



D1.3: 1st Data Management Plan

WP1 – Project Management



D1.3: 1st Data Management Plan

Document Information

Grant Agreement Number	688363	Acronym	hackAIR
Full Title	Collective awareness platform for outdoor air pollution		
Start Date	1 st January 2016	Duration	36 months
Project URL	www.hackAIR.eu		
Deliverable	D 1.3 – 1 st Data Management Plan		
Work Package	WP 1 – Project management		
Date of Delivery	Contractual	30 June 2016	Actual 30 June 2016
Nature	Report	Dissemination Level	Public
Lead Beneficiary	DRAXIS Environmental S.A.		
Responsible Authors	Stavros Tekes (DRAXIS), Ioulia Anastasiadou (DRAXIS), Panagiota Syropoulou (DRAXIS)		
Contributions from	Philipp Schneider (NILU), Laurence Claeys (VUB), Wiebke Herding (ON:SUBJECT), Anastasia Moumtzidou (CERTH), Aristeidis K. Georgoulas (DRAXIS), Kostas Kourtidis (DRAXIS), Ilias Stavrakas (DRAXIS)		

Document History

Version	Issue Date	Stage	Description	Contributor
	06/06/2016	Draft	DMP Template provided	DRAXIS
	07/06/2016	Draft	Datasets for WP4 and WP7	NILU
	14/06/2016	Draft	Datasets for WP2 and WP6	VUB
	16/06/2016	Draft	Datasets for WP8	ON:SUBJECT
	17/06/2016	Draft	Datasets WP3	CERTH
1.0	24/06/2016	Draft	First version incorporating input datasets from all WPs	DRAXIS
2.0	01/07/16	Final	Final version	DRAXIS

Disclaimer

Any dissemination of results reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

Copyright message

© hackAIR Consortium, 2016

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorised provided the source is acknowledged.



Table of Contents

1 Executive summary.....	5
2 Introduction.....	6
3 Methodology	7
3.1 Data set reference and name	7
3.2 Data set description.....	7
3.3 Standards and metadata.....	7
3.4 Data sharing.....	8
3.5 Archiving and preservation (including storage and backup).....	8
3.6 Discoverable	9
3.7 Accessible	9
3.8 Assessable and intelligible	9
3.9 Usable beyond the original purpose for which it was collected	9
3.10 Interoperable to specific quality standards	9
4 Datasets in hackAIR	10
4.1 Datasets in WP1 – Project management (DRAXIS)	10
4.2 Datasets in WP2 - Analysis and requirements (VUB)	10
4.2.1 User requirements (intake survey) data set	11
4.2.2 User requirements (workshop) data.....	12
4.3 Datasets in WP3 - Collective sensing models and tools (CERTH)	13
4.3.1 Geotagged Images Dataset.....	13
4.3.2 Web cams Dataset.....	14
4.3.3 Environmental measurements Dataset	15
4.3.4 Look-up Table	16
4.4 Datasets in WP4 - Data fusion model and reasoning services (NILU)	17
4.4.1 CAMS regional modelling results	17
4.4.2 Data fusion maps	18
4.5 Datasets in WP5 - Development of the hackAIR platform (DRAXIS)	18
4.5.1 Architecture and Integration Framework Definition Specification	18
4.6 Datasets in WP6 - Engagement strategies for user participation (VUB)	19
4.6.1 Engagement survey data set.....	19
4.6.2 Engagement qualitative insights data set	20
4.6.3 Behavioral change survey data set	21
4.6.4 Behavioral change qualitative insights data set.....	23
4.7 Datasets in WP7 - Pilot operation and evaluation (NILU)	24
4.7.1 PM concentrations by PSOC system	24

D1.3: 1st Data Management Plan

4.7.2 PM concentrations by Arduino sensor.....	25
4.7.3 PM concentrations by COTS system	25
4.7.4 Content generated through the hackAIR platform	26
4.7.5 Photos uploaded through the hackAIR mobile app	27
4.8 Datasets in WP8 - Communication, Dissemination and Exploitation (ONSUB)	28
4.8.1 Contact database	29
4.8.2 Communications monitoring data	29

1 Executive summary

The purpose of the current deliverable is to present the 1st Data Management plan of the hackAIR project. The deliverable is a collective product of work among the coordinator and the rest of the consortium partners. It includes detailed descriptions of all datasets that will be collected, processed or generated in all Work packages during the course of the 36 months of hackAIR project. The deliverable is submitted on Month 6 as required by the Open Research Data Pilot. The methodology proposed by the European Commission Guidelines has been adopted for the deliverable compilation.

The deliverable is structured in the following chapters:

Chapter 2 includes an introduction to the Deliverable.

Chapter 3 includes a description of the methodology used.

Chapter 4 includes the description of the datasets.

2 Introduction

The current deliverable D1.3 “1st Data management plan” represents the first version of the Data Management Plan (DMP) of hackAIR project. Since hackAIR project participates in the H2020 Open Research Data Pilot, the deliverable D1.3 “1st Data Management plan” is submitted on Month 6 (June 2016) of the project as required.

The purpose of this deliverable is to provide a detailed description of all the datasets that will be collected, processed or generated during the course of hackAIR project, to specify which datasets will be openly accessible, to describe the handling of research data during and after the project, to describe the methodology and standards required but also to identify whether and how data will be shared, exploited or made accessible for verification and re-use, and how they will be curated and preserved. This document is the initial of the three versions to be produced for the Data Management Plan throughout the hackAIR project’s duration, serving as a working document.

The DMP provides an analysis of the main elements of the data management policy that will be used by the hackAIR consortium with regard to all the datasets that will be generated by the project. The DMP is not a fixed document, but instead it evolves during the lifespan of the project. In this respect the 2nd version will be submitted on Month 18 (June 2017), while the 3rd and final version is due on Month 36 (December 2018).

The methodological approach used to produce the initial hackAIR Data management plan is presented in the next chapter.

3 Methodology

The methodological approach that has been used for the compilation of D1.3 follows the updated version of the “Guidelines on Data Management in Horizon 2020”¹ version 2.1 released on 15 February 2016 by the European Commission Directorate - General for Research & Innovation. Taking into account the proposed methodology, the hackAIR DMP addresses the following points below on a dataset by dataset basis:

- Data set reference and name
- Data set description
- Standards and metadata
- Data sharing
- Archiving and preservation (including storage and backup).

Additionally in order to provide further clarification for all datasets produced in the project also the following points have been addressed on a dataset by dataset basis:

- Discoverable
- Accessible
- Assessable and intelligible
- Usable beyond the original purpose for which it was collected
- Interoperable to specific quality standards

The hackAIR project coordinator (DRAXIS) has provided on time all the work package leaders and rest of partners with a dataset description template that included all the 10 abovementioned points along with instructions to fill the template and related literature. Project partners have been notified on time by the project coordinator in order to provide their input.

