAirCoach partner at least until the end of the project.
Approximated end volume of data	Unknown
Indicative associated costs for data archiving and preservation	Probably two dedicated hard disk drives will be allocated for the dataset; one for the public part and one for the private. There are no costs associated with its preservation of the data.
Indicative plan for covering the above costs	Small one-time costs covered within the MyAirCoach project.
Ethical issues and requirements	
	In the case that the data are related with a patient and not with a specific geographic location, they should be anonymized carefully

3.7 Datasets of Patient Tomography

Name	Datasets of Patient Tomography
Naming Prefix	*DS_Tomography*
Summary	A dataset of patient lung/chest tomographies will be

	collected and utilized within the MyAirCoach project in order to support the understanding and prediction of asthma condition of patients. Images resulting from modalities such as Computed Tomography (CT) will be used to the understanding of important asthma related parameters and will serve as a basis for the simulation of airflows within the lung airways.
Positioning within the MyAirCoach project	
Relation to the project objective	The MyAirCoach project is aiming to utilize Computational Fluid Dynamics and Fluid Particle Tracing for the understanding of the flow of inhaled medication and irritant particles inside the airways of the patient lungs. In this direction the availability or realistic geometric models of human lungs will be of fundamental importance in order to reach realistic results.
Related Work Packages	**WP2** Test Campaigns, measurements and clinical analysis **WP4** Computational models, intelligent information processing and DSS module **WP6** Evaluation
Description of Dataset Category	
Origin of Data	There are three types of patient tomographies used for asthma: Computed Tomography (CT), Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI). **Computed tomography (CT)** scan provides a high degree of anatomical detail and has been used in the diagnosis of various airway diseases. **High resolution computed tomography (HRCT)** is a special type of CT which allows visualization of airways and parenchyma in much greater detail than conventional CT or plain radiography.In asthma it is very useful particularly when it is difficult to achieve an effective control of disease, and in severe deterioration. **Positron Emission Tomography (PET)** can be also used in asthma diagnosis and especially in the assessment of lung inflammation in patients with atopic asthma,.**Chest Magnetic Resonance Imaging (MRI)** is a more safe and non-invasive method providing even higher resolution than the previously mentioned tomography approaches.
Nature and scale of data	Patient tomographies are actually images of patients' lungs or chest and will be in DICOM (Digital Imaging and Communications in Medicine) format providing the capability to share medical images easily and quickly.

Use by researchers and healthcare professionals	The datasets in this category can support research in the field of medical image processingand extraction of lung geometry and can form the basis for the comparative validation of different algorithmic approaches. Furthermore, the current type of datasets can be used for the extraction of significant asthma indicators that are based on the geometry of the lungs, and therefore contribute to the enhancement of modelling approaches and the medical research of asthma.
Indicative existing similar dataset	Open-Access Medical Image Repositories[25] Public Medical Image Databases – Cornell University[26] DICOM sample image sets[27] MRI and CT Data from The Visible Human Project[28] Bone and Joint CT-SCAN Data – International Society of Biomechanics[29] Sample DICOM Data - TRIPOD[30]
Indicative scientific publications	Although a variety of scientific publications are available for the application of novel image processing approaches on tomographic data and the extraction of the geometry of the airways
Standards and Metadata	
Existing suitable standards	The dataset will follow the DICOM standard[31]
Data Sharing	
Access type	In accordance with the ethical and legal requirements regarding data obtained from human participants, the dataset will be initially available to the Consortium Members and only after its careful anonymization. Personal information regarding the participants will be kept strictly private. As the project progresses and the collected data are used for the research and development processes of the project they will become available at the projects open data platform after the approval by the ethics committee of the MyAirCoach project. The inclusion of a subject's data in the public part of this dataset will be done on the basis of appropriate informed consent to data publication.
Access procedure	In the first stages of the dataset sharing, and as soon it reaches an anonymized formed, it will be shared among the consortium through the wiki page of the project. For the second stage of dataset publication, the

	anonymized data will be published through the open data platform of the project in order to be used by registered users and subsequently by any interested party aiming to use them for research and development. Anonymized DICOM images will also considered to be made publicly available through the DICOM Library[32].
Embargo periods (if any)	No preset embargo periods. Selection of the appropriate time of publication based on the research and development timeline of the project, the protection of intellectual property and the proper safeguarding of the privacy of participants
Technical mechanisms for dissemination	The public part of the datasets in this category will be accessible through the projects open data platform.
Necessary S/W and other tools for enabling re-use	The data will be only accessible through the use of software components and products that support openEHR[8]
Repository where data will be stored (institutional, etc., if already existing and identified)	The dataset will be accommodated at the wiki page of the MyAirCoach project, as well as at an Open Data Platform of the final system.
Archiving and preservation (including storage and backup)	
For how long should the data be preserved?	The public part of the dataset will be preserved online for as long as there are regular downloads within the online platform of the MyAirCoach system. After that, it would be made accessible by request in order to reduce any issues regarding the overall performance of the system. The private part of the dataset will be preserved by responsible MyAirCoach partner at least until the end of the project.
Approximated end volume of data	Unknown
Indicative associated costs for data archiving and preservation	Probably two dedicated hard disk drives will be allocated for the dataset; one for the public part and one for the private. There are no costs associated with its preservation of the data.
Indicative plan for covering the above costs	Small one-time costs covered within the MyAirCoach project.
Ethical issues and requirements	
	The collected data should be carefully anonymized for the preservation of the privacy of participants.

	All doctors' comments accompanying the assessments should be carefully reviewed and delete any sections that can be used to identify the respective patient.

3.8 Datasets of MyAirCoach Patient Models

Name	Datasets of MyAirCoach Patient Models
Naming Prefix	*DS_PatientModels*
Summary	The current type of dataset will cover the generalized patient models produced in the project's framework and which will be designed based on the results of measurement campaigns.
Positioning within the MyAirCoach project	
Relation to the project objective	One of the main objectives of MyAirCoach is the development of a personalized and accurate approach for the modelling of asthma condition of patients. Parallel to this goal, generalized patients models will be created so as to help medical researches to study the disease of asthma through computational simulation approaches.
Related Work Packages	**WP2** Test Campaigns, measurements and clinical analysis **WP4** Computational models, intelligent information processing and DSS module **WP6** Evaluation
Description of Dataset Category	
Origin of Data	Generalize models of asthma patients will be created within the MyAirCoach project as they are described in T4.1 "Patient modelling and formal representation", T4.3 "Multiscale computational modeling of airways and respiratory system" and based on the outcomes of WP2 "Test campaigns, measurements and clinical analysis"
Nature and scale of data	The dataset could be in the form of XML-based representations of the parameters involved in the myAirCoach Virtual Models, in OWL or UsiXML. Furthermore the clinical component of the models could be based on the format of electronic health records such as the openEHR framework.
Use by researchers and healthcare professionals	The datasets of this category are aiming to become a useful component for the study of asthma condition by medical researchers on the basis of computational approaches and simulation.

Indicative existing similar dataset	There have not been identified any online available datasets in this category and for any method of sensing.
Indicative scientific publications	There have not been identified any online available datasets in this category and for any method of sensing.
Standards and Metadata	
Existing suitable standards	The dataset will be accompanied with detailed documentation of its contents and of all the variables involved in the myAirCoach Patient Models (according to the D4.1 deliverable).
Data Sharing	
Access type	In accordance with the ethical and legal requirements regarding data obtained from human participants, the dataset will be initially available to the Consortium Members and only after its careful anonymization. Personal information regarding the participants will be kept strictly private. As the project progresses and the collected data are used for the research and development processes of the project they will become available at the projects open data platform after the approval by the ethics committee of the MyAirCoach project. The inclusion of a subject's data in the public part of this dataset will be done on the basis of appropriate informed consent to data publication.
Access procedure	In the first stages of the dataset sharing, and as soon it reaches an anonymized formed, it will be shared among the consortium through the wiki page of the project. For the second stage of dataset publication, the anonymized data will be published through the open data platform of the project in order to be used by registered users and subsequently by any interested party aiming to use them for research and development.
Embargo periods (if any)	No preset embargo periods. Selection of the appropriate time of publication based on the research and development timeline of the project, the protection of intellectual property and the proper safeguarding of the privacy of participants
Technical mechanisms for dissemination	The public part of the datasets in this category will be accessible through the projects open data platform.
Necessary S/W and other tools for enabling re-use	The data will be only accessible through the use of software components and products that support XML

	based data representations
Repository where data will be stored (institutional, etc., if already existing and identified)	The dataset will be accommodated at the wiki page of the MyAirCoach project, as well as at an Open Data Platform of the final system.
Archiving and preservation (including storage and backup)	
For how long should the data be preserved?	The public part of the dataset will be preserved online for as long as there are regular downloads within the online platform of the MyAirCoach system. After that, it would be made accessible by request in order to reduce any issues regarding the overall performance of the system. The private part of the dataset will be preserved by responsible MyAirCoach partner at least until the end of the project.
Approximated end volume of data	Unknown
Indicative associated costs for data archiving and preservation	Probably two dedicated hard disk drives will be allocated for the dataset; one for the public part and one for the private. There are no costs associated with its preservation of the data.
Indicative plan for covering the above costs	Small one-time costs covered within the MyAirCoach project.
Ethical issues and requirements	
	The collected data should be carefully anonymized for the preservation of the privacy of participants. All doctors' comments accompanying the assessments should be carefully reviewed and delete any sections that can be used to identify the respective patient.

3.9 Dataset of Educational and Training Content

Name	Datasets of Educational and Training Content
Naming Prefix	*DS_EducationAndTraining*
Summary	Material related to the education of patients regarding asthma disease its pathophysiology, symptoms, risk factors and indicators Material related to the training of patients regarding the proper use of different types of inhalers.

