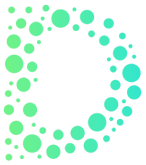

D-NOSES

Distributed Network for Odour Sensing,
Empowerment and Sustainability

Data Management Plan

D1.6 v1.0

SEPTEMBER 2018

Data Management Plan

PROJECT ACRONYM	GRANT AGREEMENT #	PROJECT TITLE
D-NOSES	789315	D-NOSES

DELIVERABLE TITLE

D1.6 Data Management Plan v1.0

AUTHORS

Francisco Sanz
Ibercivis

Louise Francis
Mapping for Change

Rosa Arias
Ibercivis

Funded by the Horizon 2020 programme of the European Union

Grant Agreement No 789315

DISSEMINATION LEVEL

✔ **P** **Public**

P Confidential, only for members of the consortium and the Commission Services

Summary

This report presents the first version of the Data Management Plan (DMP)

The objective of this report is to present the first version of the Data Management Plan for the D-NOSES Project. The main goal of the Data Management Plan is to provide an analysis of the main elements of the data management policy that the consortium will use. To create this DMP we are following the guidelines provided by the European Commission on FAIR Data Management in Horizon 2020[1] , some aspects related to the conclusions generated by the JRC Technical Report *Survey report: data management in Citizen Science Projects*[2] and others documents collected in the references.

[1] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[2] ec.europa.eu/newsroom/document.cfm?doc_id=19684

Index

SUMMARY

1 Introduction

The D-NOSES project, funded under the topic *H2020-SwafS-23-2017 Responsible Research and Innovation (RRI) in support of sustainability and governance, taking account of the international context*, will reverse the way in which odour pollution is commonly tackled. It will empower citizens to become a driving force for change through RRI, citizen science and co-creation tools to map and measure the problem, and co-design solutions with key quadruple helix stakeholders.

D-NOSES aims to kickstart a much needed collaborative journey to tackle the problem of odours at a global scale by developing coordinated local case studies in 10 European and non-European countries (pilots). Several project actions will guarantee a high impact and project sustainability. With the aim of situating odour pollution in the map, the International Odour Observatory (IOO) will be created to promote engagement and public participation. In the IOO, all relevant data and information will be gathered, mapped and made available, granting access to information to allow for the implementation of Principle 10 of Rio Declaration. The App OdourCollect will also be used to gather odour observations from engaged citizens, meaning that citizens will not only have, for the first time, access to information in odour pollution, but will become data generators.

All this means that the data will be collected from different sources and different stakeholders, as described in *Deliverable 7.2 Project website, branding and templates*, including data collected by citizens.

The results of the D-NOSES project will improve the management of odour problems, after the validation of the proposed innovative, bottom-up methodology to monitor, for the first time, the real perception of nuisance in the impact area of odour emitting activities. The analysis of the results of each pilot (at least 10 pilots in at least 10 different countries) will be used to co-create DIY Guidelines for Project Replicability and, standard criteria for future odour regulations at different levels, together with the Green Paper and the Strategic Roadmap for Governance in odour pollution, which will pave the way for capacity building and a improved governance.

D-NOSES, in its core, is a Citizen Science project. As such, there are currently no standards in data or metadata that have been released yet, although there are some initiatives leaded by CSA[3] or the *Working Group #5 of the Citizen Science COST Action CA15212*, lead by the European Citizen Science Association (ECSA[4] [5]), a partner of D-NOSES with which Ibercivis collaborates closely to further develop the above mentioned standards, working on this. These first steps are following open standards; for example: "*WG5's specific objective for the second period (1.5.2017-30.4.2018) is to contribute to develop an ontology of citizen-science projects (including a vocabulary of concepts and metadata) to support data sharing among citizen-science projects. WG5 will coordinate with activities on data and service interoperability carried out in Europe, Australia and the USA (e.g., the CSA's international Data and Metadata Working Group [http://citizenscience.org/ association/about/working-groups/]), and will take into account existing standards, namely Open Geospatial Consortium (OGC) standards (via the OGC Domain Working Group on Citizen Science), ISO/TC 211, W3C standards (semantic sensor network/Linked Data), and existing GEO/GEOSS semantic interoperability. WG5 will investigate the best format to publish the ontology.*"[6]

Partial results on how to manage data or metadata in Citizen Science projects in a FAIR way have been produced by those initiatives, which will be used in D-NOSES where possible. The outcome of our experience in producing, validating and managing citizen science data will be reported to the WGs of the Citizen Science COST Action to contribute and improve the work already done.