A detailed description of the 10 main points of the DMP is provided in the following chapters.

3.1 Data set reference and name

This point is the identifier for the dataset to be produced. The hackAIR dataset identification follows the naming: Data_<WPno>_<serial number of dataset>_<dataset title>. Example: Data_WP2_1_ user requirements (intake survey) data set.

3.2 Data set description

In this point the data that will be generated or collected is described, including references on their origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Where applicable, information on the existence (or not) of similar data and the possibilities for integration and reuse, are mentioned.

3.3 Standards and metadata

This point refers to existing suitable standards of the discipline, as well as an outline on how and what metadata will be created. Therefore, at this stage, the available data standards (if any) accompany the description of the data that will be collected and/or generated, including the description on how the data will be organised during the project, mentioning for example naming conventions, version control and folder structures.

The following questions can be considered as guidance on Data Capture Methods:

¹ European Commission, (15 February 2016), *Guidelines on Data Management in Horizon 2020*, Version 2.1

D1.3: 1st Data Management Plan

- How will the data be created?
- What standards or methodologies will you use?
- How will you structure and name your folders and files?
- How will you ensure that different versions of a dataset are easily identifiable?

As far as the metadata are concerned, the way the consortium will capture and store this information should be described. For example, for data records stored in a database with links to each item metadata can pinpoint their description and location. There are various disciplinary metadata standards², however the hackAIR consortium has identified a number of available best practices and guidelines for working with Open Data, mostly by organizations or institutions that support and promote Open Data initiatives, and will be taken into account. These include:

- Open Data Foundation³
- Open Knowledge Foundation⁴
- Open Government Standards⁵

Data will be interoperable, adhering to standards for data annotation, data exchange, compliant with available software applications, and allowing re-combinations with different datasets from different origins. Standards such as the Dublin Core and ISO/IEC 11179 Metadata Registry (MDR), which addresses issues in the metadata and data modelling space, will be taken into account.

3.4 Data sharing

This point will describe how data will be shared, including access procedures, embargo periods (if any), while it will outline the technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and, finally, define 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 will be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

Within hackAIR all personal data used in the project will be protected, while the informed consent of all participants will be a prerequisite for their participation in the project activities. When possible, the data collected in the project will be available to third parties in contexts such as scientific scrutiny and peer review. As documented in the Deliverable D1.2 Project Management Handbook, deliverables' external reviewers will sign a confidentiality declaration, which includes the following statement:

"I hereby declare that I will treat all information, contained within the above mentioned deliverable and which has been disclosed to me through the review of this deliverable, with due confidentiality."

Finally, it is expected that the hackAIR project will result in a number of publications in scientific, peer-reviewed journals. Project partners are encouraged to collaborate with each other and jointly prepare publications relevant to the hackAIR project. Scientific journals that provide open access (OA) to all their publications will be preferred, as it is required by the European Commission.

3.5 Archiving and preservation (including storage and backup)

In this point the procedures that will be put in place for long-term preservation of the data will be described, along with the indication of how long the data should be preserved, what is its approximated end volume, including a reference to the associated costs (if any) and how these are planned to be covered. This point emphasizes in the

² <http://www.dcc.ac.uk/resources/metadata-standards>

³ <http://www.opendatafoundation.org/>

⁴ <https://okfn.org/>

⁵ <http://www.opengovstandards.org/>

D1.3: 1st Data Management Plan

long-term preservation and curation of data, beyond the lifetime of the hackAIR project. Where dedicated resources are needed, these should be outlined and justified, including any relevant technical expertise, support and training that is likely to be required and how it will be acquired.

3.6 Discoverable

This point further defines whether the data and associated software produced and/or used in the project are discoverable (and readily located), and identifiable by means of a standard identification mechanism such as the Digital Object Identifier which is a serial code used to uniquely identify objects.

3.7 Accessible

This point further defines whether the data and associated software produced and/or used in the project are accessible and in what modalities, scope and licenses. Examples include licensing framework for research and education, embargo periods and commercial exploitation.

3.8 Assessable and intelligible

This point defines whether the data and associated software produced and/or used in the project are assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review. This point also clarifies whether the minimal datasets are handled together with scientific papers for the purpose of peer review, or cases where data is provided in a way that judgments can be made about their reliability and the competence of those who created them.

3.9 Usable beyond the original purpose for which it was collected

This point defines whether the data and associated software produced and/or used in the project are useable by third parties even long time after the collection of the data. It clarifies cases where the data safely stored in certified repositories for long term preservation and curation, stored together with the minimum software, metadata and documentation to make it useful or case that data useful for the wider public needs and usable for the likely purposes of non-specialists.

3.10 Interoperable to specific quality standards

This point defines whether the data and associated software produced and/or used in the project are interoperable allowing data exchange between researchers, institutions, organisations and countries. Examples include adhering to standards for data annotation, data exchange, compliant with available software applications, and allowing re-combinations with different datasets from different origins.

4 Datasets in hackAIR

4.1 Datasets in WP1 – Project management (DRAXIS)

As for the 1st hackAIR DMP WP1 does not generate research data. For the purposes of WP1 the following dataset will be generated: Contact details of project partners and advisory board.

Data set reference and name	Data_WP1_1_ Contact details of project partners and advisory board
Data set description	The database contains name, organisation and contact details for all project partners and advisory board members.
Standards and metadata	The data is stored in a simple table, with the following fields: <ul style="list-style-type: none"> - Name - Category (1: Partner, 2: Advisory Board Member) - Short description - Location - Link - Email - Comments Additional fields will be added as the project progresses.
Data sharing	The data is confidential for internal use within the consortium (as personal data is involved). The contact database is managed in a Google Spreadsheet, and linked from the project internal wiki.
Archiving and preservation (including storage and backup)	The data is collected for internal use in the project, and not intended for long-term preservation. The work package leader is keeping a quarterly backup on a separate disk.
Discoverable	No
Accessible	No
Assessable and intelligible	No
Useable beyond the original purpose for which it was collected	No
Interoperable to specific quality standards	CSV

4.2 Datasets in WP2 - Analysis and requirements (VUB)

As for the 1st hackAIR DMP WP2 generates research data. For the purposes of WP2 the following two datasets will be generated:

D1.3: 1st Data Management Plan

- User requirements (intake survey) data set
- User requirements (workshop) data

4.2.1 User requirements (intake survey) data set

Data set reference and name	Data_WP2_1_ user requirements (intake survey) data set
Data set description	<p>The user requirements (intake survey) data set will contain all relevant data for designing, running and analysing the co-creation intake survey, to define the user requirements of the hackAIR platform.</p> <p>Consortium partners involved: VUB, BUND, NILU</p> <p>The following data will be collected:</p> <ul style="list-style-type: none"> – Demographics: gender, birth year, occupation – Device ownership and internet access – Personal innovativeness (scale) – Air quality awareness (scale) <p>Data collection method: paper questionnaire</p> <p>Data subjects:</p> <ul style="list-style-type: none"> – Amount: 12 to 16 data subjects – Citizens of Berlin and Oslo <p>Data type: .xls</p> <p>No reuse of the data</p> <p>Main aggregated and anonymous research findings will be discussed in scientific research publications.</p> <p>Only participants who sign the informed consent statement at the start of the workshop will fill in the intake survey. By signing this form, they give permission for the use and disclosure of anonymous information for scientific purposes of this study at any time in the future.</p> <p>Time period:</p> <ul style="list-style-type: none"> – M6: collection intake survey data
Standards and metadata	There are no specific standards or metadata associated with these types of data.
Data sharing	Only aggregated and anonymous data and research results will be published in project reporting documents (accessible to all consortium partners via the hackAIR project Wiki - the central repository of internal information for the hackAIR project) and in public research publications from the project.
Archiving and preservation (including storage and backup)	Aggregated and anonymous data and research results will be saved on the hackAIR project wiki.
Discoverable	Not applicable
Accessible	Not applicable
Assessable and intelligible	Not applicable. In case of a report or paper submitted for publication with peer review, all research findings will be integrated into the report or paper. Datasets will never be added to the publication.
Useable beyond the original purpose for which it was	Only aggregated and anonymous research findings published in public reports or deliverables and published in academic papers will be and stay accessible from the

D1.3: 1st Data Management Plan

collected	moment of publication.
Interoperable to specific quality standards	Not applicable.

4.2.2 User requirements (workshop) data

Data set reference and name	Data_WP2_2_ user requirements (workshop) data set
Data set description	<p>The user requirements (workshop) data set will contain all relevant data for designing, running and analysing the co-creation workshop, to define the user requirements of the hackAIR platform.</p> <p>Consortium partners involved: VUB, BUND, NILU</p> <p>The following data will be collected:</p> <ul style="list-style-type: none"> • Qualitative insights into: (1) Experiences, practices and expectations with regards to measuring and retrieving air quality information, (2) Expectations with regards to the hackAIR platform and (3) Evaluation of the hackAIR platform • Contact information (name, email): Only known to the local organizers (BUND, NILU) for recruiting purposes <p>Data collection method: co-creation workshop</p> <p>Data subjects:</p> <ul style="list-style-type: none"> – Amount: 24 to 32 data subjects (12 to 16 data subjects in M6 + 12 to 16 data subjects in M10) – Citizens of Berlin and Oslo <p>Data type:</p> <ul style="list-style-type: none"> – Text – Audio-records: The workshop will be audio-recorded for post-processing. This tape will be used by the involved researchers (BUND and NILU) only for the processing of the workshop findings. It will only serve research purposes and it will by no means be released to other persons. <p>No reuse of the data (but open to discuss on request)</p> <p>All participants will be coded (by using pseudonyms) in the processing and reporting of the research results. This means that real names will not be associated in any way with the information collected or with the research findings from this study.</p> <p>Aggregated and pseudonymised research findings will be discussed in scientific research publications.</p> <p>Only participants who sign the informed consent statement at the start of the workshop will participate. By signing this form, they give permission for the use and disclosure of pseudonymised information for scientific purposes of this study at any time in the future and for the audio-recording of the workshop only for post-processing purposes.</p> <p>Time period:</p> <ul style="list-style-type: none"> – M6: workshop phase 1

D1.3: 1st Data Management Plan

	– M10: workshop phase 2
Standards and metadata	There are no specific standards or metadata associated with these types of data.
Data sharing	Aggregated and pseudonymised data and research results will be published in project reporting documents (accessible to all consortium partners via the hackAIR project Wiki - the central repository of internal information for the hackAIR project) and in external research publications from the project.
Archiving and preservation (including storage and backup)	Aggregated and pseudonymised data and research results will be saved on the hackAIR project wiki.
Discoverable	Not applicable
Accessible	Not applicable
Assessable and intelligible	Not applicable. In case of a report or paper submitted for publication with peer review, all research findings will be integrated into the report or paper. Datasets will never be added to the publication.
Useable beyond the original purpose for which it was collected	Only aggregated and pseudonymised research findings published in public reports or deliverables and published in academic papers, will be and stay accessible from the moment of publication.
Interoperable to specific quality standards	Not applicable.

4.3 Datasets in WP3 - Collective sensing models and tools (CERTH)

As for the 1st hackAIR DMP WP3 generates research data. For the purposes of WP3 the following four datasets will be generated:

- Geotagged Images Dataset
- Web cams Dataset
- Environmental measurements Dataset
- Look-up Table

4.3.1 Geotagged Images Dataset

Data set reference and name	Data_WP3_1_Geotagged_Images dataset
Data set description	<p>The dataset contains user generated images that are publicly available through social media platforms such as Instagram and Flickr. The images share the following features: they are geotagged and time stamped, and they contain a sufficiently large region that is the sky (which makes them usable as potential sources for estimating air quality information).</p> <p>Currently, the dataset contains images from several cities in Germany, Norway and some other European cities such as Brussels, Amsterdam, Thessaloniki, Athens, and spans the period between Jan 1st and Jun 1st, 2016.</p>
Standards and metadata	<p>The images will be processed and the extracted metadata that will be provided are the following:</p> <ul style="list-style-type: none"> – Image URL

D1.3: 1st Data Management Plan

	<ul style="list-style-type: none"> – Image location and timestamp – Image feature descriptors (such as the ones produced by Convolutional Neural Networks) that are useful for image classification – Mask of automatically detected sky region – R/G and G/B ratios of sky part of the image
Data sharing	<p>The original images cannot be shared due to Instagram’s and Flickr’s privacy and copyright policies.</p> <p>Hence, only the extracted image metadata, including references to the image files (URLs), will be shared.</p>
Archiving and preservation (including storage and backup)	<p>The original images will be stored only for the period necessary to provide the developed app's service following the privacy policies of the platforms.</p> <p>The extracted metadata will be archived and preserved to an online repository such as figshare or GitHub.</p>
Discoverable	<p>The original images can be retrieved from the corresponding URLs. The metadata produced will be available either through GitHub or figshare (the latter providing DOIs) that will make them discoverable and identifiable.</p> <p>Regarding the software:</p> <ul style="list-style-type: none"> – ‘sky localization’ is based on open software which can be found, on GitHub (https://github.com/BVLC/caffe/wiki/Model-Zoo#fcn) – ‘sky detection’ is using non-open software. <p>The dataset will be discoverable by querying conventional search engines (e.g. Google) with the dataset name.</p>
Accessible	The metadata and software will be searchable and referenceable.
Assessable and intelligible	Sufficient documentation of the software and data will be provided that allow assessment and review by third parties.
Useable beyond the original purpose for which it was collected	The data can be used by computer vision researchers for several tasks (e.g. image retrieval, concept detection, etc.)
Interoperable to specific quality standards	The metadata will be available in text format that will allow easy parsing and information exchange.