Positioning within the MyAirCoach project	
Relation to the project objective	A very important parameter for increased involvement of asthma patients in the management of their disease is their understanding of its fundamental nature and the ability to detect and interpret correctly symptoms of reduce control. Furthermore, the efficient training of patients regarding the proper use of their inhaler is expected to increase their adherence to the prescribed medication and help them optimize their inhaler technique.
Related Work Packages	**WP1 User Needs, system requirements , architecture** **WP2** Test Campaigns, measurements and clinical analysis **WP6** Evaluation
Description of Dataset Category	
Origin of Data	A dataset of educational and training content will be generated during the myAirCoach project lifecycle in order to support patients and clinicians in better asthma management. Registered users of the myAirCoach will also have the capability to upload similar content following the established template.
Nature and scale of data	Educational content will include information about the asthma disease, such as associated risks, allergens, physiology etc. Training content will include multimedia data concerning the proper management and treatment of the disease (e.g. proper use of the inhaler). Data can be in the form of documents, pdf files, videos, images, presentations etc.
Use by researchers and healthcare professionals	The material concentrated under the current category will be useful for patients, doctors, clinicians, Institutes of Health, as well as for researchers investigating issues related to asthma so as to help their patients to effectively manage asthma disease and correctly use their medication.
Indicative existing similar dataset	Asthma Handouts – Sutter Health[33] Asthma Education Materials – Neighborhood Health Plan[34] Instructions for Inhaler and Spacer Use[35] Inhalation protocols[36]
Indicative scientific publications	There have not been identified any online available datasets in this category and for any method of sensing.

Standards and Metadata	
Existing suitable standards	The dataset will be accompanied with detailed documentation of its contents. Existing common formats for documents, pdf files, videos, images and presentations will be utilized (e.g. pdf, doc, png).
Data Sharing	
Access type	Widely open to the entire asthma community
Access procedure	Open access within the MyAirCoach website and the open data platform of the MyAirCoach System
Embargo periods (if any)	No preset embargo periods. Selection of the appropriate time of publication based on the research and development timeline of the project, the protection of intellectual property and the proper safeguarding of the privacy of participants
Technical mechanisms for dissemination	The public part of the datasets in this category will be accessible through the projects open data platform.
Necessary S/W and other tools for enabling re-use	The dataset will be designed to allow easy reuse with commonly available tools and software libraries (e.g. Microsoft Office, Open Office, Adobe Reader, …)
Repository where data will be stored (institutional, etc., if already existing and identified)	The dataset will be accommodated at the project's website and wiki, as well as at an Open Data Platform of the final system.
Archiving and preservation (including storage and backup)	
For how long should the data be preserved?	The public part of the dataset will be preserved online for as long as there are regular downloads within the online platform of the MyAirCoach system. After that, it would be made accessible by request in order to reduce any issues regarding the overall performance of the system. The private part of the dataset will be preserved by responsible MyAirCoach partner at least until the end of the project.
Approximated end volume of data	Unknown
Indicative associated costs for data archiving and preservation	Probably two dedicated hard disk drives will be allocated for the dataset; one for the public part and one for the private. There are no costs associated with its preservation of the data.
Indicative plan for covering the above costs	Small one-time costs covered within the MyAirCoach project.

Ethical issues and requirements	
	The collected data should be carefully anonymized for the preservation of the privacy of participants. All doctors' comments accompanying the assessments should be carefully reviewed and delete any sections that can be used to identify the respective patient.

3.10 Dataset of Asthma Action Plans

Name	Datasets of Asthma Action Plans
Naming Prefix	*DS_ActionPlans*
Summary	This dataset will include templates of action plans and will be used not only for the design and development of the related electronically enhanced action plans of MyAirCoach but also serve as a repository for practitioners to use in their clinical practice.
Positioning within the MyAirCoach project	
Relation to the project objective	Action plans are the main tool for the definition of the methodology that a patient should follow for the effective management of his/her asthma disease. The asthma action plan shows patient's daily treatment, such as what kind of medicines to take and when to take them. It also describes how to control asthma long term and how to handle worsening asthma, or attacks. Moreover, the plan explains when to call the doctor or go to the emergency room. Asthma action plan are actually documents. Traditionally, provided in paper form action plans are based on a variety of templates related to the choice of the doctors towards their easy understanding by patients.
Related Work Packages	**WP2** Test Campaigns, measurements and clinical analysis **WP4** Computational models, intelligent information processing and DSS module **WP6** Evaluation
Description of Dataset Category	
Origin of Data	Templates of action plans will be collected during the measurement campaigns of the project and also from online resources towards the formation of a unified repository that will cover different medication approaches and also different languages.

Nature and scale of data	Electronic documents of action plans or detailed description of interactive electronically enhanced approaches(doc/docx or pdf files)
Use by researchers and healthcare professionals	The current dataset can be used by healthcare professionals in order to review a spectrum of action plan templates and provide their prescribed medication regiment using the most fitted template to the needs of the specific patient.
Indicative existing similar dataset	There have not been identified any online available datasets in this category and for any method of sensing.
Indicative scientific publications	There have not been identified any online available datasets in this category and for any method of sensing.
Standards and Metadata	
Existing suitable standards	There is no widely accepted template for asthma action plans. In this regard the MyAirCoach project is aiming to document the available approaches and provide a detailed review comparing their strengths and weaknesses. Although this review will serve as the guideline for the design of the related MyAirCoach components, it is also expected to help healthcare professionals in their daily practice.
Data Sharing	
Access type	Widely open to the entire asthma community
Access procedure	Open access within the MyAirCoach website and the open data platform of the MyAirCoach System
Embargo periods (if any)	No preset embargo periods. Selection of the appropriate time of publication based on the research and development timeline of the project, the protection of intellectual property and the proper safeguarding of the privacy of participants
Technical mechanisms for dissemination	The public part of the datasets in this category will be accessible through the projects open data platform.
Necessary S/W and other tools for enabling re-use	The dataset will be designed to allow easy reuse with commonly available tools and software libraries (e.g. Microsoft Office, Open Office, Adobe Reader, etc.)
Repository where data will be stored (institutional, etc., if already existing and identified)	The dataset will be accommodated at the project's website and wiki, as well as at an Open Data Platform of the final system.
Archiving and preservation (including storage and backup)	

For how long should the data be preserved?	The public part of the dataset will be preserved online for as long as there are regular downloads within the online platform of the MyAirCoach system. After that, it would be made accessible by request in order to reduce any issues regarding the overall performance of the system. The private part of the dataset will be preserved by responsible MyAirCoach partner at least until the end of the project.
Approximated end volume of data	Unknown
Indicative associated costs for data archiving and preservation	Probably two dedicated hard disk drives will be allocated for the dataset; one for the public part and one for the private. There are no costs associated with its preservation of the data.
Indicative plan for covering the above costs	Small one-time costs covered within the MyAirCoach project.
Ethical issues and requirements	
	The collected data should be carefully anonymized for the preservation of the privacy of participants. All doctors' comments accompanying the assessments should be carefully reviewed and delete any sections that can be used to identify the respective patient.

3.11 Datasets of Collected User Requirements

Name	Datasets of MyAIrCoach Measurement Campaigns
Naming Prefix	*DS_UserRequirements*
Summary	The design and implementation of the MyAirCoach system will be based on the collection and the analysis of user requirements so as to increase the usability and usefulness of the final system. The collected requirements, user inputs and analysis results can be a valuable asset for the development of devices and software systems supportingthe self-management of asthma.
Positioning within the MyAirCoach project	
Relation to the project objective	The development of the MyAirCoach system will be based on a User Centered Approach that has begun with the initial collection of user requirements and will continue

	throughout project.
Related Work Packages	Related to the entire project
Description of Dataset Category	
Origin of Data	Data collected and conclusions drawn from the User Centered Design approach of the project.
Nature and scale of data	The current category may include all previously defined types of datasets of user feedback as they will be assessed during the UCD processes defined in D1.2 "User Requirements, use cases, UCD methodology and final protocols for evaluation studies"
Use by researchers and healthcare professionals	The datasets of this category are aiming to become a useful component for the development of asthma oriented self-management software tools and devices
Indicative existing similar dataset	There have not been identified any online available datasets in this category and for any method of sensing.
Indicative scientific publications	There have not been identified any online available datasets in this category and for any method of sensing.
Standards and Metadata	
Existing suitable standards	The dataset will be accompanied with detailed documentation of its contents and of all the parameters and selected procedures during the deployment of user-feedback collection sessions.
Data Sharing	
Access type	In accordance with the ethical and legal requirements regarding data obtained from human participants, the dataset will be initially available to the Consortium Members and only after its careful anonymization. Personal information regarding the participants will be kept strictly private. As the project progresses and the collected data are used for the research and development processes of the project they will become available at the projects open data platform after the approval by the ethics committee of the MyAirCoach project. The inclusion of a subject's data in the public part of this dataset will be done on the basis of appropriate informed consent to data publication
Access procedure	In the first stages of the dataset sharing, and as soon it reaches an anonymized formed, it will be shared among the consortium through the wiki page of the project.

	For the second stage of dataset publication, the anonymized data will be published through the open data platform of the project in order to be used by registered users and subsequently by any interested party aiming to use them for research and development.
Embargo periods (if any)	No preset embargo periods. Selection of the appropriate time of publication based on the research and development timeline of the project, the protection of intellectual property and the proper safeguarding of the privacy of participants.
Technical mechanisms for dissemination	The public part of the datasets in this category will be accessible through the projects open data platform.
Necessary S/W and other tools for enabling re-use	Dependent on the dataset as it will be defined during the deployment of measurement campaigns and the practice of the responsible clinical partner.
Repository where data will be stored (institutional, etc., if already existing and identified)	The dataset will be accommodated at the wiki page of the MyAirCoach project, as well as at an Open Data Platform of the final system.
Archiving and preservation (including storage and backup)	
For how long should the data be preserved?	The public part of the dataset will be preserved online for as long as there are regular downloads within the online platform of the MyAirCoach system. After that, it would be made accessible by request in order to reduce any issues regarding the overall performance of the system. The private part of the dataset will be preserved by responsible MyAirCoach partner at least until the end of the project.
Approximated end volume of data	Unknown
Indicative associated costs for data archiving and preservation	Probably two dedicated hard disk drives will be allocated for the dataset; one for the public part and one for the private. There are no costs associated with its preservation of the data.
Indicative plan for covering the above costs	Small one-time costs covered within the MyAirCoach project.
Ethical issues and requirements	
	The collected data should be carefully anonymized for the preservation of the privacy of participants. All doctors' comments accompanying the assessments should be carefully reviewed and delete any sections that

	can be used to identify the respective patient.

