

Grant agreement no. 709443

DITOs

Doing It Together science

Coordination & Support Action

## D6.7 Final Data Management Plan

Work Package: 6

Due date of deliverable: Month 36

Actual submission date: 29/05/2019

Start date of project: 01/06/2016

Duration: 36 months

Lead beneficiary for this deliverable: UCL

Contributors: Meritum, Waag, UNIGE, UPD, RBINS

Reviewer: Myriam Fellous-Sigrist, Research Data Support Officer, UCL

Project co-funded by the European Commission within the H2020 Programme (2014-2020)		
PU	Public	X
CO	Confidential, only for members of the consortium (including the Commission Services)	
EU-RES	Classified Information: RESTREINT UE (Commission Decision 2005/444/EC)	
EU-CON	Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)	
EU-SEC	EU-SEC Classified Information: SECRET UE (Commission Decision 2005/444/EC)	



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 709443

## Disclaimer

The content of this deliverable does not reflect the official opinion of the European Union. Responsibility for the information and views expressed herein lies entirely with the author(s).

All 'Doing It Together science' (DITOs) consortium members are also committed to publish accurate and up to date information and take the greatest care to do so. However, the DITOs consortium members cannot accept liability for any inaccuracies or omissions nor do they accept liability for any direct, indirect, special, consequential or other losses or damages of any kind arising out of the use of this information.

## Copyright Notice



This work by Parties of the DITOs Consortium is licensed under a Creative Commons Attribution 4.0 International Licence (<http://creativecommons.org/licenses/by/4.0/>).

## Acknowledgement



The DITOs project has received funding from the European Union Horizon 2020 programme under grant number 709443.

## Reference

Please cite this work as:

DITOs Consortium, 2019. *Doing It Together science: D6.7 Final Data Management Plan* UCL, London.

## Document Identification Sheet

Project ref. no.	709443
Project acronym	DITOs
Project full title	Doing It Together Science
Document name	DITOs-D6.7-20190528.pdf
Security (distribution level)	Public
Contractual date of delivery	Month 36, 31.05.2019
Actual date of delivery	28.05.19
Deliverable number	D6.7
Deliverable name	Final Data Management Plan
Type	Report
Status & version	Version 1.0
Number of pages	28
WP / Task (responsible)	WP6 / T6.2 (UCL)
Author(s)	UCL: Judy Barrett
Other contributors	UCL: Muki Haklay, UPD: Imane Biaz UNIGE: Elisa Radosta Meritum: Pawel Wyszomirski RBINS: Carole Paleco WS: Tamar ter Steege
Project Officer	Colombe Warin
Abstract	Final data management plan, including curation and maintenance beyond the lifetime of the project.
Keywords	citizen science, DITOs, data management, metadata, H2020
Sent to peer reviewer	06/03/2019
Peer review completed	01/05/2019
Supervisory Board approval version 1.0	28/05/2019

## Table of Contents

1	Version Log .....	5
2	Definitions and Acronyms .....	5
3	Executive Summary .....	7
4	Introduction .....	9
4.1	Project Background & Current Status .....	9
5	Data Management Plan .....	10
5.1	Datasets Review .....	10
5.1.1	DS1_DITOs_BioDesign_Events_Evaluation .....	11
5.1.2	DS2_DITOs_Environmental_Sustainability_Events_Evaluation .....	11
5.1.3	DS3_DITOs_Capacity_Building_Tools_Results .....	11
5.1.4	DS4_DITOs_Policy_Engagement_RRI .....	12
5.1.5	DS5_DITOs_Evaluation .....	13
5.1.6	DS6_DITOs_Business_Model_Innovation_Plan .....	13
5.1.7	DS7_DITOs_Involves .....	14
5.1.8	DS8_DITOs_Promotional_Material .....	14
5.2	FAIR Data .....	14
5.2.1	Making Data findable, including Provisions for Metadata .....	14
5.2.2	Making Data openly accessible .....	15
5.2.3	Making Data interoperable .....	16
5.2.4	Increase Data Re-Use (through clarifying Licences) .....	16
5.3	Allocation of Resources .....	17
5.4	Data Security .....	17
5.5	Ethical Aspects .....	17
6	Conclusion .....	17
7	References .....	18
1	APPENDIX 1 – Citizen Science Data Curation .....	19
1.1	Introduction .....	19
1.2	Study 1 – RBINS Xperibird, Belgium .....	19
1.3	Study 2 – RBINS Leopold Park Bioblitz, Belgium .....	20
1.4	Study 3 – Meritum Air Quality Science Café .....	21
1.5	Case Study 4 – Medialab Prado Interactivos? .....	21
1.6	Case Study 5 – Into the Night .....	22
1.7	Case Study 6 – UPD i4Gamer .....	23

1.8	Recommendations for Data Collected at Citizen Science Events.....	24
2	APPENDIX 2 - Metadata to be recorded against each Dataset uploaded to Zenodo .....	24
3	APPENDIX 3 – Deletion Procedures .....	25
4	APPENDIX 4 – Anonymisation Procedures .....	27
5	APPENDIX 5 – Creative Commons Licence .....	28

## Table of Figures

<b>Figure 1</b>	Data management process (EUROPEAN COMMISSION Directorate-General for Research & Innovation, 2017).....	7
<b>Figure 2</b>	DITOs data curation and retention process.....	8
<b>Figure 3</b>	Xperibird Interactive Map .....	19
<b>Figure 4</b>	XperiBIRD.be data available via Website.....	20
<b>Figure 5</b>	Citizen Science Air Quality Data Stored on Thingspeak .....	21
<b>Figure 6</b>	- Medialab Prado project page created by participants .....	22
<b>Figure 7</b>	Globe at Night App Recording.....	22
<b>Figure 8</b>	UPD, iGAMER 2016 - Photo credits: Yves Ininahazwe .....	23

## 1 Version Log

Version	Date	Released by	Nature of Change
DRAFT v1.0	06/03/19	Judy Barrett (UCL)	Initial Draft for peer review
DRAFTv1.1	17/04/19	Judy Barrett (UCL)	Reviewer's and Muki Haklay's comments included
DRAFT v1.2	1/5/19	Judy Barrett (UCL)	Comments from Imane Biaz, Elisa Radosta, Pawel Wyszomirski i
1.0	28/5/19	Judy Barrett (UCL)	Comments from Carole Paleco

## 2 Definitions and Acronyms

Acronyms	Definitions
CA	Consortium Agreement
CC	Creative Commons
CSA	Coordination and Support Action
Data	Information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of