4.3.2 Web cams Dataset

Data set reference and name	Data_WP3_2_Webcams dataset
Data set description	<p>The set contains images extracted from static outdoor webcams. The webcams are geotagged and as a consequence the images are also geotagged and time stamped.</p> <p>The webcams targeted will cover several cities in Germany, Norway and some other European cities such as Brussels, Amsterdam, Thessaloniki, Athens.</p>
Standards and metadata	The images will be processed and the metadata that will be provided are the same as those of the Data_WP3_2_Geotagged_Images dataset, and in addition will include references (URLs) to the web pages where the web cams are contained.
Data sharing	The original images cannot be shared due to potential copyright terms set by some

D1.3: 1st Data Management Plan

	<p>of the webcam owners.</p> <p>The extracted metadata will be open.</p>
Archiving and preservation (including storage and backup)	<p>The original images will be stored only for the period necessary to provide the developed app's service following the privacy policies of the platforms.</p> <p>The extracted metadata will be archived and preserved to an online repository such as figshare or GitHub.</p>
Discoverable	<p>The original images cannot be retrieved, however the metadata produced will be available either through GitHub or figshare.</p> <p>Regarding the software:</p> <ul style="list-style-type: none"> – ‘sky localization’ is based on open software which can be found, on GitHub (https://github.com/BVLC/caffe/wiki/Model-Zoo#fcn) – ‘sky detection’ is using non-open software <p>The dataset will be discoverable by querying conventional search engines (e.g. Google) with the dataset name.</p>
Accessible	The metadata and software will be searchable and referenceable.
Assessable and intelligible	Sufficient documentation of the software and data will be provided that allow assessment and review by third parties.
Useable beyond the original purpose for which it was collected	The data (for which we have the necessary rights) can be used by computer vision researchers in several tasks (e.g. image retrieval, concept detection, etc.)
Interoperable to specific quality standards	The metadata will be available in text format that will allow easy parsing and information exchange.

4.3.3 Environmental measurements Dataset

Data set reference and name	Data_WP3_3_Environmental dataset
Data set description	<p>The dataset contains files with environmental measurements that were published in environmental web sites. The dataset contains information such as the location to which the measurements refer, the time stamp, and the air pollutant.</p> <p>The sites targeted will cover several cities in Germany, Norway and some other European cities such as Brussels, Amsterdam, Thessaloniki, Athens.</p>
Standards and metadata	<p>The web sites will be processed and the following information will be provided</p> <ul style="list-style-type: none"> – Web site URL – Air pollutant – Timestamp of measurement – Location – Value of measurement <p>This information will be available both in csv files and in a SOS server (SOS is an approved standard of the Open Geospatial Consortium) like 52North.</p>
Data sharing	The files with the environmental measurements will be available freely.
Archiving and preservation	The files will be stored on GitHub or figshare.

D1.3: 1st Data Management Plan

(including storage and backup)	
Discoverable	<p>The website resources from which this data retrieved will be cited. The software for web page information extraction is available on GitHub (https://github.com/MKLab-ITI/easIE/).</p> <p>The dataset will be discoverable by querying conventional search engines (e.g. Google) with the dataset name.</p>
Accessible	The files and software will be searchable and referenceable.
Assessable and intelligible	Sufficient documentation of the software and data will be provided that allow assessment and review by third parties.
Useable beyond the original purpose for which it was collected	N/A
Interoperable to specific quality standards	The metadata will be available in text format that will allow easy parsing and information exchange.

4.3.4 Look-up Table

Data set reference and name	Data_WP3_4_Look-up_Table
Data set description	<p>Radiative transfer calculations have been implemented using the SBDART (Santa Barbara DISORT Atmospheric Radiative Transfer) radiative transfer model in order to create a Look-up Table (LUT) with Red (700nm) / Green (550nm) band ratios and Green (550nm) / Blue band (450nm) ratios. The LUT consists of one ASCII file per geographical grid cell of 5x5 degrees (a total of 2592 files for the whole globe). Each ASCII includes R/G and G/B ratios on an hourly basis, per specified AOD₅₅₀ bins, sky viewing angles and directions relative to the sun. The input data used for the radiative transfer calculations leading to the LUT are from the well-established MACV1 aerosol climatology (AOD₅₅₀, single scattering albedo and asymmetry parameter) and the ERA-Interim reanalysis (total column ozone, water vapour, surface albedo).</p> <p>The LUT will be used in order to calculate the AOD₅₅₀ levels (particulate pollution) in the atmosphere from photos which are available in social media. After calculating the R/G and G/B ratios of the photos the LUT will allow for the attribution of the ratios to AOD₅₅₀ values.</p> <p>The radiative transfer calculations are driven by a script code which allows for the automatic generation of LUTs of various spatial and temporal resolutions depending on the needs of the project.</p>
Standards and metadata	Readme files with the parameters included in the LUT files and the method followed for the production of the LUT.
Data sharing	The LUT dataset is for use only within the project. The LUT is a product of collaboration between DRAXIS and DUTH and should remain available only to project members as the LUT could potentially be used in the future for other scientific or commercial activities.
Archiving and preservation	The LUT data files will only occupy some MBs of space and hence their archiving and

D1.3: 1st Data Management Plan

(including storage and backup)	preservation is particularly easy.
Discoverable	The LUT data are not identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier).
Accessible	The LUT data files should not be accessible to third party users.
Assessable and intelligible	The LUT data are not assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review.
Useable beyond the original purpose for which it was collected	The LUT data should not be useable by third parties even long time after the end of the project. The LUT should only be used in the future for other scientific or commercial activities from DRAXIS and DUTH.
Interoperable to specific quality standards	The LUT data could be easily used in the future by atmospheric aerosol retrieval algorithms that use photos, sky camera images, etc.