3.12 Datasets of MyAirCoach Measurement Campaigns

s

Name	Datasets of MyAirCoach Measurement Campaigns
Naming Prefix	*DS_MeasurementCampaigns*
Summary	In the context of the project, two measurement campaigns are scheduled for the initial clinical analysis of asthma condition and the evaluation and optimization of the integrated MyAirCoach system. Three different pilot sites in Europe (London, Manchester, Leiden) will participate in these processes and help for the collection of important data and conclusions regarding asthma management and the related parts of the healthcare system. The current collection of datasets is intended to collects the produced results in a common reference framework and help for the easy access and future reference.
Positioning within the MyAirCoach project	
Relation to the project objective	The measurement campaigns of the MyAirCoach project will form the information basis for the design and development of the majority of envisioned system components and also for the validation of the overall usefulness of the final integrated version of MyAirCoach.
Related Work Packages	**WP2** Test Campaigns, measurements and clinical analysis **WP6** Evaluation
Description of Dataset Category	
Origin of Data	Data collected and conclusions drawn from the measurements campaigns of the project.
Nature and scale of data	The current category may include all previously defined types of datasets in addition to documents or any other types of data collected by the clinical partners in during the campaigns.
Use by researchers and healthcare professionals	The datasets of this category are aiming to become a useful component for the study of asthma condition by medical researchers and hopefully be extended by the input of other projects in the field of asthma related research.
Indicative existing similar	There have not been identified any online available

dataset	datasets in this category and for any method of sensing.
Indicative scientific publications	There have not been identified any online available datasets in this category and for any method of sensing.
Standards and Metadata	
Existing suitable standards	The dataset will be accompanied with detailed documentation of its contents and of all the parameters and selected procedures during the deployment of the campaigns
Data Sharing	
Access type	In accordance with the ethical and legal requirements regarding data obtained from human participants, the dataset will be initially available to the Consortium Members and only after its careful anonymization. Personal information regarding the participants will be kept strictly private. As the project progresses and the collected data are used for the research and development processes of the project they will become available at the projects open data platform after the approval by the ethics committee of the MyAirCoach project. The inclusion of a subject's data in the public part of this dataset will be done on the basis of appropriate informed consent to data publication.
Access procedure	In the first stages of the dataset sharing, and as soon it reaches an anonymized formed, it will be shared among the consortium through the wiki page of the project. For the second stage of dataset publication, the anonymized data will be published through the open data platform of the project in order to be used by registered users and subsequently by any interested party aiming to use them for research and development.
Embargo periods (if any)	No preset embargo periods. Selection of the appropriate time of publication based on the research and development timeline of the project, the protection of intellectual property and the proper safeguarding of the privacy of participants
Technical mechanisms for dissemination	The public part of the datasets in this category will be accessible through the projects open data platform.
Necessary S/W and other tools for enabling re-use	Dependent on the dataset as it will be defined during the deployment of measurement campaigns and the practice of the responsible clinical partner.

Repository where data will be stored (institutional, etc., if already existing and identified)	The dataset will be accommodated at the wiki page of the MyAirCoach project, as well as at an Open Data Platform of the final system.
Archiving and preservation (including storage and backup)	
For how long should the data be preserved?	The public part of the dataset will be preserved online for as long as there are regular downloads within the online platform of the MyAirCoach system. After that, it would be made accessible by request in order to reduce any issues regarding the overall performance of the system. The private part of the dataset will be preserved by responsible MyAirCoach partner at least until the end of the project.
Approximated end volume of data	Unknown
Indicative associated costs for data archiving and preservation	Probably two dedicated hard disk drives will be allocated for the dataset; one for the public part and one for the private. There are no costs associated with its preservation of the data.
Indicative plan for covering the above costs	Small one-time costs covered within the MyAirCoach project.
Ethical issues and requirements	
	The collected data should be carefully anonymized for the preservation of the privacy of participants. All doctors' comments accompanying the assessments should be carefully reviewed and delete any sections that can be used to identify the respective patient.

4 MyAirCoach Open Access Platform

In order to provide the required framework for the sharing of information generated by the MyAirCoach project a relevant knowledge portal of the project was implemented by CERTH were all partners may use for uploading and sharing documents and data produced during the project duration. After the assurance of anonymity and the protection of the privacy of involved users (patients, healthcare professionals) the data produced and collected can be published through the dissemination channels of the project and mainly through the project's website.

Finally, the open access to the MyAirCoach data will continue to being available even after the completion of the project timeline as an independent open access framework for the data of the project.

4.1 MyAirCoach Open Access Data Management Demonstrator

The open access platform of MyAirCoach has been designed and implemented as an appropriate web component with the support of two fundamental accessing views. The first view (members area) is addressed to registered members of the system such as health care professionals/researchers who in addition to the data collected of their patients they can have access to anonymized health records and all the knowledge generated within the MyAirCoach project. Furthermore, these users will be able to upload their data to the open access framework and share them with the entire asthma and research community.

The second view of the system is intended for unregistered users who need to get access to the datasets and publications of MyAirCoach without the need of registering to the system as a user. In this case only anonymized data will be made available to them and they will not be able to upload any type of data to the system.

Figure 1 illustrated the login page of the MyAirCoach platform showing the two different ways of accessing the data of the myAircoach project.

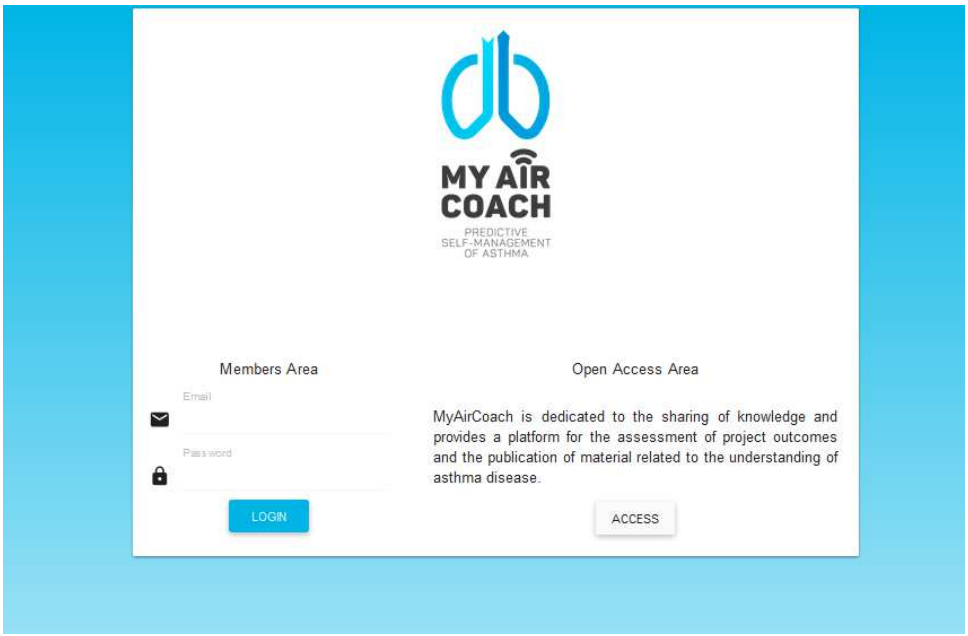

Figure 1: Login page of the MyAirCoach data management Platform

After their login to the platform, the users can be provided with the supported functionalities of the system. The selection of the open data option (in the left panel of the page) should lead to an introduction page which is describing the aim and the main purpose of the myAirCoach open data repository as well as with some instructions for those who are interested in the study and the understanding of asthma disease.

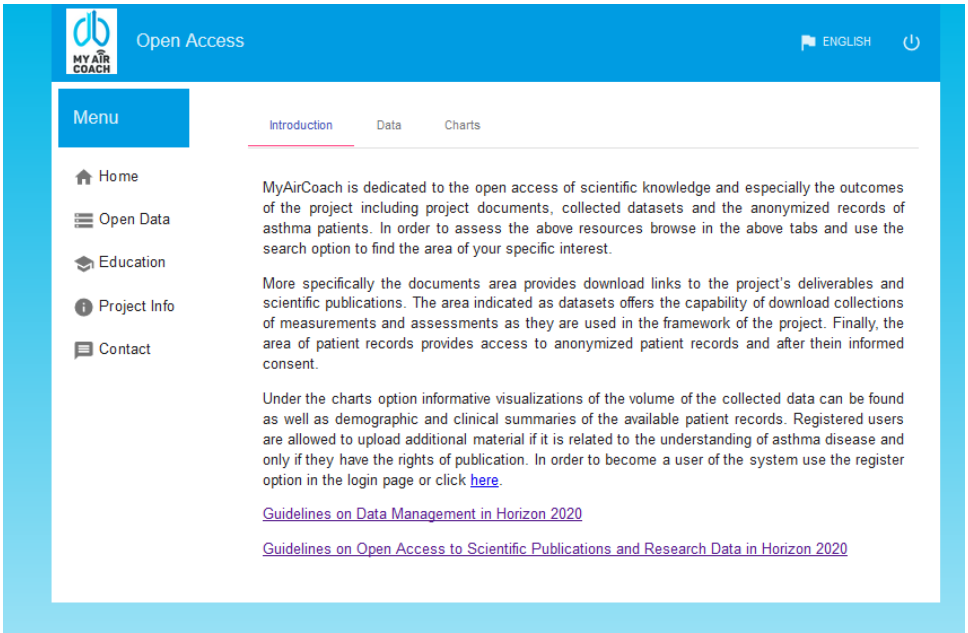

Figure 2: Home page of the MyAirCoach Open Data management platform functionalities

The selection of the data option functionality in the top menu of the web page can provide to the user access to the main part of the open data repository where he/she can access the stored documents, datasets and the anonymized patient records.