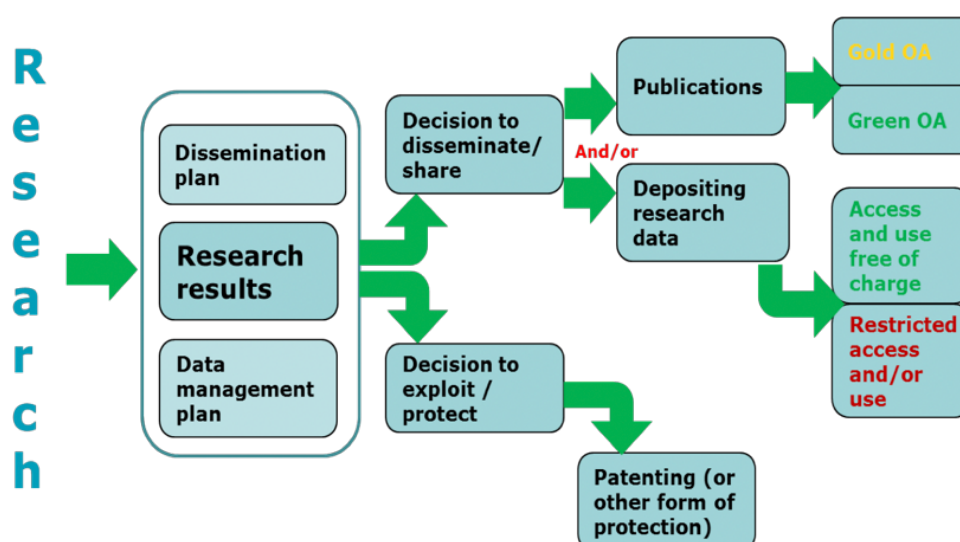
	data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form. (European Commission, 2016)
Dataset	A grouping of data
Digital Curation	Selection, preservation, maintenance and archiving of electronically stored data
DITOs	Doing It Together science
DMP	Data Management Plan
DOI	Digital Object Identifier
DS	Data Set
EC	European Commission
ECSA	European Citizen Science Association / Verein der Europäischen Bürgerwissenschaften
eutema	EUTEMA GMBH
FAIR	Findable, Accessible, Interoperable and Reusable
GA	Grant Agreement
H2020	Horizon 2020 Programme
IPR	Intellectual Property Rights
KI	Kersnikova Institute
Meritum	Centrum Szkolen i Rozwoju Osobistego Meritum
Metadata	A description of data
MP	Medialab Prado, Madrid
Open Access	Access that is free to all and free of any restrictions
Open Data	Data that can be freely used, shared and built on by anyone for any purpose
OpenAIRE	Open Access Infrastructure for Research in Europe
PPSR	Public Participation in Scientific Research
RBINS	Institut Royal des Sciences Naturelles de Belgique
Repository	A location in which data is stored or managed
RRI	Responsible Research and Innovation
Tekiu	Tekiu Limited
UCL	University College London
UNIGE	Universite de Geneve
UPD	Universite Paris Descartes
WS	Waag

### 3 Executive Summary

The European Commission 2017 defines data as “... information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation..... examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images.”

(EUROPEAN COMMISSION Directorate-General for Research & Innovation, 2017)

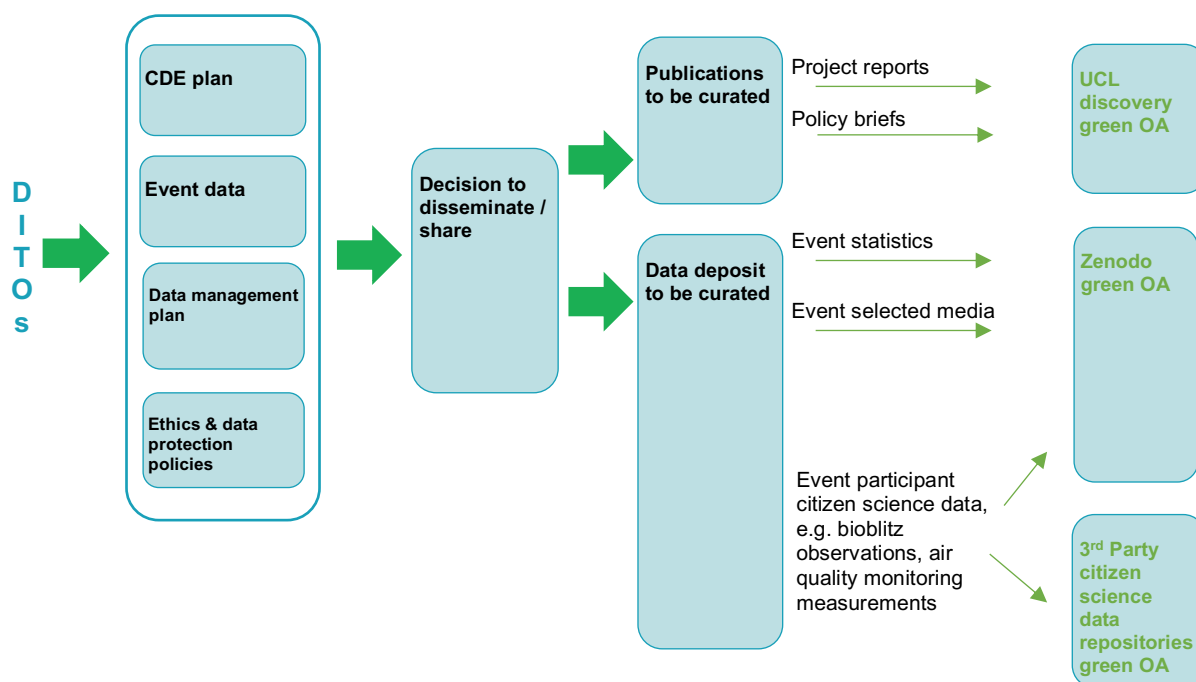
The following diagram illustrates this process:



**Figure 1** Data management process (EUROPEAN COMMISSION Directorate-General for Research & Innovation, 2017)

The primary role of DITOs is as a Coordination and Support Action rather than a research project, and as such it is not aimed to produce a set of data that has been analysed and studied by the project’s consortium. However, DITOs activities have generated some datasets which may be of potential interest to future research, so a data management plan is relevant.

