

# IMPETUS

## Deliverable D7.1

### Data management plan v1

<b>Project Title</b> (grant agreement No)	IMPETUS Grant Agreement 101058677		
<b>Project Acronym</b> (EC Call)	IMPETUS HORIZON-WIDERA-2021-ERA-01		
<b>WP No &amp; Title</b>	WP7, Project Management		
<b>WP Leaders</b>	Gefion Thuermer (KCL)		
<b>Deliverable Lead Beneficiary</b>	8 - KCL		
<b>Contractual delivery date</b>	31/12/2022	<b>Actual Delivery date</b>	16/12/2022
<b>Delayed</b>	No		
<b>Partner(s)</b> contributing to this deliverable	KCL, AE, T6ECO, NESTA		
<b>Authors</b>	Gefion Thuermer (KCL)		
<b>Contributors</b>	Antonella Passani (T6Eco) Vanessa Hanneschläger (AE) Meera Shah (Nesta) Aleks Berditchevskaia (Nesta)		
<b>Acknowledgements</b> (not grant participants)	n/a		
<b>Reviewers</b>	Vanessa Hanneschläger (AE)		



Funded by  
the European Union

Agency

IMPETUS is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101058677. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive

(REA). Neither the European Union nor the granting authority can be held responsible for them.

## Log of changes

Date	Mvm	Who	Description
15/08/2022		Gefion Thuermer (KCL)	Initial draft for comments from partners
21/22/2022		Gefion Thuermer (KCL)	First full draft including all partner datasets and comments.

## Table of contents

<b>1. Executive Summary</b>	<b>2</b>
<b>2. Contribution towards project objectives</b>	<b>4</b>
<b>3. Introduction</b>	<b>Error! Bookmark not defined.</b>
<b>4. Description of work accomplished</b>	<b>7</b>
4.1 Description of work accomplished subheading	7
4.1.1 Description of work accomplished subheading	7
<b>5. Results</b>	<b>4</b>
<b>7. Conclusions</b>	<b>Error! Bookmark not defined.</b>
<b>8. Impact</b>	<b>10</b>
<b>9. Next Steps</b>	<b>11</b>
<b>10. Deviation from Description of Action</b>	<b>12</b>

# 1. Executive Summary

This deliverable consists of the first, preliminary data management plan for the IMPETUS project. It outlines all the data the project already has or currently foresees it will develop, collate, or use. As the project has not yet collected much data, and systems are still being set up for the most data heavy parts of the project (the applications for the open calls, and the CSIs themselves), the plan is preliminary, and provides an overview of what data the project and partners anticipate that they will have, and these are anticipated to be handled. The data management plan will be consistently updated throughout the lifetime of the project, and future iterations of it will be submitted at M24 and M48.

## 2. Contribution towards project objectives

The data management plan is part of the administrative and ethical implementation of the project. It outlines all the data the project does or will use, how it will be collected and used, where it will be stored, and whether or not it will be published. A data management plan by its nature needs to be a living document that follows the project as it develops, and captures changes to data as they occur. The plan covers all data across all project WPs.

## 3. Methods

The data management plan was developed using the DMP online at <https://dmponline.dcc.ac.uk>.

A first draft was developed based on the project GA and initial discussions about the implementation of all the work packages. This was then supplemented by all WP leads to capture data across the entire project.

The plan contains seven parts:

1. **Data Storage:** An overview of the different systems used within IMPETUS to manage data.
2. **Data summary:** This provides an overview of all datasets currently held or anticipated across all IMPETUS WPs. It includes details about data sources, formats, the use and storage of data, and any plans for their publication.
3. **FAIR data:** This provides an overview of how data will be made findable, accessible, interoperable, and reusable.
4. **Allocation of resources:** This provides an overview of the cost for data processing and how they are covered.
5. **Data security:** This provides an overview of how data security is ensured, especially with regards to data sharing beyond the projects' boundaries.
6. **Ethical aspects:** This provides an indication of any ethical considerations related to data.
7. **Other:** This contains additional information relevant to the current and future processing of data within IMPETUS.

## 4. Impetus - Initial DMP

### 4.1. Data Storage

IMPETUS uses several data storage and publication platforms for different purposes:

1. **SharePoint** is the central repository for the project, hosted by coordinator Zabala. This is where all project data and documents are stored and collated.