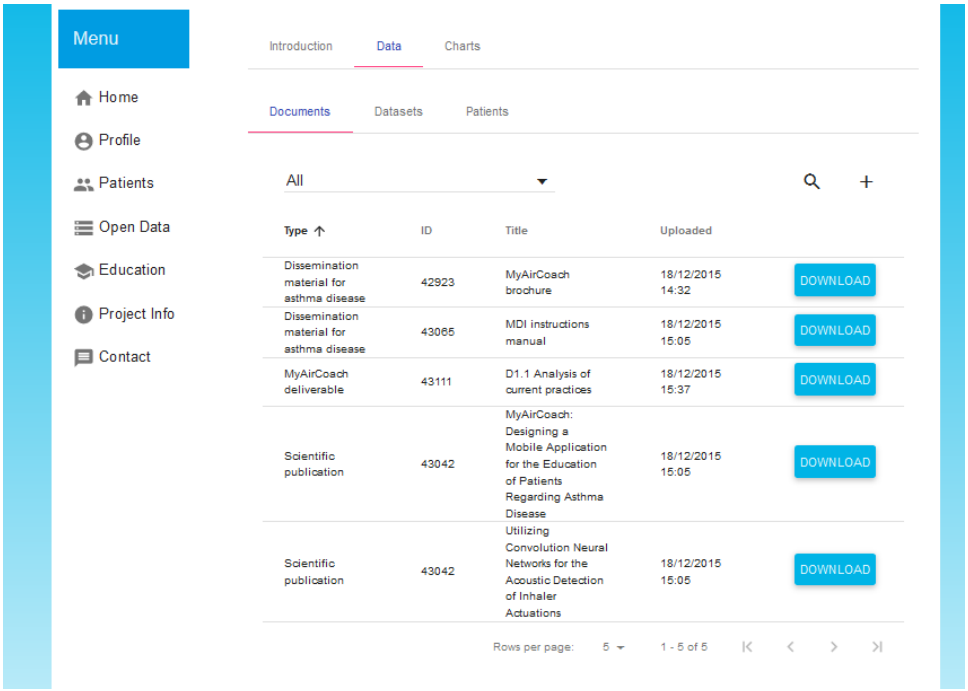

Figure 3: Documents repository of the MyAirCoach platform

The documents repository of the platform can provide to the interested users, access to the following outcomes of the project:

- **MyAirCoach public deliverables** as they will be produced throughout the project and summarize the important results and strategies selected.

- **Scientific publications (from the myAircoach consortium but also from external sources)** as they will translate the results of the project to scientific knowledge to be used by medical researchers and information technology specialists.
- **Appropriate Dissemination material for the asthma disease,** including educational and health content, etc..

In order to support the usability, usefulness and accessibility of the data a metadata template was adopted for the description of every uploaded document as presented in thefollowing Figure 4. It should be underlined that only the creators/owners of the document and the system administrator may have the right to edit/modify any of the stored information to the repository.

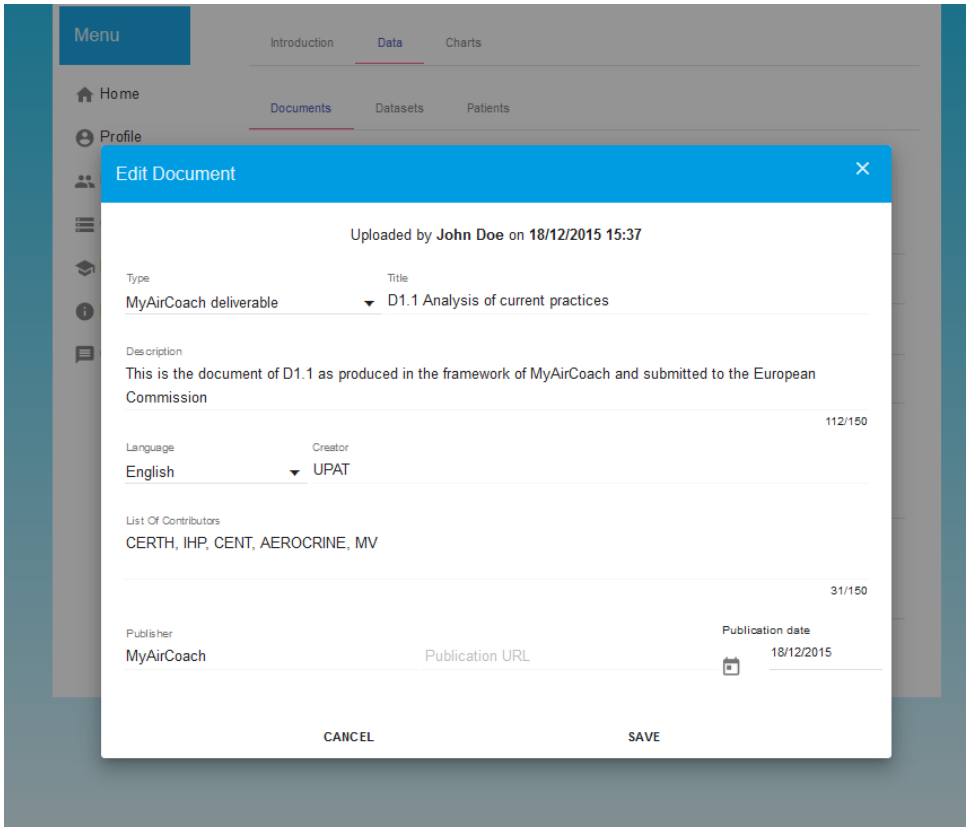

Figure 4: Indicative example of the adopted document metadata template

Furthermore, and following the same approach, the platform can provide to registered users the ability to upload a new document on the platform with the explicit requirement of filling in the most important parameters of document description as shown in the Figure 5.

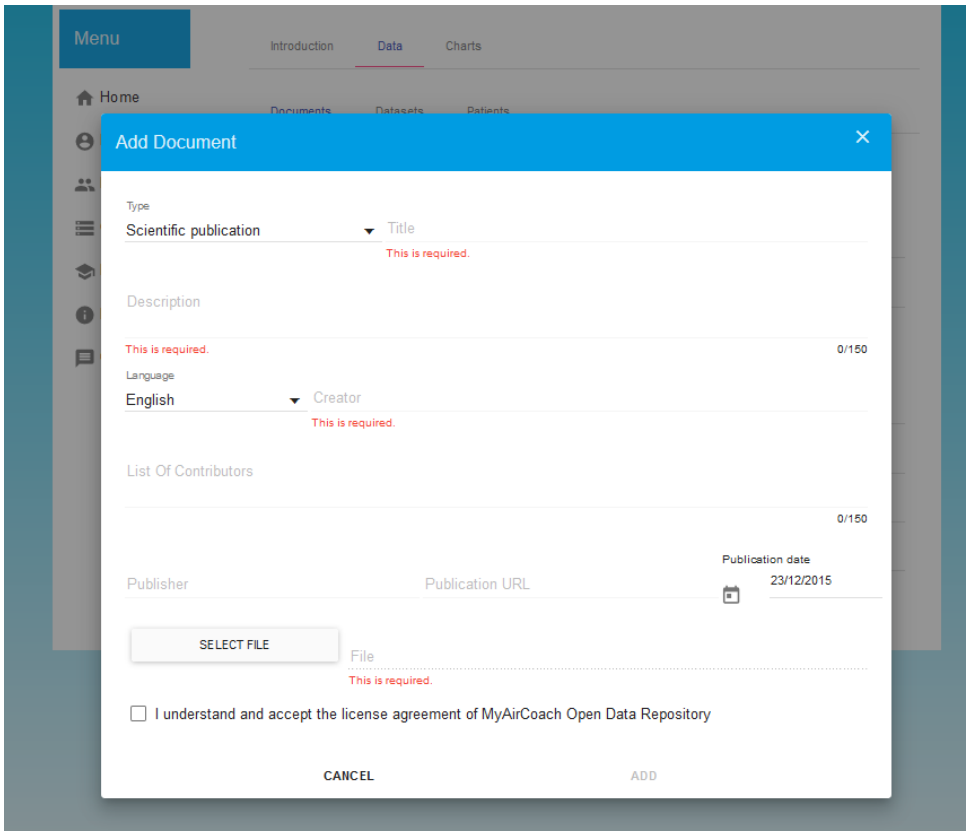

Figure 5: Template for the uploading of a new document on the MyAirCoach system

The following figures describe the same functionalities as above but for the case of the additional datasets that can be supported from the MyAirCoach open access platform. More specifically, the currently available categories of datasets include:

- **Inhaler usage measurements** as they relate to the measurements during the actual use of inhalers by patients.
- **Physiology measurements** as they relate to the physiological assessments of healthcare professionals or measurements of physiological parameters through the use of sensing devices in the patients environment.
- **Exhaled NiOX measurements** as they relate to the use of modern Forced Exhaled Nitric Oxide devices in the clinical environment or in the patients home environment.
- **Nutritional assessments** as they relate to the collection of data related to the nutritional habits of patients or the guidelines of doctors.
- **Lifestyle measurements** as they relate to the collection of data from questionnaires and sensing devices regarding the activity levels of patients and also the advice of healthcare professionals in this area.
- **Environmental measurements** as they relate to the collection of information regarding environmental conditions and pollution levers in the vicinity of asthma patients.

- **Patient tomography data**as they relate to the 3D imaging of patient lungs and respiratory tract.
- **Patient models**as they are related to the modeling framework of MyAirCoach and the general and anonymized patient models produced within the project's framework.
- **Educational and training content** documents and interactive material aiming to educate patients regarding the condition of asthma and help them use their inhalers correctly.
- **Asthma action plans** action plan templated in document form or interactive computer/smartphone based approaches for the description of the prescribed methodology for the effective self-management of asthma.