The following diagram summarises the final DITOs data management process and curation strategy:



**Figure 2** DITOs data curation and retention process

As this diagram shows, project data will be stored in one of three ways:

- Project reports, papers, policy briefs and research insights will be stored in the UCL Discovery repository. Associated metadata will include attribution to the funded 'EU', project acronym 'DITOs', full name 'Doing It Together Science', and grant ID '709443' along with IRIS (UCL research information system) links to the UCL contributors, while associating deliverables with the consortium as a whole.
- Event data will be stored in a series of datasets in the OpenAire Zenodo repository under the Doing It Together Science label. Associated metadata will include 'EU', 'DITOs', '709443'
- Citizen Science project outputs, which are not a primary output of the CSA grant, but a 'by product' of the activities which contributed to the grant will be stored in established repositories such as github and the iNaturalist project database, or, if no such repository exists, will be stored in the OpenAire Zenodo repository under the Doing It Together Science project. Associated metadata will include 'EU', 'DITOs', '709443'. As agreed in the initial project data management plan, these elements are beyond the scope of the project; they are listed in APPENDIX 1 – Citizen Science Data Curation for completeness

#### Summary of results:

- The initial data management plan proved a satisfactory, workable solution for the CSA and it is recommended that future similar projects follow such procedures.
- The use of UCL Discovery for reports has proved very satisfactory in ensuring project outputs are 'FINDABLE' or searchable (See D6.8 CDE for examples of download statistics / dissemination records).
- There is a huge variance in the facilities available for the open storage and access of citizen scientists' observations / outputs. Further research into this is recommended (it was beyond the scope of the CSA).

The DITOs Data Management Plan (DMP) is Deliverable 6.7 (D6.7) from the coordination and support action (CSA), Doing It Together science (DITOs), grant agreement (GA) 709443.

## 4 Introduction

This deliverable is the final version of the DMP and complements D6.3, produced in M6 of the project. D6.3 contained details of dataset curation DURING the project; this document:

- identifies data to be retained beyond the life of the project; and
- defines the curation and maintenance of the data, including evaluation against the need for data to be findable, accessible, interoperable, and reusable (FAIR), in accordance with the Horizon 2020 Open Research Data pilot.

### 4.1 Project Background & Current Status

DITOs was a 3-year citizen science CSA (June 2016 – May 2019) which increased citizen science awareness and involvement by

- increasing the public's confidence in and level of engagement with science; and
- informing and encouraging policy makers to understand the benefits and challenges of citizen science

The project was a study of how to perform and evaluate citizen science outreach activities, with many of the deliverables being guidelines and proposals on how to run future projects and ensure that citizen science continues to grow and flourish.

This means that there were two categories of dataset within DITOs:

- those concerned with data on how to run citizen science activities (e.g. event attendance statistics and mailing lists) and also

- those concerned with data generated from the citizen science activities performed during the project (such as air quality monitoring measurements).

D6.3 (Initial Data Management Plan) defined the data that would be stored during the project and archived after the project, alongside associated processes, for the first category, but stated the following in respect of the second category:

*“The production of individual data plans for each individual citizen science project is beyond the scope of a CSA DMP. It was anticipated that some partners might identify ‘showcase’ projects for which data management of results may be evaluated, or schemas such as PPSR\_CORE (Public Participation in Scientific Research), currently under development by the Citizen Science Association, may be trialled, with appropriate feedback noted, but this is beyond the scope of the DITOs aims.”*

Consequently, this report focuses on the first category with the second category being outlined in APPENDIX 1 – Citizen Science Data Curation.

The initial DMP (D6.3) identified 8 datasets to be curated during the project, along with a proposed retention plan (see **Error! Reference source not found.**). It is confirmed that these datasets were successfully curated throughout the project in accordance with the D6.3 principles.

In accordance with Guidelines on FAIR Data management in Horizon 2020 (European Commission, 2016), this deliverable will:

- identify the DITOs datasets which should be retained for future research reference or input;
- review how these datasets will be made ‘FAIR’ (Findable, Accessible, Interoperable and Reusable) during their retention;
- define the allocation of resources (costs and responsibility) for data management after the project; and
- define procedures for data security (including data recovery as well as safe storage) during the project and for long term preservation.

## 5 Data Management Plan

### 5.1 Datasets Review

Each dataset of the eight datasets utilised during the project is considered below and its components reviewed for retention. In identifying datasets for retention, the overarching question has been whether the data will be of use for future research reference or input, so data robustness is of key importance. It should be noted that all data was gathered and recorded in accordance with the ethics and data protection deliverables of WP7. Deletion procedures for electronic data are specified in APPENDIX 3 – Deletion Procedures; paper documents will be securely shredded.

### 5.1.1 DS1\_DITOs\_BioDesign\_Events\_Evaluation

This dataset consisted of Biodesign event statistics and evaluations (qualitative and quantitative). The following elements were originally identified for post-DITOs curation.

Element	Post-Project Curation
Reports	Uploaded to UCL Discovery (green OA) as pdf files with appropriate metadata.
Anonymised participant satisfaction surveys	A selection may be included as an appendix to project reports (UCL Discovery), but the actual paper documents will be securely shredded.
Summaries of participant satisfaction	Uploaded to Zenodo (green OA) as csv files with appropriate metadata.
Event photographs	A selection may be included within project reports (UCL Discovery) but individual photographs will be deleted.

### 5.1.2 DS2\_DITOs\_Environmental\_Sustainability\_Events\_Evaluation

This dataset consisted of Environmental Sustainability event statistics and evaluations (qualitative and quantitative). The following elements were originally identified for post-DITOs curation.

Element	Post-Project Curation
Reports	Uploaded to UCL Discovery (green OA) as pdf files with appropriate metadata.
Anonymised participant satisfaction surveys	A selection may be included as an appendix to project reports (UCL Discovery), but the actual paper documents will be securely shredded.
Summaries of participant satisfaction	Uploaded to Zenodo (green OA) as csv files with appropriate metadata.
Event photographs	A selection may be included within project reports (UCL discovery) but individual photographs will be deleted.

### 5.1.3 DS3\_DITOs\_Capacity\_Building\_Tools\_Results

This dataset consisted of the project website (knowledge sharing platform) and its supporting tools (e.g. event mapping tools, discussion forums) and online engagement statistics, event database (project-wide) plus individual capacity-building event data.

D6.3 originally planned that the source code of the knowledge sharing platform and its online access statistics would be retained. However, this is not seen as being relevant to future research so this is no longer considered a data retention candidate. The underlying event database is now considered of interest to future researchers as it contains audience social profiles, so this will now be retained.

Element	Post-Project Curation
Event database	Uploaded to Zenodo (green OA) as csv files with appropriate metadata.
Knowledge database (from knowledge sharing platform)	Instructables from the science bus will be uploaded to Zenodo as creative commons pdf files.  Any general citizen science promotional material will be curated by ECSA for future use as appropriate (see D6.8 for details).  The Citizen Science Map will be hosted at Citizen cyberlab in Switzerland, as part of the Citizen Science Solution Kit.
Anonymised participant satisfaction surveys	A selection may be included as an appendix to project reports, but the actual paper documents will be securely shredded.
Summaries of participant satisfaction	Uploaded to Zenodo (green OA) as csv files with appropriate metadata.