4.4 Datasets in WP4 - Data fusion model and reasoning services (NILU)

For the purposes of WP4 the following two datasets will be generated:

- CAMS regional modelling results
- Data fusion maps

4.4.1 CAMS regional modelling results

Data set reference and name	Data_WP4_1_CAMS_Modelling_Results
Data set description	This dataset is operationally produced by the Copernicus Atmosphere Monitoring Service (CAMS). More specifically it is the CAMS Regional Ensemble Forecast for Europe. The dataset is used within WP4 to support the spatial mapping of the hackAIR observations. It is used only temporarily in hackAIR for processing and then discarded.
Standards and metadata	NetCDF
Data sharing	The original dataset will not be shared with the public as part of hackAIR. They are merely used as a supporting dataset within the processing of the hackAIR observations.
Archiving and preservation (including storage and backup)	hackAIR will not archive or preserve this dataset as this is being done as part of CAMS. The dataset will only be temporarily stored on hackAIR servers.
Discoverable	Not discoverable by public on hackAIR servers.
Accessible	Not accessible by public on hackAIR servers.
Assessable and intelligible	Not accessible by public on hackAIR servers.
Useable beyond the original purpose for which it was collected	Not applicable.
Interoperable to specific quality standards	Not applicable.

4.4.2 Data fusion maps

Data set reference and name	Data_WP4_2_Fused_Maps
Data set description	This dataset is the primary outcome of WP4. It includes the data fusion maps that will be produced by combining the hackAIR observations with forecast from the CAMS regional model ensemble (Data_WP4_1_CAMS_Modelling_Results).
Standards and metadata	The dataset will be exported in the standard GeoTIFF format and then subsequently imported into a database.
Data sharing	The data will be shared in the form of online maps on the official hackAIR website.
Archiving and preservation (including storage and backup)	The fused maps will be archived on the hackAIR server.
Discoverable	No.
Accessible	The code written for generating the data fusion maps will be made available as open source (license type to be determined).
Assessable and intelligible	The dataset will be accompanied by a deliverable describing the inputs, the processing methodology, and the output data format.
Useable beyond the original purpose for which it was collected	The dataset will be stored together with the corresponding code and its documentation.
Interoperable to specific quality standards	The code used for producing this dataset is written in the standard open-source R programming language. The fused maps themselves will be provided in the GeoTIFF format, which is a public domain metadata standard which allows georeferencing information to be embedded within a TIFF file.

4.5 Datasets in WP5 - Development of the hackAIR platform (DRAXIS)

As for the 1st hackAIR DMP WP5 does not generated research data. For the purposes of WP5 the following dataset will be generated:

- Architecture and Integration Framework Definition Specification

4.5.1 Architecture and Integration Framework Definition Specification

Data set reference and name	Data_WP5_1_Architecture_and_Integration_Framework_Definition_Specification
Data set description	Functional and non-functional requirements, hardware requirements, component descriptions (inputs & outputs), component dependencies, API descriptions, information flow diagram, internal and external interfaces and testing procedures. All technical partners were asked to answer a set of questions, based on which further online and face to face discussions took place in order to form the final document. This will be the basis upon which the system will be built.
Standards and metadata	There are no specific standards or metadata associated with these types of data.
Data sharing	The data will be available in D5.1: Architecture and Integration Framework Definition Specification. The dissemination level of D5.1 is public. It will be available through the hackAIR wiki for the members of the consortium and when the project decides to publicize deliverables, it will be uploaded along with the other public deliverables to

D1.3: 1st Data Management Plan

	the project website or anywhere else the consortium decides.
Archiving and preservation (including storage and backup)	All data will be saved in the DRAXIS premises and will be shared with all partners using the hackAIR wiki.
Discoverable	It will become both discoverable and accessible to the public once it is delivered to the EU and the consortium decides to do so.
Accessible	Not accessible to the public during the creation process. It will become both discoverable and accessible to the public once it is delivered to the EU and the consortium decides to do so.
Assessable and intelligible	Sufficient information will be provided within the deliverable to allow assessment and review by third parties.
Useable beyond the original purpose for which it was collected	Could be used as example for engineers who want to build similar systems
Interoperable to specific quality standards	N/A

4.6 Datasets in WP6 - Engagement strategies for user participation (VUB)

As for the 1st hackAIR DMP WP6 generates research data. For the purposes of WP6 the following four datasets will be generated:

- Engagement survey data set
- Engagement qualitative insights data set
- Behavioral change survey data set
- Behavioral change qualitative insights data set

4.6.1 Engagement survey data set

Data set reference and name	Data_WP6_1_Engagement survey data set
Data set description	<p>The engagement survey data set will contain relevant quantitative data for designing, running and analysing the engagement strategies.</p> <p>Consortium partners involved: VUB</p> <p>The following data will be collected:</p> <ul style="list-style-type: none"> – Demographic information: only used for the profiling of the research participants – Air quality awareness (use of standardized scales) – Motivations to start and to continue engaging into hackAIR and other citizen science initiatives (use of standardized scales) <p>Data collection method: online survey</p> <p>Data subjects:</p> <ul style="list-style-type: none"> – Amount: 100 up to 150 <p>Data type: SPSS format .sav /MS Office .xls format</p> <p>Reuse of the anonymized dataset</p>

D1.3: 1st Data Management Plan

	Anonymous and aggregated research data will be published in internal project reports (accessible to all consortium partners via the hackAIR project Wiki) and in external scientific research publications from the project. Time period: M4-20
Standards and metadata	There are no specific standards or metadata associated with these types of data.
Data sharing	Aggregated and anonymized data and research results will be published in project reporting documents (accessible to all consortium partners via the hackAIR project Wiki) and in external research publications from the project. The anonymized dataset can be provided upon request.
Archiving and preservation (including storage and backup)	Aggregated and anonymized data and research results will be saved on the hackAIR project wiki. This wiki is the central repository of internal information for the hackAIR project.
Discoverable	Not applicable
Accessible	Not applicable
Assessable and intelligible	In case of a report or paper submitted for publication with peer review, aggregated and anonymized research findings will be integrated into the report or paper. In the contexts of scientific scrutiny and peer review, the full anonymized dataset can be made accessible to third parties.
Useable beyond the original purpose for which it was collected	Only aggregated and anonymized research findings published in public reports or deliverables and published in academic papers, will be and stay accessible from the moment of publication.
Interoperable to specific quality standards	Not applicable.