IMPETUS is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101058677. *Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.*

Adjacent to this space, we will create a sharing platform that allows us to hold and exchange data with selected CSIs leading up to and during the accelerator without making it accessible to the entire consortium.

2. **Google Drive** is used to exchange documents with the citizen panel, and store information the panel needs to access.
3. **EasyChair** is used to manage applications for the accelerator calls, including reviews and contact information of applicants.
4. The **online platform for nominations** is hosted by Ars Electronica and holds all applications and relevant review data about applicants for the European Union Prize for Citizen Science.
5. **A Zenodo community** will be used to publish outputs, data, papers and resources from the Project and the supported CSIs. It is publicly available under <https://zenodo.org/communities/impetus/>.
6. Further project data is stored on (or synchronised from SharePoint onto) **local servers and machines within the partner organisations**.

In addition, the project uses an array of communication and ad-hoc systems, including the project website, Slack, and Doodle.

## 4.2. Data summary

Datasets in Impetus:

1. List of experts (WP1-4)
2. Outreach contacts (WP4, 6)
3. Prospective applicant queries (WP1)
4. Applicant details, proposal documents, and reviews (WP1, 6, 7)
5. Participant CSIs' contracts, plan, and deliverables (WP2, 6)
6. Prize nominees (WP3, 6)
7. Impact assessment reports, surveys and interviews (WP5)
8. Surveys and interviews for policy research (WP4)
9. Citizen panel (WP1, 3, 6)

### 1. List of experts

This will be collated from the combined network of all partners, and overseen by SfC. It will consist of names, email addresses, roles and organisations for experts in various fields relevant to citizen science. The list will be held in .xlsx on the project's SharePoint as well as on the involved partners' local data storage systems.

It is required to conduct core functions of the project. The experts will support the application review process (WP1), mentor and provide training to selected CSIs (WP2), support the nomination for and selection of winners of the European Union Prize for Citizen Science (WP3), and engage in policy workshops (WP4).

This data is held with consent from the experts. It will only be used internally for the duration of the project, and not published (aside from those experts who join the prize jury).

## **2. Outreach contacts**

This will be collated from the combined network of all partners, and overseen by EUSEA.

It is required to conduct core functions of the project. We will need to disseminate our open calls for the accelerator and prizes to relevant stakeholders in the European citizen science community.

It will consist of names, email addresses, roles and organisations of stakeholders in the various fields of citizen science. The list will be held in .xlsx format on the project's SharePoint as well as on the involved partners' local data storage systems.

This data is held with consent from the contacts. It will only be used internally for the duration of the project, and not published.

## **3. Prospective applicants' contact data (Accelerator)**

This will be collected through email, an event registration platform (like Eventbrite), and a webinar platform (like MS Teams).

It is required to adequately support prospective applicants, and complement the project FAQ.

The data will consist of names, email addresses, and organisations of prospective applicants. Tabular data (like webinar registrations) will be held in .xlsx format on the project's SharePoint as well as on the involved partners' local data storage systems. Email queries and corresponding contact details will be held locally in email systems of KCL and/or Zabala.

The data is collated for contract fulfilment, and will only be used internally for the duration of the project. The queries submitted will be used to supplement the open call FAQ on the project website, without exposing any details of who submitted them.

## **4. Applicant details and reviews**

This will be collected through the online platform EasyChair.

It is required to conduct the selection process in WP1, and enable contract negotiation and onboarding of selected projects in WP2/7. We expect applicants to be citizen science projects and researchers from European and Associated countries.

The data will consist of applicants' names, email and phone details, organisation and role, as well as an application form explaining their proposed project idea, impact, and budget. It will be held primarily on the EasyChair system, with exports in .xlsx, .pdf, .docx format stored on the project's SharePoint as well as on the involved partners' local data storage systems.

Selected applications will be shared with reviewers from the experts list (see 1.) for assessment. All reviewers will be bound by data protection requirements.

The data will be used internally, to engage with applicants and select the most promising CSIs. Anonymised data will be used for research, and summaries of the data (e.g. applicants' geographic spread) will be shared on the Call Dashboard (T1.4). The data is

collected for contract fulfilment, and stored until all project tasks (including analysis and potential publications) are completed. Anonymised and published portions of the data may be published and remain available on Zenodo.