#### 5.1.4 DS4\_DITOs\_Policy\_Engagement\_RRI

This dataset consisted of policy engagement event data (including discovery trips) and reports (including policy briefs / research insights)

Element	Post-Project Curation
Photographs and recordings	Photographs may be included in project reports but will not be kept individually.  No recordings have been identified by partners that will be 'research-worthy', so these will not be uploaded to Zenodo.  Any general citizen science promotional / informative videos, such as the Nuit de la Science 'what is citizen science?' video will be handed over to ECSA for future use as they see fit (see D6.8 for details).

Element	Post-Project Curation
Reports	Uploaded to UCL Discovery (green OA) as pdf files with appropriate metadata.
Anonymised participant satisfaction surveys	A selection may be included as an appendix to project reports, but the actual paper documents will be securely shredded.
Summaries of participant satisfaction	Uploaded to Zenodo (green OA) as csv files with appropriate metadata.

### 5.1.5 DS5\_DITOs\_Evaluation

This dataset consisted of evaluation data gathered from facilitator interviews and partner event statistics. The source interview data will be deleted; anonymised quotes or summaries may be included in deliverable reports but will not be retained other than within the reports.

Element	Post-Project Curation
Evaluation reports	Uploaded to UCL Discovery (green OA) as pdf files with appropriate metadata.
Evaluation summary data	Summary data will be derived from the events database, which will be stored as part of the DS3_DITOs_Capacity_Building_Tools_Results dataset, so no additional data will be stored here.

### 5.1.6 DS6\_DITOs\_Business\_Model\_Innovation\_Plan

This dataset consists of project outputs, such as deliverables. It is not research data but has been included in the plan for completeness of data.

Element	Post-Project Curation
CSA deliverables	Uploaded to UCL Discovery (green OA) as pdf files with appropriate metadata.
Academic publications	Uploaded to UCL Discovery (green OA) as pdf files with appropriate metadata.

Element	Post-Project Curation
Dissemination record	Uploaded to Zenodo (green OA) as csv file with appropriate metadata.

### 5.1.7 DS7\_DITOs\_Involvees

In the long term, no information from this dataset (which was consortium contact and project administrative data) will be retained. Organisational and legacy information may be retained for a short while under the curation of the PI after the project (for a maximum of 3 years), then it will be securely deleted.

### 5.1.8 DS8\_DITOs\_Promotional\_Material

This dataset consisted of event materials and outputs (such as Project Newsletters, blogs, tweets, social media posts, presentations, photographs, videos) which had been analysed/referred to by papers/research.

Following a review of project publications, papers and reports, no such material has been identified which is needed to further substantiate the material, so it is not proposed to archive any of this material. It may be retained for a short while for a short while under the curation of the PI after the project (for a maximum of 3 years), then it will be securely deleted.

## 5.2 FAIR Data

This section is written using the standard H2020 template for FAIR data given in Guidelines on FAIR Data management in Horizon 2020 (European Commission, 2016).

For ease of understanding and to ensure completeness of the H2020 DMP requirements, the template prompts are included below as questions (Q), with the answers (A) underneath.

### 5.2.1 Making Data findable, including Provisions for Metadata

Q: Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

A: All data uploaded to Zenodo will have an associated metadata document (stored as a .txt file) which describes key aspects of the data, as defined in APPENDIX 2 - Metadata to be recorded against each Dataset uploaded to Zenodo and UCL Discovery stores with appropriate metadata as part of the upload record.

Q: What naming conventions do you follow?

A: Project deliverables are assigned a unique identifier DITOs-D9.9-YYYYMMDD.

All other files are named DITOs -YYYYMMDD-nnnnnnnn where YYYYMMDD is the date of the file and nnnnnn is a brief description of the content. In addition, items uploaded to UCL Discovery and Zenodo will be assigned a unique Digital Object Identifier (DOI).

Q: Will search keywords be provided that optimize possibilities for re-use?

A: Yes, an allowance has been made for this in project metadata.

Q: Do you provide clear version numbers?

A: Filenames will include creation date in YYYYMMDD format. Deliverable reports will have a data identification sheet and version log.

Q: What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

A: Every Dataset will have an associated text document (with the name DITOs-YYYYMMDD-nnnnnnnn-metadata where nnnnnnn is the same as the dataset to which it refers) with its associated metadata.

### 5.2.2 Making Data openly accessible

Q: Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

A: All data identified in this plan for retention will be openly accessible.

Q: How will the data be made accessible (e.g. by deposition in a repository)?

A: Printed reports and publications will be uploaded into UCL Discovery. Datasets (with associated metadata) will be uploaded to Zenodo in compliance with OpenAire recommendations

Q: What methods or software tools are needed to access the data?

A: Data will be published using standard file formats for preservation (pdf, csv) according to <http://data-archive.ac.uk/create-manage/format/formats-table> (accessed 19/10/2016).

Q: Is documentation about the software needed to access the data included?

A: With the exception of the knowledge sharing platform, all data will be accessed using standard tools according to <http://data-archive.ac.uk/create-manage/format/formats-table> (accessed 19/10/2016).

Q: Is it possible to include the relevant software (e.g. in open source code)?

A: N/A as no software will be archived.

Q: Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible. Have you explored appropriate arrangements with the identified repository?

A: Yes, within UCL Discovery and Zenodo

Q: If there are restrictions on use, how will access be provided?

A: No restrictions.

Q: Is there a need for a data access committee?

A: No.

Q: Are there well described conditions for access (i.e. a machine-readable license)?

A: A Creative Commons (CC) licensing for all data will be included in the metadata.

Q: How will the identity of the person accessing the data be ascertained?

A: This will be determined through the UCL Discovery and Zenodo permission and recording mechanisms; both platforms are ODA compliant.

### **5.2.3 Making Data interoperable**

Q: Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

A: Yes – we use standard file formats as noted above.

Q: What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

A: Every dataset will have metadata record that is provided with it.

Q: Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

A: Citizen Science is by definition inter-disciplinary so PPSR\_CORE will be inter-disciplinary, too.

Q: In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

A: This is not anticipated to be an issue.

### **5.2.4 Increase Data Re-Use (through clarifying Licences)**

Q: How will the data be licensed to permit the widest re-use possible?

A: CC Licenses will be used for all data to be preserved – as per APPENDIX 5 – Creative Commons Licence.

Q: When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

A: The data will be made available for re-use from the end of the project.

Q: Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

A: All archived data will be reusable.

Q: How long is it intended that the data remains re-usable?