[3] https://www.wilsoncenter.org/sites/default/files/wilson_171204_meta_data_f2.pdf
[4] https://www.cs-eu.net/sites/default/files/media/2018/04/2018.03%20WG%20meeting%20in%20Milan%20%28COST%20Action%20CA%2015212%29%20-%20minutes.pdf
[5] https://www.cs-eu.net/sites/default/files/media/2018/06/COST-WG5-GenevaDeclaration-Report-2018.pdf
[6] https://www.cs-eu.net/sites/default/files/media/2018/04/2018.03%20WG%20meeting%20in%20Milan%20%28COST%20Action%20CA%2015212%29%20-%20minutes.pdf

2
Data Summary

Within the D-NOSES project there will be two main sources of data. The first one, directly produced by the consortium, will consist mainly of documentation relevant to methodologies, data analysis, metadata definitions, etc. This type of data will be presented under free licenses whenever possible such as Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)[7]. This data will be available under the D-NOSES main web page[8].

On the other hand, some of the data will be generated by citizens reporting odour episodes through our app. Right now, a new version of the app is under development, including a potent back office for validation purposes of the data gathered in the pilot case studies. In this document we will refer to the current - legacy - version of the app - OdourCollect, which was developed in 2016 after receiving funding in the context of the *MyGeoss Project - Applications for your Environment*, from the Joint Research Center[9] . This data is stored in an SQL database and can be downloaded in an anonymized way under CC BY-SA 4.0 License. The tables in annex I describe this dataset.

In addition, the Community Maps platform will enable citizens to map cases where they are affected by odour issues in their communities and other information deemed relevant for the different pilots. Community Maps supports constructing digital representations of physical space through participatory action. Its map interface provides a way in which to add new data as well as editing and deleting existing data. Community Maps is a single-page front-end application built on top of GeoKey to which it connects via the public API. Is it able to retrieve and store public and private information that is visualised onto the map. If private information is to be used, OAuth2 authentication is required to authorise the user. At the core of the back-end is a postgreSQL relational database system with geospatial capabilities that stores all information relevant to run the platform.

[7] https://creativecommons.org/licenses/by-sa/4.0/?
[8] dnoses.eu
[9] http://digitalearthlab.jrc.ec.europa.eu/app/odourcollect

3
Fair Data

The term **FAIR** was launched at a Lorentz Workshop in 2014, and the resulting principles were published in 2016[10]. The term FAIR describes a set of guiding principles to make data **Findable, Accessible, Interoperable, and Reusable**. We will follow the guidelines described below, but we are aware that, as stated in the *H2020 Programme Guidelines on Fair Data Management in Horizon 2020*[11], "*participating in the ORD Pilot does not necessarily mean opening up all your research data. Rather ORD pilot follows the principle "as open as possible, as closed as necessary" and focuses on encouraging sound data management as an essential part of research best practice*".

The current version of this deliverable reflects the D-NOSES Data Management Plan as designed at this stage of the Project. It has to be taken into account that we are still in the process of developing some of the project tools, such as the International Odour Observatory, where Community Maps will be integrated, and the new version of the App OdourCollect. We will be defining further issues in relation to data management, both in terms of openness and data/metadata ontologies, and updates on the Data Management Plan will be provided as new versions of the current deliverable, always guaranteeing the the project data is FAIR. It is foreseen to have the final version of the deliverable before the first Reporting Period, once all the project tools will be created, running and validated.

DATA FINDABLE

The concept of findability refers to the ability to locate information by other users, it means that we will provide the necessary metadata to help in the identification of the different datasets generated in each pilot and those provided by citizens outside these pilots. As provided by the open document *The FAIR Guiding Principles for scientific data management and stewardship*[12], published by Mark D. Wilkinshon et al. in Nature, we will, where possible:

- assign a globally unique and persistent identifier to (meta)data
- describe data with rich metadata

[10] https://www.force11.org/group/fairgroup/fairprinciples
[11] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
[12] https://www.nature.com/articles/sdata201618#bx2

- include clearly and explicitly in the metadata the identifier of the data it describes
- register on index (meta)data in a searchable resource

When possible, the data will be stored in SQL database, anonymized and linked to the web page. It will be downloadable openly in csv format during the life of the project. Periodically, anonymized (meta)data will be uploaded to Zenodo, providing a DOI (Digital Object Identifier) for each dataset generated. Using DOI will allow us to *edit/update the record's files after they have been published*[13]

We will search for other datasets which can be used for the purposes of the project, such as meteorological data.