4.6.2 Engagement qualitative insights data set

Data set reference and name	Data_WP6_2_Engagement qualitative insights data set
Data set description	<p>The engagement qualitative insights data set will contain all relevant qualitative data for designing, running and analysing the engagement strategies.</p> <p>Consortium partners involved: VUB</p> <p>The following data will be collected:</p> <ul style="list-style-type: none"> – Demographic information: only used for the profiling of the research participants – Insights into current air quality awareness and into motivations to start and to continue engaging into hackAIR and other citizen science initiatives – Contact information: Only collected and used for recruiting purposes <p>Data collection method: focus group, expert interviews</p> <p>Data subjects:</p> <ul style="list-style-type: none"> – Amount: 20 test subjects (approximately) – European citizens and domain experts

D1.3: 1st Data Management Plan

	<p>Data type:</p> <ul style="list-style-type: none"> – Text – Audio-records: The focus groups and interviews might be audio-recorded for post-processing. This tape will be used by the involved researchers only for the processing of the workshop findings. It will only serve research purposes and it will by no means be released to other persons. <p>No reuse of the data (but open to discuss on request)</p> <p>All focus group participants will be coded (by using pseudonyms) in the processing and reporting of the research results. This means that real names will not be associated in any way with the information collected or with the research findings from this study.</p> <p>Aggregated and pseudonymised research findings will be discussed in scientific research publications.</p> <p>Only focus group participants and expert interviewees who sign the informed consent statement at the start of the interview will participate. By signing this form, they give permission for the use and disclosure of pseudonymised information for scientific purposes of this study at any time in the future and for the audio-recording of the workshop only for post-processing purposes.</p> <p>Time period: M4-20</p>
Standards and metadata	There are no specific standards or metadata associated with these types of data.
Data sharing	Aggregated and pseudonymised data and research results will be published in project reporting documents (accessible to all consortium partners via the hackAIR project Wiki) and in external research publications from the project.
Archiving and preservation (including storage and backup)	Aggregated and pseudonymised data and research results will be saved on the hackAIR project wiki. This wiki is the central repository of internal information for the hackAIR project.
Discoverable	Not applicable
Accessible	Not applicable
Assessable and intelligible	Not applicable. In case of a report or paper submitted for publication with peer review, all research findings will be integrated into the report or paper. Datasets will never be added to the publication.
Useable beyond the original purpose for which it was collected	Only aggregated and anonymized research findings published in public reports or deliverables and published in academic papers, will be and stay accessible from the moment of publication.
Interoperable to specific quality standards	Not applicable.

4.6.3 Behavioral change survey data set

Data set reference and name	Data_WP6_3_Behavioral change survey data set
Data set description	The behavioural change survey data set will contain all relevant quantitative data for designing, running and analysing the behavioural change strategy.

D1.3: 1st Data Management Plan

	<p>Consortium partners involved: VUB</p> <p>The following data will be collected:</p> <ul style="list-style-type: none"> – Demographic information: only used for the profiling of the research participants – Self-reported data on the factors that measure behavioral change (use of standardized scales) – Contact information: Only collected and used for recruiting purposes <p>Data collection method: 2 online surveys (1 pre-test and 1 post-test)</p> <p>Data subjects:</p> <ul style="list-style-type: none"> – Amount: 100 <p>Data type: Data type: SPSS format .sav /MS Office .xls format</p> <p>Reuse of the anonymized dataset</p> <p>Anonymous and aggregated research data will be published in internal project reports (accessible to all consortium partners via the hackAIR project Wiki) and in external scientific research publications from the project.</p> <p>Time period: M4-20</p>
<p>Standards and metadata</p>	<p>Aggregated and anonymized data and research results will be published in project reporting documents (accessible to all consortium partners via the hackAIR project Wiki) and in external research publications from the project.</p> <p>The anonymized dataset can be provided upon request.</p>
<p>Data sharing</p>	<p>Aggregated and anonymized data and research results will be saved on the hackAIR project wiki. This wiki is the central repository of internal information for the hackAIR project.</p>
<p>Archiving and preservation (including storage and backup)</p>	<p>Not applicable</p>
<p>Discoverable</p>	<p>Not applicable</p>
<p>Accessible</p>	<p>In case of a report or paper submitted for publication with peer review, aggregated and anonymized research findings will be integrated into the report or paper. In the contexts of scientific scrutiny and peer review, the full anonymized dataset can be made accessible to third parties.</p>
<p>Assessable and intelligible</p>	<p>Only aggregated and anonymized research findings published in public reports or deliverables and published in academic papers, will be and stay accessible from the moment of publication.</p>
<p>Useable beyond the original purpose for which it was collected</p>	<p>Not applicable.</p>
<p>Interoperable to specific quality standards</p>	<p>Not applicable.</p>

4.6.4 Behavioral change qualitative insights data set

Data set reference and name	Data_WP6_4_Behavioral change qualitative insights data set
Data set description	<p>The behavioural change data set will contain all relevant qualitative data for designing, running and analysing the behavioural change strategy.</p> <p>Consortium partners involved: VUB</p> <p>The following data will be collected:</p> <ul style="list-style-type: none"> – Demographic information: only used for the profiling of the research participants – Insights into facilitators and barriers for behavioral change – Contact information: Only collected and used for recruiting purposes <p>Data collection method: focus group</p> <p>Data subjects:</p> <ul style="list-style-type: none"> – Amount: 10 to 20 test subjects – Users who have successfully been engaged into using hackAIR platform during the pilot tests (identified in WP7). <p>Data type:</p> <ul style="list-style-type: none"> – Text – Audio-records: The focus groups and interviews might be audio-recorded for post-processing. This tape will be used by the involved researchers only for the processing of the workshop findings. It will only serve research purposes and it will by no means be released to other persons. <p>No reuse of the data (but open to discuss on request)</p> <p>All focus group participants will be coded (by using pseudonyms) in the processing and reporting of the research results. This means that real names will not be associated in any way with the information collected or with the research findings from this study.</p> <p>Aggregated and pseudonymised research findings will be discussed in scientific research publications.</p> <p>Only focus group participants and expert interviewees who sign the informed consent statement at the start of the interview will participate. By signing this form, they give permission for the use and disclosure of pseudonymised information for scientific purposes of this study at any time in the future and for the audio-recording of the workshop only for post-processing purposes.</p> <p>Time period: M4-20</p>
Standards and metadata	There are no specific standards or metadata associated with these types of data.
Data sharing	Aggregated and pseudonymised data and research results will be published in project reporting documents (accessible to all consortium partners via the hackAIR project Wiki) and in external research publications from the project.
Archiving and preservation (including storage and backup)	Aggregated and pseudonymised data and research results will be saved on the hackAIR project wiki. This wiki is the central repository of internal information for the hackAIR project.
Discoverable	Not applicable

D1.3: 1st Data Management Plan

Accessible	Not applicable
Assessable and intelligible	Not applicable. In case of a report or paper submitted for publication with peer review, all research findings will be integrated into the report or paper. Datasets will never be added to the publication.
Useable beyond the original purpose for which it was collected	Only aggregated and anonymized research findings published in public reports or deliverables and published in academic papers, will be and stay accessible from the moment of publication.
Interoperable to specific quality standards	Not applicable.