A: 10+ years.

Q: Are data quality assurance processes described?

A: DITOs has a quality assurance board; document D6.4 'Self-Assessment' address procedures to ensure data quality.

### **5.3 Allocation of Resources**

Both UCL Discovery and Zenodo are free repositories and have institutional underwriting for longevity and resilience. The resources needed for data preparation and secure deletion prior to the end of the project have been covered by existing project funds. No further expenditure is envisaged.

### **5.4 Data Security**

Citizen Science is inherently about open access to data for download and reuse. After the end of the project, all data was appropriately anonymised/redacted and published under the CC Attribution 4.0 International License – see APPENDIX 5 – Creative Commons Licence.

It is anticipated that all data will be available for at least 10 years after the end of the project – both UCL Discovery and Zenodo have long-term institutional support so should ensure longevity of data.

### **5.5 Ethical Aspects**

All data has been anonymised before release; Event participants signed disclaimers and permissions as identified in D7.1/D7.2.

## **6 Conclusion**

This DMP successfully identifies the data that is will be stored after the DITOs CSA in accordance with FAIR principles.

## 7 References

European Commision. (2016, October 20). *Guidelines on FAIR Data Management in Horizon 2020*. Retrieved from European Commision: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

European Commission. (2016). *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 v3.1*. Retrieved November 03, 2016, from [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

## 1 APPENDIX 1 – Citizen Science Data Curation

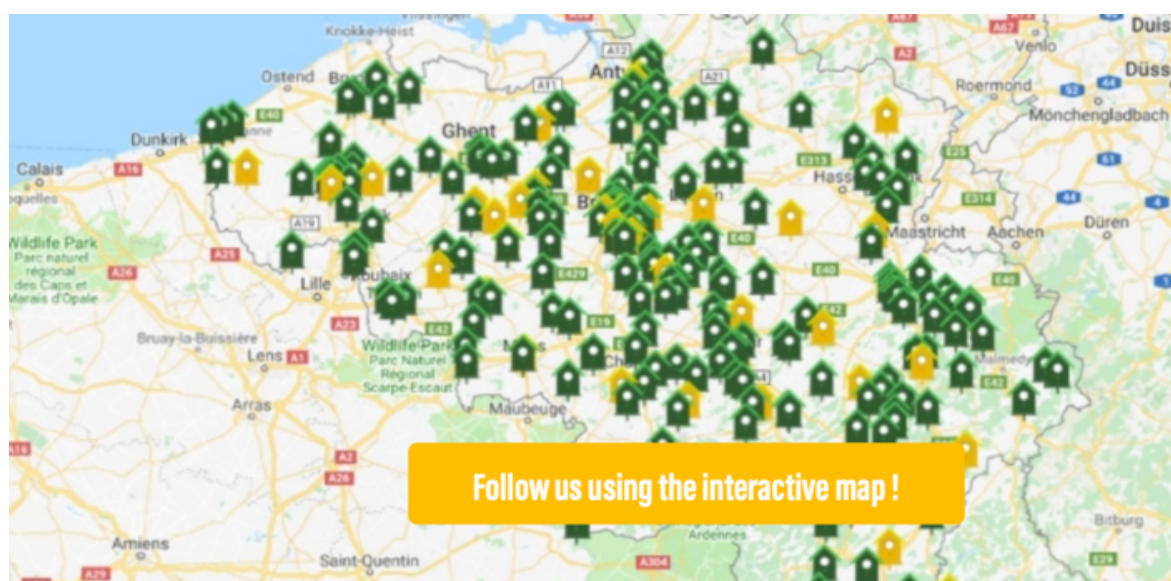
### 1.1 Introduction

This appendix contains a sample of citizen science projects that were run during the project with an overview of scientific data collected, its curation and access. As agreed in D6.2, the curation and retention is beyond the scope of the deliverable, but it provides very interesting and valuable background information so is included in this appendix.

### 1.2 Study 1 – RBINS Xperibird, Belgium

Xperibird is a DITOs-funded school based environmental project where schools set up and install birdboxes, then use remote cameras to monitor and record nesting activity.

Data (structured observations, photos and videos) are uploaded directly onto the RBINS website, and the results are universally accessible, through an interactive map and through the website.



**Figure 3** Xperibird Interactive Map

Participating schools receive a guide on how to enter their data and what will happen to it, namely that RBINS scientists will use it for their research. This is reinforced on the project website. Structured data is recorded against each school, identifying it as the 'originator' of the data.

Koolskamp									
École Communale de Bütgenbach, 4750 Bütgenbach	afgesloten	/	Eurasian Blue Tit	11	30/04/2018 to 30/04/2018	9	14/05/2018 to 15/05/2018	7	03/06/2018 to 05/06/2018
Sint-Jozefscollege, 3200 Aarschot	afgesloten	/	Great Tit	10		10		10	
École communale d'Aywaille, 4920 Aywaille	afgesloten	/	Eurasian Blue Tit	11	22/04/2018 to 29/04/2018	9	11/05/2018 to 12/05/2018	9	30/05/2018 to 30/05/2018

**Figure 4** XperiBIRD.be data available via Website

RBINS store the underlying data locally in csv files, backed up in accordance with the museum's data policies. Whilst the data is available in an interactive format via the project website, it is not yet available in a universally accessible structured format (such as a csv file) with supporting meta data on an open access repository. The museum recognises this as a potential future development, subject to finding appropriate funding. In addition, whilst the originator (the school name) of the data is recorded, there is no formal protocol in place for RBINS scientists to acknowledge the schools in their output.

The main objective XperiBIRD.be is the development of science capital in future adult citizens and it has certainly achieved this objective. However, there is scope for further development in making the data universally accessible; it is hoped that future outreach projects will be able to achieve this.

### 1.3 Study 2 – RBINS Leopold Park Bioblitz, Belgium

During this bioblitz (see D2.1 for bioblitz description), volunteers (citizen scientists) collected physical specimens from the Leopold Park, alongside RBINS scientists. Specimens were photographed, prepared by volunteers and registered by RBINS curators to form part of the museum collection (RBINS collection n°: 33884). Data was catalogued in accordance with the museum's established cataloguing protocols.