DATA ACCESSIBILITY

Four main tools will be used to provide access to the project data:

- The project web page (see more details on the structure and the contents on Deliverable 7.2)
- The International Odour Observatory (see more details on the structure and the contents on Deliverable 7.2)
- The OdourCollect mobile App, to generate collaborative odour maps
- The D-NOSES Community Mapping tools, which will integrate odour observations with other relevant project data and make it available online for public access.

As in the previous point, following D. Wilkinshon et al., we will follow the following rules where possible :

- (meta)data will be retrievable by their identifier using a standardized communication protocol.
    - the protocol is open, free and universally implementable
    - the protocol allows for an authentication and authorization procedure, where necessary
- metadata will be accessible, even when the data are no longer available.

DATA INTEROPERABILITY

As previously stated, D-NOSES biggest challenge in relation to data management is that data standards and / or metadata have not yet been defined in Citizen Science

[13] https://help.zenodo.org/

projects. However, we will follow the partial results that have come out of the above mentioned Working Groups of the Citizen Science COST Action. In particular, within D-NOSES, when possible:

- (meta)data will use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- (meta)data will use vocabularies that follow FAIR principles
- (meta)data will include qualified references to other (meta)data

DATA RE-USE

On a case by case basis, it will be agreed between all consortium partners when the data produced by the consortium and/or data produced by the engaged citizens will be licensed under Creative Commons International CC BY 4.0, with no embargo to enable re-use. Exceptions may occur in some of the pilots in relation to specific requirements of the different stakeholders in the quadruple helix for each country. In those cases, other re-use licenses will be adopted to fulfill all requirements. In particular, D-NOSES will follow the following guidelines, when possible:

- Meta(data) will be richly described with a plurality of accurate and relevant attributes
    - (meta)data will be released with a clear and accessible data usage license
    - (meta)data will be associated with detailed provenance
    - (meta)data will meet domain-relevant community standards[14]

[14] As previously mentioned, the Citizen Science Community is working right now in the definition of meta data and ontologies. The D-NOSES approach is to follow the partial outcomes from the Citizen Science COST Action Working Groups, when possible, and contribute to the definition of standards on Citizen Science metadata.

4
Allocation of Resources and Data Security

The consortium will use Ibercivis servers to store the data in a SQL database in a FAIR way. Regarding Community Maps/Geokey, data will be collected and stored on Mapping for Change servers, from where they will be pushed to the Ibercivis' server using the GK API. When required, data controller/data processing agreements will be established.

The data obtained during the project, when possible, will also be uploaded anonymized to the free-of-charge Zenodo repository. The handling of the local servers and Zenodo repository, as well as all data management issues related to the project, falls in the responsibility of the Coordinator. The data is guaranteed for 15 years on unfunded effort by Ibercivis.

Francisco Sanz, the Executive Director of Ibercivis, is the responsible for Data Management within the D-NOSES project, specifically for this deliverable D1.6, and also for the associated Ethics deliverables D8.1 (informed consent procedures for the identification and recruitment of research participants) and D8.2 (in relation to collection and processing of personal data). He will also take care of the revision of this document before M15 (v1.1) and M36 (v1.2). The PI of each partner will have the responsibility for implementing the data management plan in relation to the project actions. Each D-NOSES partner shall be responsible for following the policies described in this DMP.

The data will be stored on Ibercivis Foundation's servers, on hosts with RAID 1 hard disk system and daily backups. This guarantees its conservation for any eventuality arising.

5
Ethical Aspects

The D-NOSES consortium further confirms that each partner will check with their national legislation/practice and their local ethics committee that provides guidelines on data protection and privacy issues in terms of both data protection and research procedures in relation to any of the proposed public engagement and potential volunteer research activities. Any procedures for electronic data protection and privacy will conform to Directive (EU) 2016/680 and Regulation (EU)2016/679 on the protection of personal data, and its enactments in the national legislations.

Ethical approval for studies with volunteer participants (such as correlation of public feedback) will be sought from the University of Zaragoza Ethics Committees in line with institutional procedures at ECSA or UCL (the partners with extensive experience in volunteer research). There will be:

- No collection of data on a citizen without permission.
- Information will only be used for the purposes covered by agreement, and will not be retained except as required for these purposes.
- Information will not be made public or provided to third parties without explicit permission.
- Contractual and technical controls will be applied to prevent information becoming inadvertently available to third parties.

Informed consent will be obtained from any volunteer, especially those participating in WP5 with the provision of on-line forms supported by the necessary information for the individual to make a voluntary informed decision about whether or not to participate in any of the evaluation feedback sessions. Any electronic information collected and mined will be anonymised to prevent the identification of individual subjects unless express permission is granted.