4.7 Datasets in WP7 - Pilot operation and evaluation (NILU)

As for the 1st hackAIR DMP WP7 generates research data. For the purposes of WP7 the following five datasets will be generated:

- PM concentrations by PSOC system
- PM concentrations by Arduino sensor
- PM concentrations by COTS system
- Content generated through the hackAIR platform
- Photos uploaded through the hackAIR mobile app

4.7.1 PM concentrations by PSOC system

Data set reference and name	Data_WP7_1_PM_concentrations_by_PSOC_system
Data set description	PM measurements by PSOC sensors
Standards and metadata	Values will be provided in $\mu\text{g}/\text{m}^3$ and/or in five levels (low, mid-low, middle, mid-high, high). For each value, the following metadata will be provided: Date, time, location (using the GPS unit of the smartphone), sensor id, and user id.
Data sharing	The data (PM measurements and metadata) will be available to other Researchers through the Open Access to Research Data Pilot. The measurements and metadata will be made available for use by the hackAIR applications through the secure API that we will create.
Archiving and preservation (including storage and backup)	The data along with the metadata will be stored securely in the hackAIR database and/or storage system. Database and storage areas will be set to take differential daily backups and a complete weekly one keeping up to the three last weeks.
Discoverable	These measurements will be discoverable through the hackAIR platform, only for users who are registered through the platform.
Accessible	The data along with the metadata will be stored securely in the hackAIR database and/or storage system. Database and storage areas will be set to take differential daily backups and a complete weekly one keeping up to the three last weeks.
Assessable and intelligible	Sufficient documentation will be provided that allows assessment and review by third parties.

D1.3: 1st Data Management Plan

Useable beyond the original purpose for which it was collected	The dataset will be saved for a long time after gathering them for statistical reasons and will be available to other research groups upon request. Measurements can be shared with other researchers along with the geospatial information and time they were made. Metadata will not contain the information of who contributed it to hackAIR.
Interoperable to specific quality standards	Anyone who wants to have access to the measurements and the metadata will be able to use them as a whole or a subset of it as it fits to his/her purpose.

4.7.2 PM concentrations by Arduino sensor

Data set reference and name	Data_WP7_2_PM_concentrations_by_Arduino_sensor
Data set description	PM measurements by Arduino sensors
Standards and metadata	Values will be provided in µg/m ³ and/or in five levels (low, mid-low, middle, mid-high, high). For each value, the following metadata will be provided: Date, time, location (using user's profile geolocation details), sensor id, and user id.
Data sharing	The data (PM measurements and metadata) will be available to other Researchers through the Open Access to Research Data Pilot. The measurements and metadata will be made available for use by the hackAIR applications through the secure API that we will create.
Archiving and preservation (including storage and backup)	The data along with the metadata will be stored securely in the hackAIR database and/or storage system. Database and storage areas will be set to take differential daily backups and a complete weekly one keeping up to the three last weeks.
Discoverable	These measurements will be discoverable through the hackAIR platform, only for users who are registered through the platform.
Accessible	The data along with the metadata will be stored securely in the hackAIR database and/or storage system. Database and storage areas will be set to take differential daily backups and a complete weekly one keeping up to the three last weeks.
Assessable and intelligible	Sufficient documentation will be provided that allows assessment and review by third parties.
Useable beyond the original purpose for which it was collected	The dataset will be saved for a long time after gathering them for statistical reasons and will be available to other research groups upon request. Measurements can be shared with other researchers along with the geospatial information and time they were made. Metadata will not contain the information of who contributed it to hackAIR.
Interoperable to specific quality standards	Anyone who wants to have access to the measurements and the metadata will be able to use them as a whole or a subset of it as it fits to his/her purpose.

4.7.3 PM concentrations by COTS system

Data set reference and name	Data_WP7_3_PM_concentrations_by_COTS_system
-----------------------------	---

D1.3: 1st Data Management Plan

Data set description	PM measurements by COTS sensors
Standards and metadata	Values will be provided in five levels (low, mid-low, middle, mid-high, high). For each of the levels, the following metadata will be provided: Date, time, location (using the GPS unit of the smartphone), sensor id, and user id.
Data sharing	The data (PM measurements and metadata) will be available to other Researchers through the Open Access to Research Data Pilot. The measurements and metadata will be made available for use by the hackAIR applications through the secure API that we will create.
Archiving and preservation (including storage and backup)	The data along with the metadata will be stored securely in the hackAIR database and/or storage system. Database and storage areas will be set to take differential daily backups and a complete weekly one keeping up to the three last weeks.
Discoverable	These measurements will be discoverable through the hackAIR platform, only for users who are registered through the platform.
Accessible	The data along with the metadata will be stored securely in the hackAIR database and/or storage system. Database and storage areas will be set to take differential daily backups and a complete weekly one keeping up to the three last weeks.
Assessable and intelligible	Sufficient documentation will be provided that allows assessment and review by third parties.
Useable beyond the original purpose for which it was collected	The dataset will be saved for a long time after gathering them for statistical reasons and will be available to other research groups upon request. Measurements can be shared with other researchers along with the geospatial information and time they were made. Metadata will not contain the information of who contributed it to hackAIR.
Interoperable to specific quality standards	Anyone who wants to have access to the measurements and the metadata will be able to use them as a whole or a subset of it as it fits to his/her purpose.