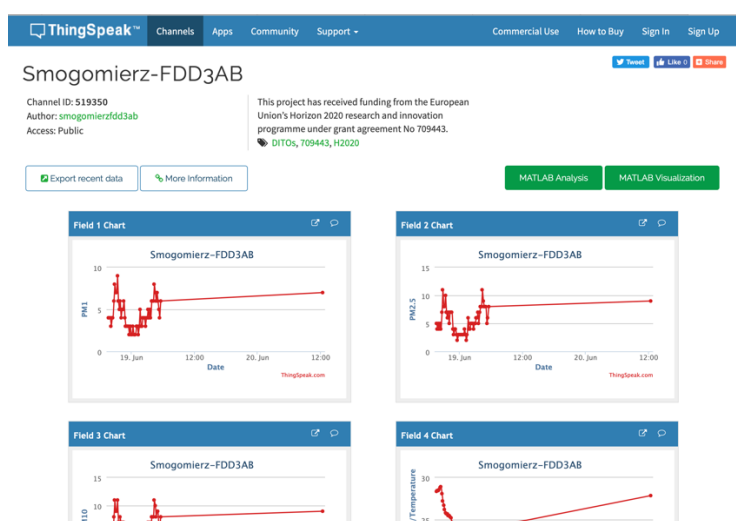
As part of the engagement process, scientists verbally explained to participants that their data and specimens would be used in this way. RBINS has the names of the citizen scientists who participated in the bioblitz, but do not have details of who collected individual samples, so individual recognition of who collected what cannot be made retrospectively.

As the purpose of the event was citizen engagement and progression in citizen knowledge, the lack of individual recognition for specimen collection is not highly significant nor practical given the volume of specimens collected. However, it is something that should be considered for future large-scale data collection or observation citizen science projects.

## 1.4 Study 3 – Meritum Air Quality Science Café

During this event, (see D2.1 for science café description), volunteers used an app to measure and record air quality in Kochlowice, Poland. Participants were given a written statement explaining that the data they collected would be uploaded to a central repository, thingspeak.com and made openly accessible to other scientists and citizen scientists who are members of the thingspeak community. Data is tagged with the project acronym and grant agreement.

Due to the structure of the app, data is tagged with the reference of the collector device, rather than being associated with the citizen scientist.



**Figure 5** Citizen Science Air Quality Data Stored on Thingspeak

The app is a prototype app and Meritum are seeking additional funding to develop it further. This future development will also include supplementary data management considerations such as contributor acknowledgement and appropriate metadata tags. In addition to the app data, the app code is stored on Github<sup>1</sup> under a Creative Commons Licence.

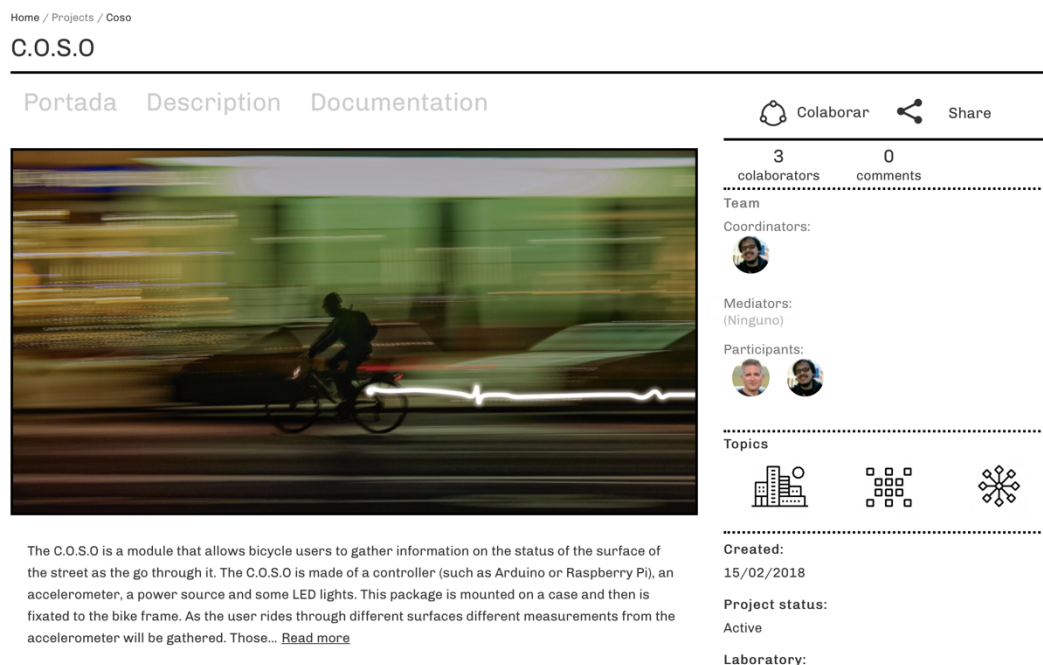
## 1.5 Case Study 4 – Medialab Prado Interactivos?

During this annual event, participants work in teams to develop prototypes and concepts on a theme (past themes have included urban transformations, green futures and food security). The event does not produce data per se; it is an event where participants collaborate in project teams to produce innovative solutions which address sustainable development goals. The blueprints and prototypes are documented by the teams and information is presented via the MediaLab Prado website and archived or backed up in accordance with Medialab protocols.

As the teams themselves document the outcomes, they have true autonomy over what is recorded and in who is acknowledged. The material on the website is open access.

<sup>1</sup> <https://github.com/hackerspace-silesia/Smogomierz>

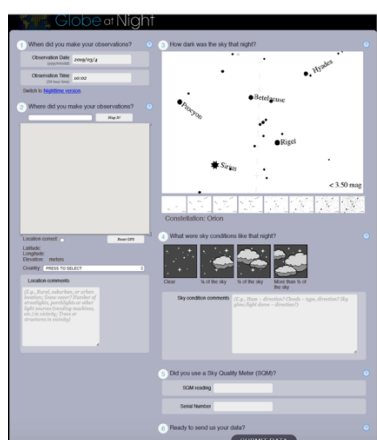
There is no structured record kept of the data with associated metadata – this is a consequence of allowing the participants total freedom in deciding what to record about their project outputs. Given more resources, Medialab Prado would like to put more time into developing protocols and mechanisms to record data in a more structured way with appropriate metadata.



**Figure 6 - Medialab Prado project page created by participants**

## 1.6 Case Study 5 – Into the Night

This project involved citizens across the UK measuring night-time light pollution, optionally using sky quality meters. Participants either used the Globe at Night web app<sup>2</sup> via mobile devices or recorded observations on paper for UCL organisers to upload. The established citizen science app (which started life as a NASA educational project) is used worldwide and has a target of 15,000 data points in 2019.