### **5. Participant CSIs' contracts, plan, and deliverables**

This will be collected directly from engagement with selected CSIs.

It is required to complete the selection process in WP1, is the subject of contract negotiation and onboarding of selected projects in WP2/7, and ongoing support and assessment of projects in WPs 2/4/5.

The data will consist of project teams' names, email and phone details, organisation and role, as well as a work plan explaining their project idea, prospected impact, and budget. It will be held in .xlsx, .pdf, .docx format on a restricted section of the project SharePoint, which will be shared with the respective applicants.

The data will be used internally for contract fulfilment, to engage with and support selected CSIs (WP2/4), as well as to enable the impact assessment report (WP5). It will remain stored at Zabala beyond the project duration in line with legal obligations in Spain.

### **6. Prize nominees and winners**

This will be collected through the online platform for nominations by Ars Electronica. Ars Electronica provides in-kind an already existing, tested and regularly updated submission tool. The platform uses secure online communication (https) and is hosted on servers of Ars Electronica complying with GDPR standards.

It is required to conduct the selection process in WP3, and enable reviews and selection of prize winners. We expect nominees to be citizen science projects and researchers from European and Associated countries.

The data will consist of nominees' names, email and phone details, organization and role, biographical details as well as details about the nominated project and its impact. The personal data will be collected at minimum about the submitter, but potentially also about the other project collaborators. It will be held primarily on the nomination system, with exports in .xlsx, .pdf, .docx format stored on the project's SharePoint as well as on the involved partners' local data storage systems. Selected nominations will be shared through the submission platform with jury members from the experts list (see 1.) for assessment. All jury members will be bound by data protection requirements.

Ars Electronica will maintain a closed archive of all nominations data, including physical data (such as physical submission material or printed forms) and digital data (such as submission data) that is only available for specific team members via the intranet of Ars Electronica. The archive is controlled through a user management system with different types of users and access rights, which makes it possible to assign specific access rights to specific users. All data is stored on Ars Electronica servers.

The data will be used internally, to assess nominated CSIs and select the best ones for the European Union Prize for Citizen Science, as well as honorary mentions. Anonymised data will be used for research, and summaries of the data (e.g. nominations' geographic spread) will be shared with the public. Details about the winning projects will be used for dissemination. Consent (GDPR compliance) and usage rights (copyright compliance) from submitters for this type of data use will be collected upon submission.

## 7. Impact assessment surveys and interviews

This will be collected directly from engagement with selected CSIs through online spreadsheets, online questionnaires and, hoc-developed forms such as the impact assessment canvas and through interviews.

It is required to enable and complete the impact assessment in WP5.

The data will consist of interviews and surveys among project participants and other data collected through forms and online spreadsheets. It will be held in .xlsx, .pdf, .docx, .mp3, .wav, .mp4 format on the project's SharePoint as well as on the involved partners' local data storage systems.

The data will be used for research purposes. Only anonymised data will be published on Zenodo; however, it is not clear at this stage whether it will be possible to anonymise the data sufficiently to allow for publication.

## 8. Surveys and interviews for policy research

This will be collected directly from engagement with selected CSIs.

It is required to carry out the action research and policy work for WP4.

The data will consist of interviews and surveys among project teams and selected experts. It will be held in .xlsx, .pdf, .docx, .mp3, .wav, .mp4 format on the project's SharePoint as well as on the involved partners' local data storage systems.

The data will be used for research. Only anonymised data will be published on Zenodo; however, it is not clear at this stage whether it will be possible to anonymise the data sufficiently to allow for publication. Non-anonymised data may be published if participants consent.

## 9. Citizen Panel

Applications for participation in the citizen panel will be collected through an online form on Microsoft Forms.

It is required to recruit a citizen panel which has a core function in the project, across WP1 (challenge definition) and WP3 (prize nomination and selection). It enables the selection of a citizen panel based on a set of criteria, including gender and membership in underrepresented groups. We expect applicants to be citizen scientist and other citizen science stakeholders from European and Associated countries.