4.7.4 Content generated through the hackAIR platform

Data set reference and name	Data_WP7_4_Content_generated_through_the_hackAIR_platform
Data set description	The data that is generated by the users of the hackAIR platform like users' personal information, user perception of the atmosphere on a given day, posts, comments, detailed log of user actions (login, logout, account creation, visits on specific parts of the app), information on user devices (which browsers and mobile phones) and number of mobile downloads.
Standards and metadata	As part of any stored data, meaningful metadata (time and date of posts and comments, owner, dates for the logs and the user perception statement) will be generated to assist the discoverability of the data and related information. Any personal data will be anonymized and encrypted. The following standards will be used: <ul style="list-style-type: none"> • RSA for generating public keys

D1.3: 1st Data Management Plan

	<ul style="list-style-type: none"> • AES for private data encryption • SHA hashes for storing passwords
Data sharing	<p>hackAIR users will be able to view all comments and posts they and fellow users have created.</p> <p>Data will be available for use by the hackAIR applications through the secure API that we will create.</p>
Archiving and preservation (including storage and backup)	<p>Data will be saved on the hackAIR database server. Whenever updates need to be performed, the old data will be overwritten and a detailed audit action will be kept in the log file containing the changed text for security reasons.</p> <p>Database will be set to take differential daily backups and a complete weekly one keeping up to the three last weeks.</p>
Discoverable	<p>Discoverability will be possible only for the administrator of the app for all the data and for some of the data (user name, anonymous contributed user perception of the atmosphere, posts and comments) only for registered users.</p> <p>The database will not be discoverable to other network machines operating on the same LAN, VLAN with the DB server or other networks. Therefore only users with access to the server (hackAIR technical team members) will be able to discover the database.</p>
Accessible	<p>Some of these data (posts, comments, users' display names and users' perception of the atmosphere) will be accessible through the hackAIR platform, only for users who are registered through the platform.</p> <p>On the technical level only authorized hackAIR technical team members will have access to the database.</p>
Assessable and intelligible	N/A
Useable beyond the original purpose for which it was collected	Specific logging data will be used to identify the level of participation of a specific group of test users to measure behavioral change (WP6 T6.2).
Interoperable to specific quality standards	N/A

4.7.5 Photos uploaded through the hackAIR mobile app

Data set reference and name	Data_WP7_5_Photos_uploaded_through_the_hackAIR_mobile_app
Data set description	hackAIR users will be asked to upload pictures depicting the sky. These pictures will be geotagged and time stamped and will be stored in the hackAIR DB or a hackAIR designated secure network storage area. They will be used to calculate the AOD for the specific location.
Standards and metadata	As part of any stored data, meaningful metadata (image URL, timestamp and location) will be generated to help internal users locate the data and related information. Those data will be saved in the hackAIR database.
Data sharing	<p>The data (metadata and photos) will be available to other Researchers through the Open Access to Research Data Pilot.</p> <p>The photos and metadata will be made available for use by the hackAIR applications</p>

D1.3: 1st Data Management Plan

	<p>through the secure API that we will create. For cases that there will need to be a mass use of photos (e.g. for conversion reasons) applications will be granted the right to read only from the storage area and bypass the API.</p> <p>Discoverability will be possible only for the administrator of the app for all the data. Some of the images might be posted to social media networks like Instagram only if the user authorizes that action.</p>
Archiving and preservation (including storage and backup)	<p>The photos along with the metadata will be stored securely in the hackAIR database and/or storage system. The photos might be resized and changed in format in order to reduce storage but also be able to be reused by hackAIR if needed.</p> <p>Database and storage areas will be set to take differential daily backups and a complete weekly one keeping up to the three last weeks.</p>
Discoverable	<p>The database will not be discoverable to other network machines operating on the same LAN, VLAN with the DB server or other networks. Therefore only users with access to the server (hackAIR technical team members) will be able to discover the database.</p> <p>The images folder will not be discoverable by systems or persons in the same or other servers in the same LAN/VLAN as the storage/database server.</p>
Accessible	<p>These images might (a pending decision) be accessible through the hackAIR platform, only for users who are registered through the platform.</p> <p>On the technical level only authorized hackAIR technical team members will have access to the database and database storage.</p>
Assessable and intelligible	N/A
Useable beyond the original purpose for which it was collected	<p>The dataset will be saved for a long time after gathering them for statistical reasons and will be available to other research groups upon request. Images can be shared with other researchers along with the geospatial information and timestamp the images were taken at. Images will not contain the information of who contributed it to hackAIR.</p>
Interoperable to specific quality standards	<p>Anyone who wants to have access to the photos and the metadata will be able to use them as a whole or a subset of it as it fits to his/her purpose.</p>

4.8 Datasets in WP8 - Communication, Dissemination and Exploitation (ONSUB)

Work package 8 (Communication, Dissemination and Exploitation) does not generate research data. For the purposes of implementing and monitoring the communications strategy, the work package leader manages the following two datasets (described in detail below):

- Contact database
- Communications monitoring data

In addition, the work package leader is gathering information from project partners to develop the two strategies included in the work package (D8.1 Communication and Dissemination Strategy and D8.2/D8.8 Sustainability and Exploitation Strategy) through interviews, workshops and questionnaires. The results of this process are included directly in the strategy documents and not described further in this document.

4.8.1 Contact database

Data set reference and name	Data_WP8_1_Contact database
Data set description	The database contains name, organisation and contact details for all relevant contacts of the project. This includes members of the network of interest, related initiatives, participants of events, recipients of the newsletter etc.
Standards and metadata	<p>The data is stored in a simple table, with the following fields:</p> <ul style="list-style-type: none"> - Name - Category (1: org health/environment, 2: research/monitoring, 3: data providers) - Short description - Location - Link - Core group? (yes, possibly, no) - Similar project? (yes, no) - Contact person (only visible for team) - E-Mail (only visible for team) - hackAIR partner - Comments <p>Additional fields will be added as the project progresses.</p>
Data sharing	The data is confidential for internal use within the consortium (as personal data is involved). The contact database is managed in a Google Spreadsheet, and linked from the project internal wiki.
Archiving and preservation (including storage and backup)	The data is collected for internal use in the project, and not intended for long-term preservation. The work package leader is keeping a quarterly backup on a separate disk.
Discoverable	No
Accessible	No
Assessable and intelligible	No
Useable beyond the original purpose for which it was collected	No
Interoperable to specific quality standards	CSV

4.8.2 Communications monitoring data

Data set reference and name	Data_WP8_2_Communications monitoring data
Data set description	The database contains quarterly information on key communications indicators, including the number of stakeholders on contact list, visits to project website, newsletters, social media impressions, media impressions and events. Data sources include: automatic monitoring of Google Alerts, Twitter Analytics, Piwik and partner reports.

D1.3: 1st Data Management Plan

Standards and metadata	The data is stored in Google Spreadsheets, with a separate table for each source of data. The format for each entry is: <ul style="list-style-type: none">- ID- Date- Description- Partner- Estimated Number of People Reached- Link (if applicable)
Data sharing	The raw data is internal to the work package leader (as it contains personal information); a summary will be made available to partners as part of the biannual internal reports.
Archiving and preservation (including storage and backup)	The data is collected for internal use in the project, and not intended for long-term preservation. The work package leader is keeping a quarterly backup on a separate disk.
Discoverable	No
Accessible	No
Assessable and intelligible	No
Useable beyond the original purpose for which it was collected	No
Interoperable to specific quality standards	CSV