**Figure 7 Globe at Night App Recording**

<sup>2</sup> [www.globeatnight.org/webapp](http://www.globeatnight.org/webapp)

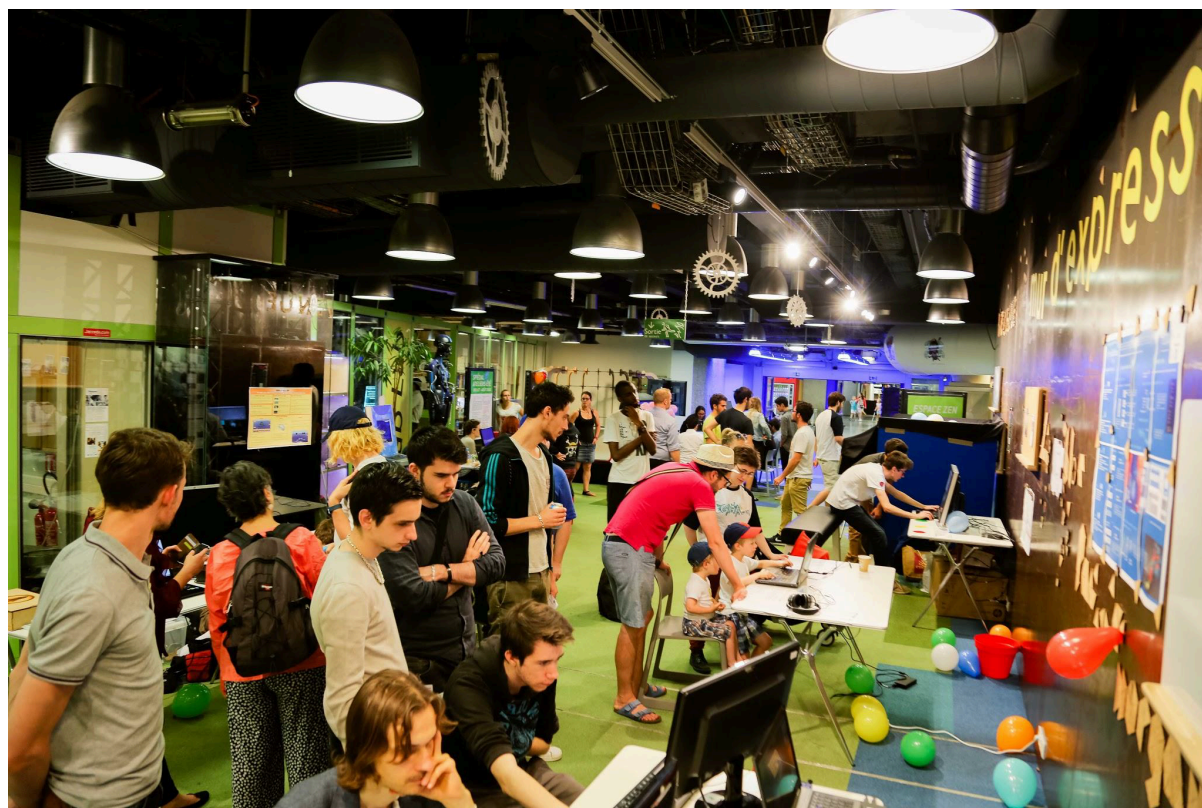
The paper version offered contributors the chance to record their name, but nobody chose to do this. All those who used the paper version were informed that the data would be put into the globeatnight database. On the app itself, there is nowhere for contributors to record their name, just location, date and time.

The (unvetted) raw data collected is downloadable from the website for use by anyone, in various formats under the Open Database License of Open Knowledge International<sup>3</sup>. This licence stipulates that data must be freely available for reuse and redistribution, including being machine-readable and for universal, non-restricted access.

## 1.7 Case Study 6 – UPD i4Gamer

UPD ran a series of gaming events where volunteers ‘played’ online games designed by UPD students and researchers. Participants were made fully aware that the data the game generated or collected would be uploaded to UPD servers for scientific research.

Data that is collected in this way is stored according to the associated project’s data plan – contributors may/may not be acknowledged, and data may/may not be open access. As stated earlier, it is beyond the scope of the DITOs CSA to review these procedures beyond making general recommendations for good practice for event organisers and scientists.



**Figure 8** UPD, iGAMER 2016 - Photo credits: Yves Ininahazwe

<sup>3</sup> [www.okfn.org](http://www.okfn.org)

## 1.8 Recommendations for Data Collected at Citizen Science Events

As part of the DITOs Toolkits produced as part of D1.3, it will be recommended that citizen science event organisers ensure they have protocols in place to ensure that individual citizen scientist's contributions can be attributed to data and specimens collected.

It is recommended that the following considerations are introduced at the initial event planning stage to ensure responsible data management for citizen science events

- 1) The name of the contributing citizen scientist / organisation should be stored alongside each data record / physical specimen
- 2) Appropriate metadata (can be a simple text file) is agreed beforehand to ensure the data is appropriately structured

An appropriate open access database is identified, and protocols are put in place to ensure all concerned (citizen scientists and scientists) are in accordance with this procedure.

## 2 APPENDIX 2 - Metadata to be recorded against each Dataset uploaded to Zenodo

The following should be uploaded as a txt file with each dataset uploaded to Zendo, or recorded as metadata for each report/publication uploaded to UCL Discovery.

Field	Value
Project Name	DITOs
Principal Investigator	Muki Haklay
Project Description	<p>Citizens have a major role to play in addressing the challenges to a sustainable future. It is by 'doing science together' that we combine our resources and expertise to raise awareness, build capacity, and innovative lasting solutions grounded in society. We address the call for Pan-European public outreach in science with and for society, through a tangible 'Do It Together' (DIY) method for wide and deep public engagement and participation in science.</p> <p>This project will support and build upon DIY, grassroots, and frugal innovation initiatives so that in the short and medium term we sustain localised capacity building, and in the long term the effects of these grassroots efforts channel into policy makers at different levels, from external advice to societal inputs, regarding appropriate research and innovation policies. 'Doing-It-Together science' (DITOs) aims to build the institutional and policy</p>

	foundations for deep public engagement in science and technology in Europe.
Funding Sources	EU Horizon 2020 grant agreement 709443
Responsible Organisation	Partner name
Contact Person	Partner Contact Name
Contact Email	Partner Email
Data Overview	Dataset description as given by this document
Technical Information of Files	For qualitative data, description and format of each element
Coding Instrument	The name of the mechanism for accessing files
Data Collection Start Date	
Data Collection End Date	
Confidentiality Classification	Public
Data Subject	Events/Methodology/Technology/People
Keywords	List of keywords, to include H2020, EU, grant agreement no, DITOs, Doing It Together Science, SWAFS, Science with and for Society, citizen science.