The data consists of applicants' names, email, some demographic details (i.e. gender and diversity criteria), as well as an application form explaining their experience and network in citizen science. We explicitly collect very limited personal data, focusing on whether applicants are part of underrepresented groups without further details about which exact groups they belong to. It will be held in .xlsx, .csv, .pdf, .docx format stored on the project's SharePoint as well as on the involved partners' local data storage systems.

The data will be used internally, to engage with applicants, and select the most appropriate panellists. Anonymised data will be used for research, and summaries of the data (e.g. applicants' geographic spread) will be shared in project deliverables.

Profiles of the panel members are available on the IMPETUS website, and a contact list of the selected panel members is available to the panel on a shared Google Drive folder.



## 4.3. FAIR data

### 2.1 Making data findable, including provisions for metadata:

We will publish data about the prize-winning projects every in the Ars Electronica catalogue, published in print as well as at [archive.aec.at/print/](https://archive.aec.at/print/).

We do not yet have specific plans for publications of other data, but will revisit this point and update this plan as appropriate once this changes. Any data that we decide to publish, will be available on Zenodo.

### 2.2 Making data openly accessible:

The prize-winning projects are published every year that will generate text, videos, images, authorship information and further contents that will be published online and in the Ars Electronica catalogue at <https://archive.aec.at/>, published in print as well as at [archive.aec.at/print/](https://archive.aec.at/print/).

The Ars Electronica catalogue is stored in paper in the physical archive of Ars Electronica and online in the Ars Electronica Print Archive [archive.aec.at/print/](https://archive.aec.at/print/). The Ars archive, used for long-term storage of submissions, consists of a public archive, accessible to everyone via internet browser, and includes the work description and metadata, specifically: title, category, prize, submission year, name, nationality, catalogue text, catalogue biography, catalogue credits, and any material.

Survey data from the policy and/or impact assessment research may be published if it is possible to anonymise it. We cannot make this determination at this stage, and will revisit the question at a later stage. Any additional data that we do publish will be openly available without restrictions on Zendo, to which we will also invite CSIs to publish their respective datasets. Data access will require a web browser, spreadsheet software like MS Excel, and / or pdf readers.

### 2.3 Making data interoperable:

This does not apply to our data yet. We will publish data with relevant metadata, and consider existing schemas for any research data we publish in the future-.

### 2.4 Increase data re-use (through clarifying licenses):

Where possible, data will be published using creative commons (CC-BY) licenses.

## 4.4. Allocation of resources

We do not currently foresee any costs associated with data management and publication. The Zenodo community is free of charge, thus ensuring long-term availability of published data. Any additional required resources will be covered out of the project budget by the partner responsible for the data.

## 4.5. Data security

All data will be held and kept secure on the project's SharePoint as well as on the involved partners' local data storage systems. These are only accessible to project members, and include solutions for backup and recovery.

The only personal data that is shared outside of the project consortium is application data for the accelerator and prize, for the purpose of assessment through external experts.

Applicants will consent to this process by submitting their application. Practically, data will not be transferred directly, but access to the respective assessment platforms (EasyChair for the accelerator, the ArsElectronica platform for the prizes) will be granted.

## 4.6. Ethical aspects

IMPETUS is in the process of developing an ethics framework for both the project itself, and the CSIs we will support. This will cover how the project collects and processes data in their own activities, and how equity, diversity and inclusion are implemented in our interactions. With regards to data, the main implication is that we will need to collect some personal, and potentially sensitive personal data, in order to assess the degree to which we meet our own benchmarks for inclusion (such as recruiting a majority of non-male led CSIs). The details of these processes cannot be completed before the submission of this plan, but will be updated at a later date, including a link to the relevant project ethics documentation; this will also be addressed in deliverables D8.2-4 over the lifetime of the project.

## 4.7. Other

### Data Justice

Where data is collected from external sources - such as citizen science projects - we endeavour to ensure that this data is used for the benefit of the people who provided it. To this end, we will build on our own past research (Thuermer at al., "*When data meets citizens: an investigation of citizen engagement in data-driven innovation programmes*"), to develop a template for a Data Justice Plan that we will invite all the CSIs we support to complete alongside their general data documentation. This will be part of the ethical guidance provided to CSIs when they join the accelerator.

## 5. Next steps

This is a preliminary version of the IMPETUS data management plan that will evolve as the project progresses. It will be updated regularly as the project and its data develop. Future iterations will be submitted for review at M24 and M48.