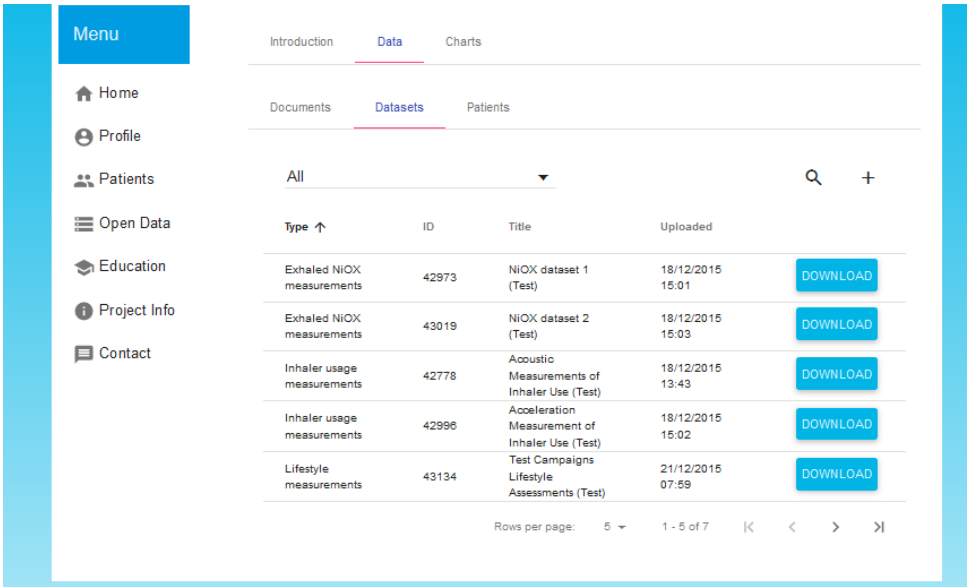

Figure 6: Dataset repository of the MyAirCoach platform

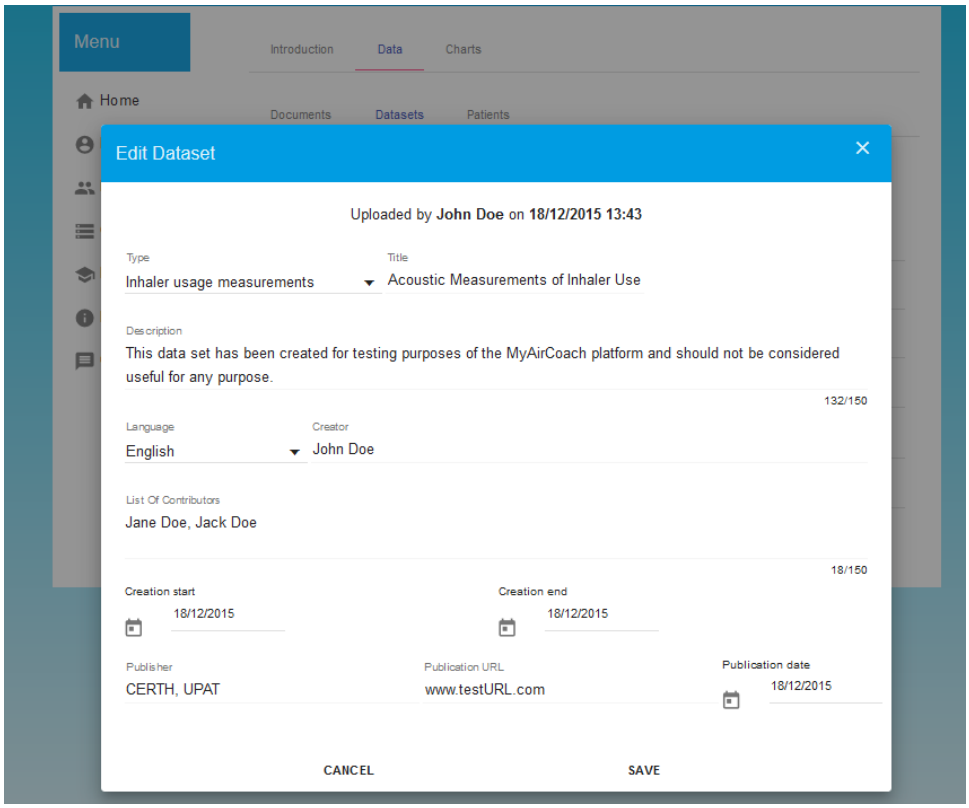

Figure 7: Indicative Example of Dataset Metadata

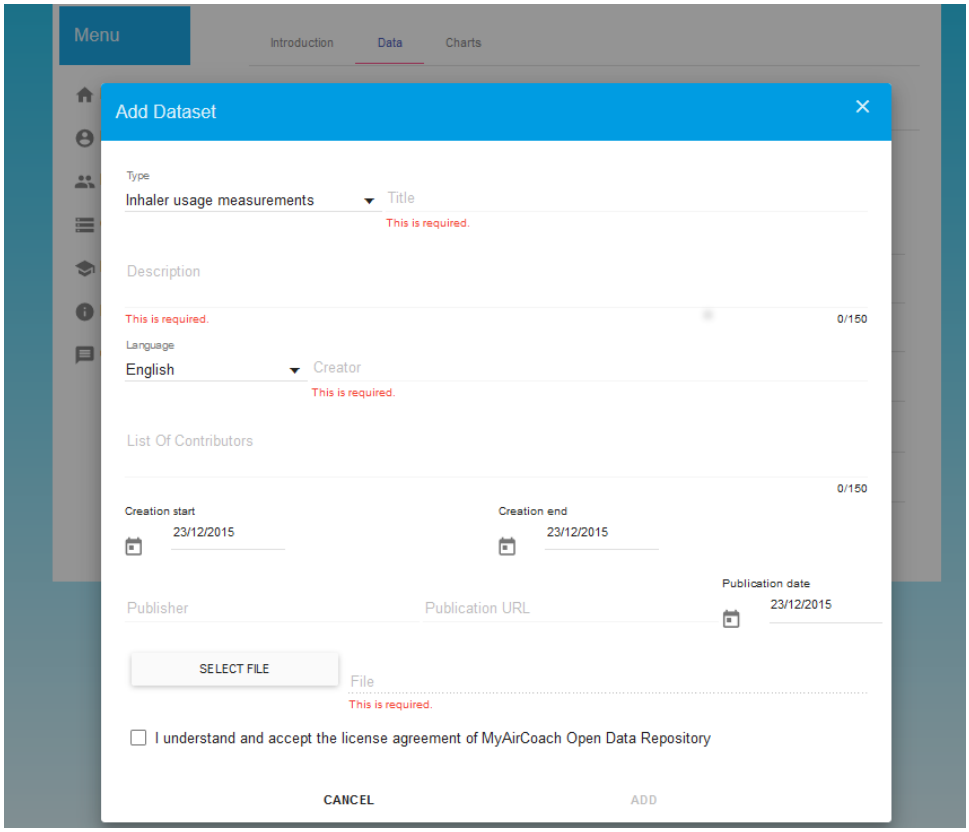

Figure 8: Template for the uploading of datasets on the MyAirCoach platform

Finally, the MyAirCoach open data repository of can give access to anonymised records of patients (after they have given their informed consent). The patients data can be accessed directly through the platform and/or to be downloaded in a standardized data format such as openEHR or HL7. The following Figure presents is presented an example of a list of test patient records created for demonstration the purposes.

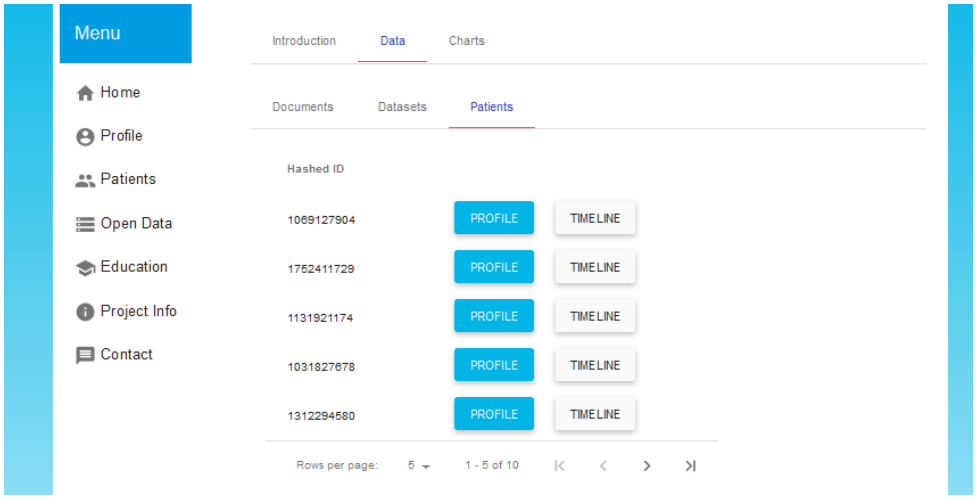

Figure 9:MyAirCoach repository of anonymised patient records

As seen in the following Figure 9, the system can provide to the users two alternative options of accessing patients's data: the profile view and the timeline view. Thus, on the one hand, the profile selection view of the patients electronic health record separated in tabs of different health assessments (Figure 10) while on the other hand, the timeline view sorts the assessments based on their time and is aiming to allow doctors to better understand the overall evolution of the patients health (Figure 11).