More details on the Ethics requirements in relation to informed consent procedures and protection of personal data will be provided in deliverables 8.1 and 8.2.

6
Other

Each partner will provide the Project Coordinator copies of opinion or confirmation by the competent Institutional Data Protection Officer and/or authorization or notification by the National Data Protection Authority must be submitted (which ever applies according to the Data Protection Directive and the national law). As focused in the JRC document *Survey report: data management in Citizen Science Projects*, we will pay attention not only to legal aspects regarding different countries but also to cultural aspects. We will follow also legislation on personal data from GDPR (2016/679). Two deliverables - D8.1 and D8.2 - will cover all aspects related with GDPR.

7

Definitions, acronyms and abbreviations

CSV: Comma **S**eparated **V**alues is a text file that uses a comma to separate value

CSA: Citizen **S**cience **A**ssociation

DMP: Data Management Plan

D-NOSES: Distributed Network for Odour Sensing, Empowerment and Sustainability

DOI: Digital Object Identifier is a persistent identifier used to uniquely identify objects, standardized by the ISO

ECSA: European Citizen Science Association

FAIR: Research data that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution.

JRC: Joint Research Centre

Metadata: data that provides information about other data. Three types of metadata can be distinguished, including descriptive metadata, structural metadata and administrative metadata.

OGC: Open Geospatial Consortium

Open data: Research data that that can be freely used, re-used and redistributed by anyone for any purpose. Open data is free of restrictions from copyright, patents or other mechanisms of control.

PPSR_CORE: Citizen Science and **P**ublic **P**articipation in **S**cientific **R**esearch

RRI: Responsible research and Innovation

SQL: Structured **Q**uery **L**anguage is a domain-specific language used in programming and designed for managing data held in a relational database management system

W3C: World Wide Web Consortium

WP: Work package

8
References

Anne Bowser et al. (2017, Dec) *Citizen Science Association Data & Metadata Working Group*, retrieved from https://www.wilsoncenter.org/sites/default/files/wilson_171204_meta_data_f2.pdf

H2020 Programme (2016, 26th July) *Guidelines on FAIR Data Management in Horizon 2020*, retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Mark D. Wilkinson et al. (2016, 15th March) *The FAIR Guiding Principles for scientific data management and stewardship*, Scientific Data 3, Article number: 160018(2016), retrieved from https://www.nature.com/articles/sdata201618

Openaire Project, *Openaire Frequently Asked Questions*, retrieved from https://www.openaire.eu/support/faq

Luigi Ceccaroni et al.(2018, 6th June) *Minutes of WG5 workshop en Geneva: On the citizen-science ontology, standards & data*, retrieved from https://www.cs-eu.net/sites/default/files/media/2018/06/COST-WG5-GenevaDeclaration-Report-2018.pdf

Luigi Ceccaroni et al. (2018, 26th-27th March) *Minutes of WG5 workshop in Milan: Towards a new ontology of citizen science, retrieved from* https://www.cs-eu.net/sites/default/files/media/2018/04/2018.03%20WG%20meeting%20in%20Milan%20%28COST%20Action%20CA%2015212%29%20-%20minutes.pdf

Sven Schade et al. (2016) Survey report: data management in Citizen Science projects, retrieved from https://ec.europa.eu/newsroom/document.cfm?doc_id=19684

Annex I

Odour Collect table description.

User description		
Field	Type	Description
id	int(11)	Unique identifier
username	varchar(255)	User name
email	varchar(255)	Email of the user
signup_date	timestamp	Date of signup

Report description		
Field	Type	Description
id	int(11)	Unique identifier
type	int(11)	Type of odour
intensity	int(2)	Intensity of odour
annoyance	int(2)	annoyance of odour
cloud	int(2)	Is weather cloudy?
rain	int(2)	Is weather rainy?
wind	int(2)	Is weather windy?
origin	varchar(255)	Do you know origin of odour?
duration	varchar(255)	Do you know duration of odour?
pic	longblob	Can you upload a picture?
report_date	timestamp	Date of report
latlng	varchar(255)	Latitude + Longitude

latitude	decimal(10,8)	Latitude
longitude	decimal(11,8)	Longitude
user_id	int(11)	User doing the report

Comment		
Field	Type	Description
id	int(11)	Unique identifier
comment	varchar(255)	Comment to the episode
comment_date	timestamp	Date of comment
user_id	int(11)	User doing the comment
report_id	int(11)	Id of the report commented