### 3 APPENDIX 3 – Deletion Procedures

The consortium will follow the current UCL procedures for secure data deletion.<sup>4</sup>

#### Secure data destruction methods for electronic media

To prevent unauthorised access to data, it is important that the data must be rendered unreadable when the device on which it resides is disposed of or recycled. The appropriate method to use depends on the type of media. There are three main categories of media for these purposes: hard disks, CDs/DVDs and USB drives. Flash memory cards and floppy disks can be treated the same as USB drives. Techniques for destroying all data on these types of media are described below. There are also utilities for erasing individual sensitive files without wiping an entire disk.

#### Hard Disks

<sup>4</sup> <https://www.ucl.ac.uk/drupal/site-information-security/technical-advice>, accessed 17/4/19

Simply deleting files from a computer's hard drive or other storage media is almost never sufficient, as 'delete' simply changes indexing information about a file and the data itself remains on the disk. Emptying the 'recycle bin' or the 'trash' folder of deleted files is also usually ineffective, as the pointers to the deleted files are removed but the data itself still remains on the storage media as unallocated space. There are many widely available programs that can restore data that has been deleted in this way. A plain "format" command is somewhat better, but a determined person could still access sensitive data.

UCL CST recommend that data should be overwritten several times using software tools prior to disposal or recycling of the media. The following free software tool satisfies UCL requirements for secure deletion:

- Darik's Boot and Nuke (DBAN)<sup>5</sup> is a self-contained boot disk that securely wipes hard disks. It does not matter which operating system (eg: Windows, Mac OS, Linux) is on the hard disk to be wiped. To use DBAN, visit the site below, download the .iso file and burn it to CD/DVD as an image. Then boot the target machine from the CD/DVD and follow the prompts. DBAN will allow you to choose from a list of hard disks that it detects and will completely destroy the data on the drives you select. DBAN's default 'DoD short' method overwrites data 3 times and should be sufficient for all purposes. No private data recovery company claims to be able to reconstruct completely overwritten data.

### CDs/DVDs

It is recommended that optical media such as CDs and DVDs are physically shredded.

- USB Drives (including flash memory cards and floppy disks)

USB drives can be erased using DBAN in the same way as hard disks, but it is usually faster and more convenient to use a secure deletion utility specific to your operating system. Free or built-in products for the main operating systems are listed below. These are also recommended for erasing sensitive files from a hard disk without wiping the entire disk. An example of when this should be done is when sensitive data has been stored unencrypted on a laptop hard drive

- Windows  
Eraser<sup>6</sup> is a free file removal utility for all versions of Windows.

Once installed, Eraser adds an 'Erase' option to the right-click menu, so you can simply right-click on files or folders in Windows Explorer and select 'Erase'. If sensitive files have already been deleted in the normal way, you should run the Eraser program and use the 'Unused disk space' option on the drive in question to ensure all traces of the files are removed from the drive.

- Macintosh

---

<sup>5</sup> <http://www.dban.org/>

<sup>6</sup> <http://eraser.heidi.ie/>

Macintosh users have built-in options for secure deletion. For files you've deleted by dragging them to the Trash, use Secure Empty Trash from the Finder menu. It will overwrite and delete files in your Trash folder. For whole file systems, use the Disk Utility, which can be found in the /Applications/Utilities/ folder.

- Unix/Linux  
Wipe<sup>7</sup> is a secure file removal utility for Unix/Linux.

## 4 APPENDIX 4 – Anonymisation Procedures

The consortium will follow the anonymisation procedures laid down by the UK Data Service.<sup>8</sup>

### Primary anonymisation techniques

Remove direct identifiers from a dataset. Such identifiers are often not necessary for secondary research.

*Example:* Remove respondents' names or replace with a code; remove addresses, postcode information, institution and telephone numbers.

Aggregate or reduce the precision of a variable such as age or place of residence. As a general rule, report the lowest level of geo-referencing that will not potentially breach respondent confidentiality. The exact scale depends on the type of data collected, but very detailed geo-references like full postcodes, names of small towns or villages are likely to be problematic. Coded or categorical variables which may be potentially revealing can be aggregated into broader codes. If aggregation of a disclosive variable is not possible, consider whether it should be removed from the dataset.

*Example:* Record the year of birth rather than the day, month and year; record postcode sectors (first 3 or 4 digits) rather than full postcodes; aggregate detailed 'unit group' standard occupational classification employment codes up to 'minor group' codes by removing the last digit.

Generalise the meaning of a detailed text variable by replacing potentially disclosive free-text responses with more general text.

*Example:* Detailed areas of medical expertise could indirectly identify a doctor. The expertise variable could be replaced by more general text or be coded into generic responses such as 'one area of medical speciality', 'two or more areas of medical speciality', etc.

Restrict the upper or lower ranges of a continuous variable to hide outliers if the values for certain individuals are unusual or atypical within the wider group researched. In such circumstances the unusually large or small values might be collapsed into a single code, even if the other responses are kept as actual quantities, or one might code all responses.

---

<sup>7</sup> <http://wipe.sourceforge.net>

<sup>8</sup> <https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation> accessed 17/04/19

*Example:* Annual salary could be 'top-coded' to avoid identifying highly paid individuals. A top code of £100,000 or more could be applied, even if lower incomes are not coded into groups.

Anonymise relational data where relations between variables in related or linked datasets or in combination with other publicly available outputs may disclose identities.

*Example:* In confidential interviews on farms the names of farmers have been replaced with codes and other confidential information on the nature of the farm businesses and their locations have been disguised to anonymise the data.

However, if related biodiversity data collected on the same farms, using the same farmer codes, contain detailed locations for biodiversity data alone the location would not be confidential. Farmers could be identified by combining the two datasets.

The link between farmer codes and biodiversity location data should be removed, for example by using separate codes for farmer interviews and for farm locations.

Anonymise geo-referenced data by replacing point coordinates with non-disclosing features or variables; or, preferably, keep geo-references intact and impose access restrictions on the data instead.

Point data may fix the position of individuals, organisations or businesses studied, which could disclose their identity. Point coordinates may be replaced by larger, non-disclosing geographical areas such as polygon features (km<sup>2</sup> grid, postcode district, county), or linear features (random line, road, river). Point data can also be replaced by meaningful alternative variables that typify the geographical position and represent the reason why the locality was selected for the research, such as poverty index, population density, altitude, vegetation type. In this way, the value of data is maintained, whilst removing disclosing geo-references.

A better option may be to keep detailed spatial references intact and to impose access controls on the data instead.

Procedures to anonymise any research data that are destined for sharing or archiving should always be considered together with appropriate informed consent procedures.

## 5 APPENDIX 5 – Creative Commons Licence

All publicly-available reports contained the following creative commons text:

### Copyright Notice



This work by Parties of the DITOs Consortium is licensed under a Creative Commons Attribution 4.0 International Licence (<http://creativecommons.org/licenses/by/4.0/>).