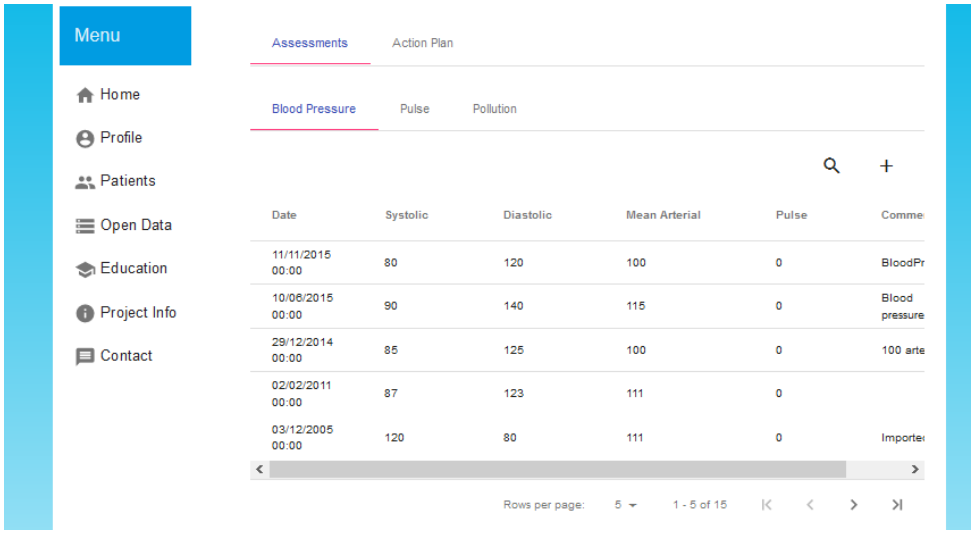

Figure 10: Profile view of the patient's record

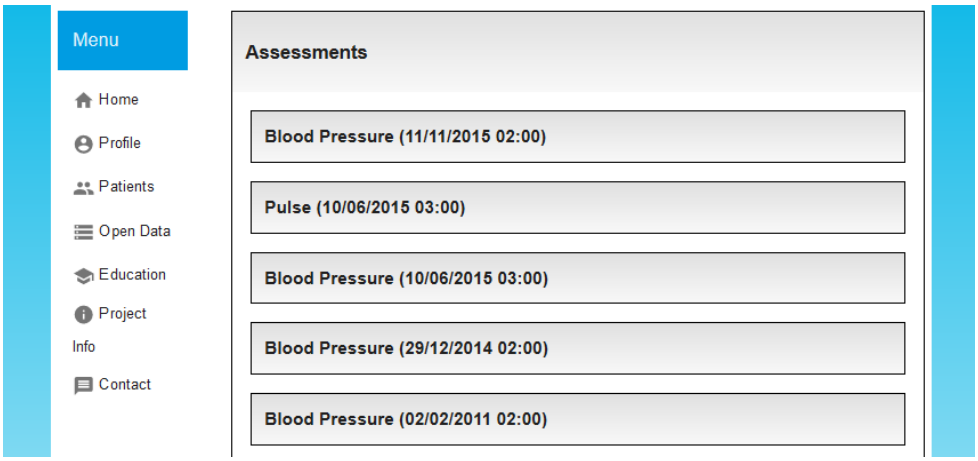

Figure 11: Timeline view of patient record

In order to better explain the anonymisation of patient records in the open data repository Figure 12 shows the patient record as it will be viewed by the responsible doctor. In this case and in addition to the presentation of assessments the name and demographic data of the patient are also available for the doctor to use.

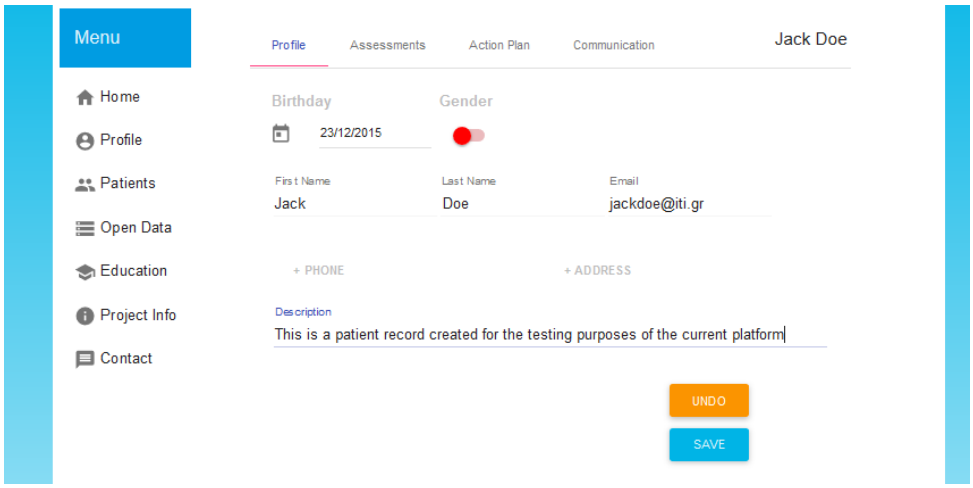

Figure 12: Non anonymised view of a test patient record

The third and final selection on the top menu of the open data platform is used in order to visualise important parameters of the datasets collected and help to understand how the MyAirCoach repository will be evolving through the timeline of the project.

More specifically the document and dataset charts include pie charts for the visualisation of the relative percentage for the defined types of documents or datasets and the number of datasets uploaded as a function of time. Figure 13 and Figure 14 show indicative examples of these visualisations based on the testing data and the evaluation of the platform before the integration with the MyAirCoach system.

Furthermore, informative diagrams are also available as a summarisation of the available anonymised patient records as can be seen in Figure 15. As presented the initial version of the supported charts include the distribution of demographic data among the entire dataset (age and Gender) as well as the distribution of important clinical parameters as they are assessed in the last exam of the patient.

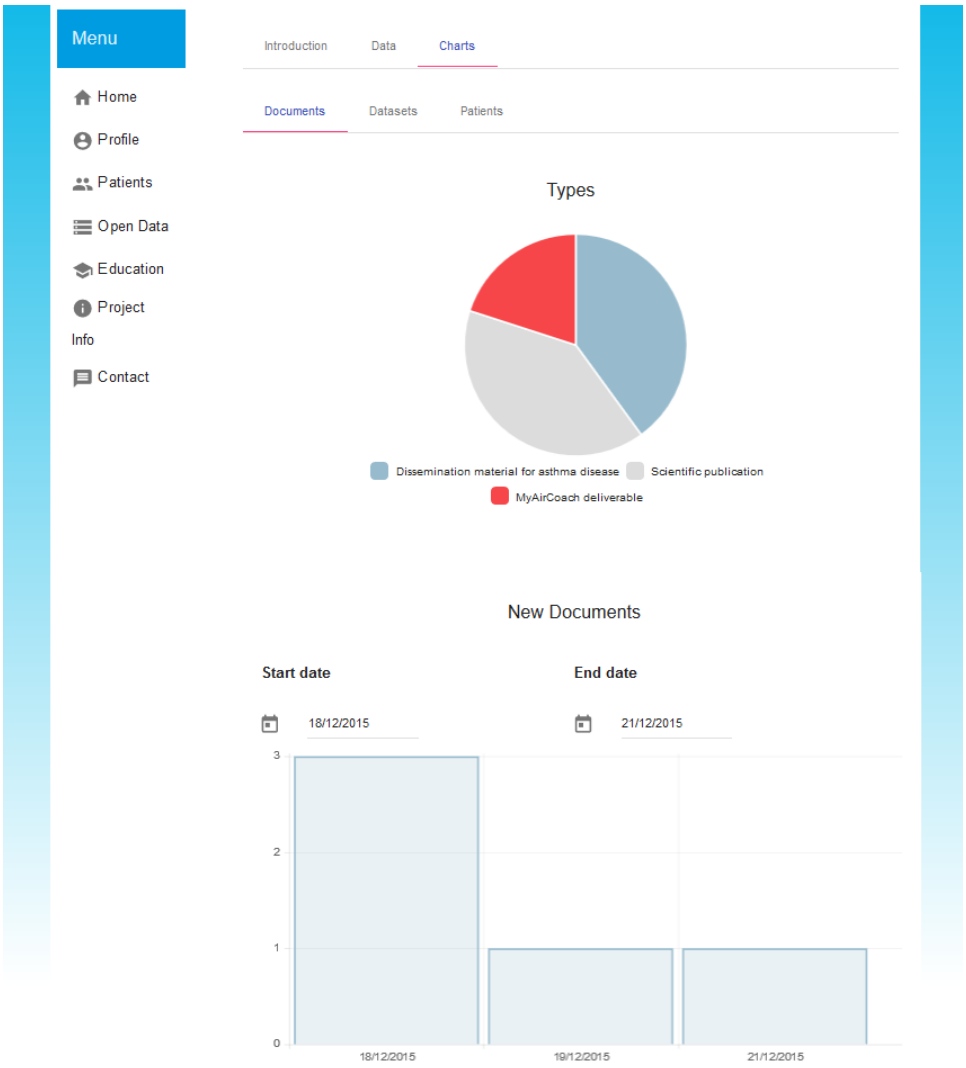

Figure 13: Charts for the visualization of uploaded documents

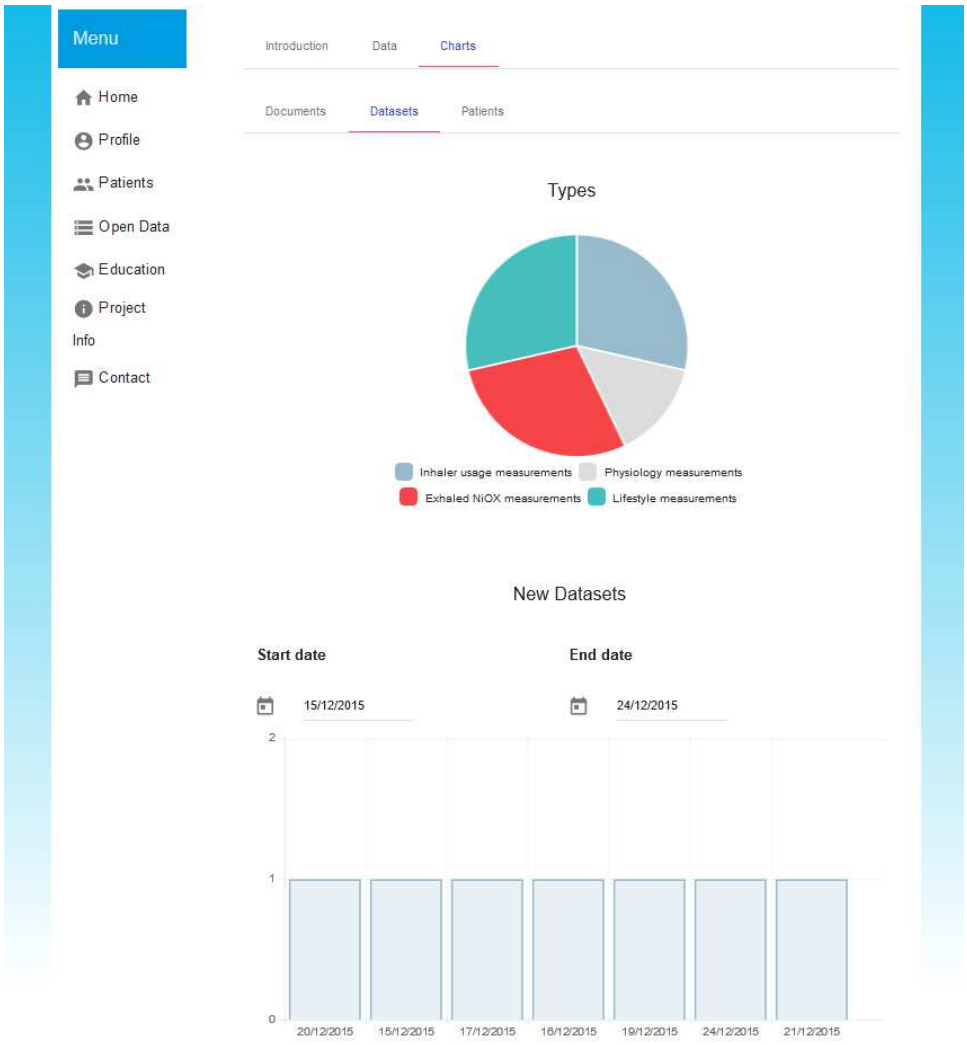

Figure 14: Charts for the visualization of uploaded datasets

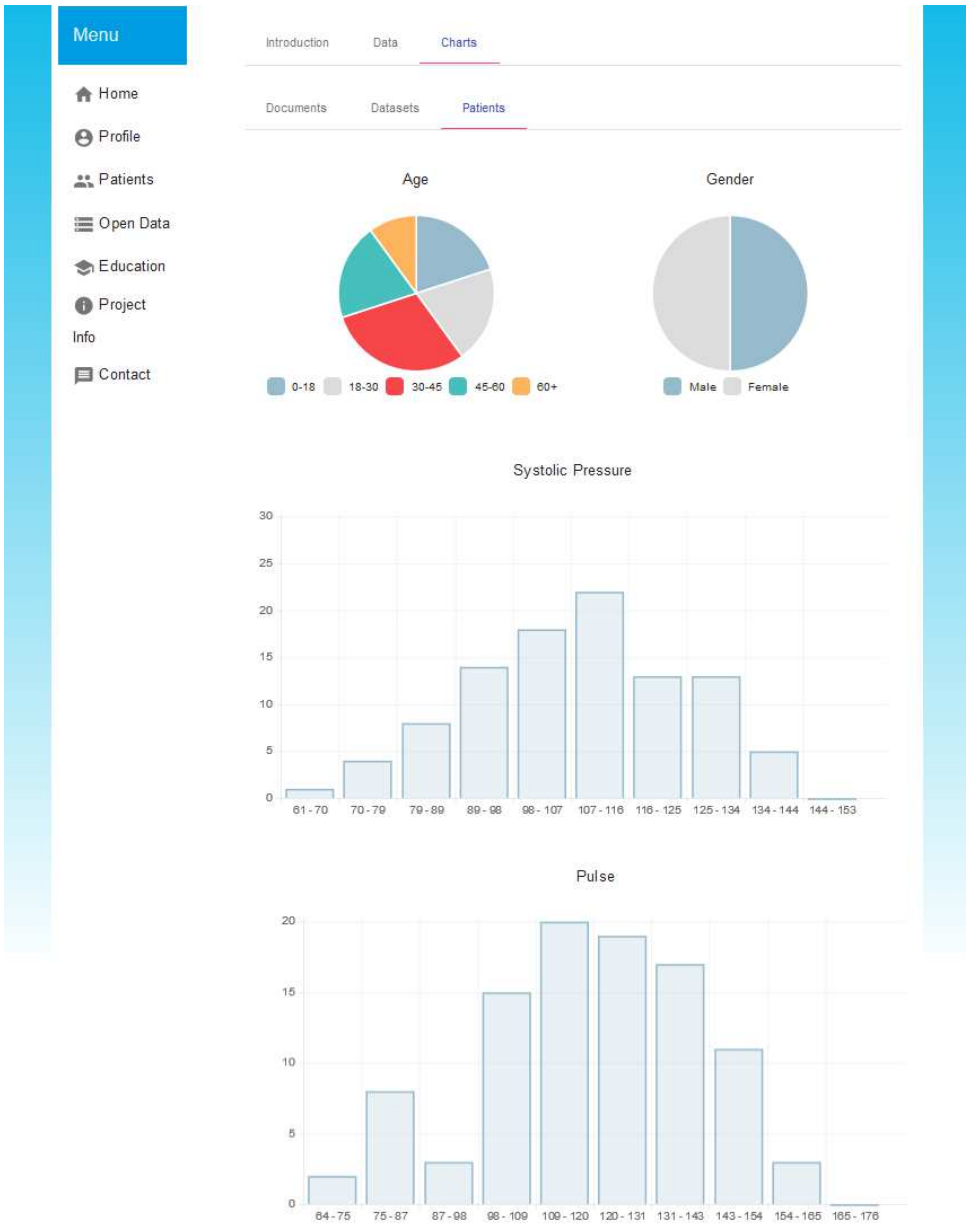

Figure 15: Charts for the visualization of available anonymised patient records

4.2 Conformance to EU Commission Guidelines

The following table summarizes the proposed solutions of MyAirCoach for the addressing of the data management aspects as described by EU commission.

Table 8: Conformance with the EU Commission Data Management Plan Guidelines

Aspect	MyAirCoach Solution
Discoverable	The documents and datasets of the project will be made available through a diverse and side number of dissemination channels in order to support their discoverability. Furthermore, all scientific publications of the project will provide links to the respective datasets on the online open data platform of MyAirCoach
Accessible	The knowledge created within MyAirCoach, both in terms of documents and datasets, will be easily accessible from the website of the project and the open data repository as demonstrated in the previous section
Assessable and intelligible	The metadata provided for its document and dataset uploaded on the MyAirCoach platform together with the provided searching tool will allow their easy access and understanding so as to be used by researchers and be subjected to scientific review.
Usable beyond the original purpose for which it was collected	The inclusion of a diverse set of datasets and documents in the same platform is expected to increase the visibility of the available data and also support their use beyond their initial purpose and by researchers outside the project's consortium.
Interoperable to specific quality standards	The suggested file formats for every type of document and dataset indicate the project's objective to remove any standardization barriers that may prevent a number of users from assessing the data. Furthermore, the selected file formats are supported by free software packages and open source programming libraries that allow their use without additional costs.

4.3 Conformance to Principles of Medical Information Security

The following table summarizes the proposed solutions of MyAirCoach for the addressing of issues of medical information security

Table 9: Conformance with theHarvard Research Data Security Policy

Principle	Description
Access control.	The medical records of patients will be only accessible to

	their doctors and family members as identified by the patient. Furthermore, and after the informed consent of the patient an anonymized version of their record will be made available
Record opening	MyAirCoach records will be accessible by the patients themselves. In addition the open data repository will be also available to all users.
Control	The uploading of data or editing will be subjected to a detailed scheme of permissions and all uploaded data will be characterized by the name of their creator
Consent and notification	Informed consent of patients will be required before any type of publication or sharing of information within the consortium or with external users.
Persistence	No deletion functionalities of health record will be provided to any type of users. If a user requires the deletion of his/her health record or uploaded data a request should be sent to the ethical committee of the project for review.
Attribution	All uploaded data and changes will be marked with the user id of the respective creator. An audit trail will be kept in when deletions are performed, and after the approval of the ethical committee of the project.
Information flow	No information flow will be available between records within the MyAirCoach framework.
Aggregation control	Patients will have the control of the users that have access to their medical record, either through the anonymized or the detailed view.
Trusted Computing Base	Information technology experts will supervise the proper function of the system and report any risks for privacy and data security.

5 Conclusions

The purpose of the current deliverable of the MyAirCoach project is to support the data management life cycle for all the data that will be collected, processed or generated by the project. The datamanagement plan of the project consists of a detailed analysis of the datasets that the partners of the MyAirCoachproject plan to collect and use. Foreseen datasets contain inhaler usage measurements, physiology assessments, exhaled Nitric Oxidemeasurements, environmental measurements, patient tomography data, virtual models etc.

Each dataset was separately analyzed, with emphasis given on the nature of the data, the accessibility and its possible access type, as well as any ethical issues that may arise from manipulating sensitive personal information. This deliverable will serve as a guide

to build the infrastructure for efficiently managing, storing and distributing the amount of data collected, especially concerning the portions of the MyAirCoach datasets that will be made publicly available.

Furthermore a detailed demonstrator of the online open data management platform of the project is presented, showing the main functionalities implemented in this stage of the project and how it will be integrated with the online version of the MyAirCoach system. Furthermore, the user Centered Design and Development processes of the MyAirCoach together with the planned evaluation task will allow the optimization of the open data platform and towards its use from researchers outside the project's consortium and after the completion of the project activities.

Appendix 1: Deposit License Agreement

In order to guarantee the proper function of the online open data repository of MyAirCoach a License Agreement was prepared based on the respective document of the 3TU Datacentrum.[37]

The following parties are involved in this Licence Agreement:
1. The organization or person authorized to transfer and deposit the digital dataset/document(s),hereafter referred to as the Depositor
2. The organization that is authorized to archive and manage the digital dataset/document(s), hereafterreferred to as the Repository
The Depositor is:
The person or legal entity registered as such with the Repository
The Repository is:
MyAirCoach open access repository

This Licence Agreement is subject to the following provisions:

1. Licence

a. The Depositor grants the Repository a non-exclusive license for digital data files, hereafter referred to as 'dataset/document'.
b. The Repository is authorized to include the dataset/document in its data archive. The Repository shall transfer the content of the dataset/document to an available carrier, through any method and in any form.
c. The Repository is authorized to make the dataset/document (or substantial parts thereof) available to third parties by means of online transmission. In addition, the Repository has the right, on the instruction of third parties or otherwise, to make a copy of the dataset/document or to grant third parties permission to download a copy.

2. The Depositor

a. The Depositor declares that he is a holder of rights to the dataset/document, or the only holder of rights to the dataset/document, under the Databases act and where relevant the Copyright Actor otherwise, and/or is entitled to act in the present matter with the permission of other parties that hold rights.
b. By depositing a dataset/document the Depositor does not transfer ownership. The Depositor retains the right to deposit the dataset/document elsewhere in its present or future version(s). The Depositor retains all moral rights in the dataset/document including the right to be acknowledged as creator.
c. The Depositor indemnifies the Repository against all claims made by other parties against the Repository with regard to the dataset/document, the transfer of the dataset/document, and the form and/or content of the dataset/document.

3. The dataset/document

a. The dataset/document to which the license relates consists of all the databases, documentation and other data files and documents that form part of this dataset/document, which have been transferred by the Depositor.
b. The Depositor declares that the dataset/document corresponds to the specification provided.
c. The Depositor declares that the dataset/document contains no data or other elements that are contrary to European law.
d. The Depositor will supply the dataset/document by means of a method and medium deemed acceptable by the Repository.

4. The Repository

a. The Repository shall ensure, to the best of its ability and resources, that the deposited dataset/document is archived in a sustainable manner and remains legible and accessible.
b. The Repository shall, as far as possible, preserve the dataset/document unchanged in its original software format, taking account of current technology and the costs of implementation. The Repository has the right to modify the format and/or functionality of the dataset/document if this is necessary in order to facilitate the digital sustainability, distribution or re-use of the dataset/document.
c. If the access category “Temporary restriction: Embargo”, as specified at the end of this Agreement, is selected, the Repository shall, to the best of its ability and resources, ensure that effective technical and other measures are in place to prevent unauthorized third parties from gaining access to and/or consulting the dataset/document or substantial parts thereof.

5. Removal of dataset/documents

a. If sufficient weighty grounds exist, the Repository has the right to remove the dataset/document from the archive wholly or in part, or to restrict or prevent access to the dataset/document on a temporary or permanent basis. The Repository shall inform the Depositor in such cases.

6. Availability to third parties

a. The Repository shall make the dataset/document available to third parties in accordance with the access conditions agreed with the Depositor: "Open access", or the “Temporary restriction: Embargo”.
b. The Repository shall make the dataset/document available only to third parties who have agreed to comply with the General Conditions of Use.
c. Notwithstanding the above, the Repository can make the dataset/document (or substantial parts thereof) available to third parties:
    - if the Repository is required to do so by legislation or regulations, a court decision, or by a regulatory or other institution
    - if this is necessary for the preservation of the dataset/document and/or the data archive
    - (to a similar institution) if the Repository ceases to exist and/or its activities in the field of data archiving are terminated
d. The Repository shall publish the metadata and make them freely available, on the basis of the documentation that the Depositor provides with the

dataset/document. The term metadata refers to the information that describes the digital files.

e. The general information about the research and the metadata relating to the dataset/document shall be included in the Repository's databases and publications that are freely accessible to all persons.

7. Provisions relating to use by third parties

a. The Repository shall require third parties to whom the dataset/document (or substantial parts thereof) is made available to include in the research results a clear reference to the dataset/document from which data have been used. The reference must comply with the General Conditions of Use.

b. The Repository shall require parties to which a dataset/document is made available to grant a non-exclusive license for the dataset/document(s) they create using the dataset/document that has been made available.

8. Liability

a. The Repository accepts no liability in the event that all or part of a dataset/document is lost.

b. The Repository accepts no liability for any damage or losses resulting from acts or omissions by third parties to whom the Repository has made the dataset/document available.

c. The Repository accepts no responsibility for mistakes, omissions, or legal infringements within the deposited dataset/document.

9. Term and termination of the Agreement

a. This Agreement shall come into effect on the date on which the Repository receives the dataset/document (hereafter the deposit date) and shall remain valid for an indefinite period. If the repository decides not to include the dataset/document in its data archive, this Agreement is cancelled. The Repository notifies the Depositor of publication or non-inclusion of the dataset/document in its data archive. Cancellation of this Agreement is subject to a period of notice of six months, and notice shall be given in writing. It is possible to change the agreed access category at any time during the term of the Agreement.

b. Notwithstanding point (a), this Agreement shall end when the dataset/document is removed from the data archive in accordance with Article 5 of this Agreement.

c. If the Repository ceases to exist or terminates its data-archiving activities, the Repository shall attempt to transfer the data files to a similar organization that will continue the Agreement with the Depositor under similar conditions if possible.

10. Jurisdiction

MyAirCoach open data platform is entitled, but not obliged, to act independently against violations of the Copyright Act and/or any other intellectual property right of the holder(s) of rights to the dataset/document and/or the data from the dataset/document.

11. Applicable law

European law is applicable to this agreement.

The Depositor hereby agrees to the above provisions and the general code(s) of conduct referred to in this document.

References

[1]Guidelines on Data Management in Horizon 2020.Version 2.0. October 30, 2015. Available at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf . Assessed 2015

[2] A Bill Governing Collection, Use and Disclosure of Personal Health Information', British Medical Association 1995

[3]The Caldicott Report. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1769982/ (Assessed 2015)

[4]Harvard Research Data Security Policy. Available at: http://vpr.harvard.edu/pages/harvard-research-data-security-policy (Assessed 2015)

[5]Seheult, Jansen N., et al. "The Acoustic Features of Inhalation can be Used to Quantify Aerosol Delivery from a Diskus™ Dry Powder Inhaler." Pharmaceutical research 31.10 (2014): 2735-2747.

[6]D'Arcy, Shona, et al. "A method to assess adherence in inhaler use through analysis of acoustic recordings of inhaler events." (2014): e98701.

[7]Taylor, Terence E., et al. "An acoustic method to automatically detect pressurized metered dose inhaler actuations." Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE.IEEE, 2014.

[8]OpenEHR framework. Available at: http://www.openehr.org/. (Assessed 2015)

[9] Health Level Seven. Available at: http://www.hl7.org/implement/standards/ (Assessed 2015)

[10]Dweik, Raed A., et al. "An official ATS clinical practice guideline: interpretation of exhaled nitric oxide levels (FENO) for clinical applications." American journal of respiratory and critical care medicine 184.5 (2011): 602-615.

[11]NICE diagnostics guidance. Measuring fractional exhaled nitric oxide concentration in asthma: NIOX MINO, NIOX VERO and NObreath. Available at: https://www.nice.org.uk/guidance/dg12 (Assessed 2015)

[12]National Health and Nutrition Examination Survey.Exhaled Nitric Oxide. Available at: http://wwwn.cdc.gov/nchs/nhanes/2009-2010/ENX_F.htm. (Assessed 2015)

[13]Harnan, S., M. Essat, T. Gomersall, P. Tappenden, R. Wong, R. Lawson, I. Pavord, and M. Everard."Exhaled Nitric Oxide For The Diagnosis Of Asthma In Adults And Children: A Systematic Review."Value in Health 18, no. 7 (2015): A345.\

[14]Petsky, H. L., Kew, K. M., Turner, C., Kynaston, J. A., & Chang, A. B. (2015). Exhaled nitric oxide levels to guide treatment for adults with asthma. The Cochrane Library.

[15]Petsky, H. L., Kew, K. M., Kynaston, J. A., Turner, C., & Chang, A. B. (2015). Exhaled nitric oxide levels to guide treatment for children with asthma. The Cochrane Library.

[16] London Air Quality Network – King's College London. Available at: http://www.londonair.org.uk/london/asp/datadownload.asp (Assessed 2015)

[17]Air Quality – The City of London. Available at: http://www.cityoflondon.gov.uk/business/environmental-health/environmental-protection/air-quality/Pages/Air-Quality.aspx (Assessed 2015)

[18]Air quality information and campaigns – Manchester City Council. Available at: http://www.manchester.gov.uk/info/100006/environmental_problems/2942/air_quality_information_and_campaigns/3 (Assessed 2015)
[19]Weather data for research and projects – University of Reading. Available at: http://www.greatairmanchester.org.uk/ (Assessed 2015)
[20]Weather data for research and projects – University of Reading. Available at: http://www.met.rdg.ac.uk/~brugge/data_sources.html (Assessed 2015)
[21]Historical monthly open data for UK meteorological stations – Met Office. Available at: https://data.gov.uk/dataset/historic-monthly-meteorological-station-data (Assessed 2015)
[22] UK Humidity open datasets. Available at: https://data.gov.uk/data/search?tags=humidity (Assessed 2015)
[23]Scatena, R., Q. Liu, C. Holm, N. P. Grant, X. Yan, W. Sessa, G. L. Chupp, and H. Rajeevan. "Effect Of Atmospheric Conditions On Asthma Control And Gene Expression In The Airway." Am J RespirCrit Care Med 191 (2015): A6270.

[24]Hebbern, Christopher, and SabitCakmak. "Synoptic weather types and aeroallergens modify the effect of air pollution on hospitalisations for asthma hospitalisations in Canadian cities." Environmental Pollution 204 (2015): 9-16.

[25]Open-Access Medical Image Repositories . Available by Dr. Stephen Aylward at: http://www.aylward.org/notes/open-access-medical-image-repositories (Assessed 2015)
[26]Public Medical Image Databases – Cornell University.Vision and Image Analysis Group. Available at: http://www.via.cornell.edu/databases/ (Assessed 2015)
[27]DICOM sample image sets. Available by OsiriX Imaging Software at: http://www.osirix-viewer.com/datasets/ (Assessed 2015)
[28]MRI and CT Data from The Visible Human Project .Available by the U.S. National Library of Medicine at: https://www.nlm.nih.gov/research/visible/getting_data.html (Assessed 2015)
[29]Bone and Joint CT-SCAN Data – International Society of Biomechanics. Available at: http://isbweb.org/data/vsj/ (Assessed 2015)
[30]Sample DICOM Data – TRIPOD. Available at: http://members.tripod.com/clunis_immensus/free3d/sample_data.htm (Assessed 2015)
[31]DICOM standard 2015. Available at: http://dicom.nema.org/standard.html (Assessed 2015)

[32]Dicom Library. Available at: http://www.dicomlibrary.com/ (Assessed 2015)

[33] Asthma Handouts – Sutter Health .Palo Alto Medical Foundation. Available at: http://www.pamf.org/asthma/education/handouts.html (Assessed 2015)
[34] Asthma Education Materials – Neighborhood Health Plan. Available at: https://www.nhp.org/provider/clinical/Pages/Asthma-Education-Materials.aspx (Assessed 2015)
[35]Global Initiative for asthma.Instructions for Inhaler and Spacer Use. Available at: http://www.ginasthma.org/Instructions-for-Inhaler-and-Spacer-Use (Assessed 2015)

[36]ADMIT. Inhalation Protocols. Available at: http://www.admit-inhalers.org/index.php?option=com_content&view=article&id=179&Itemid=350&lang=en (Assessed 2015)

[37]3TU Datacentrum.Uploading License Agreement. Available at: http://datacentrum.3tu.nl/en/publishing-research/uploading-data/ (Assessed 2015)