

Hungry EcoCities S+T+ARTS Residencies

Deliverable 6.2 – Project Data Management Plan

Version 1.0

Grant Agreement Number	101069990
Project title	AI-enabled artistic solutions for sustainable food systems
Start date of the project	Sep 1 st 2022
Duration of the project	42 months
Date of submission of the report	February 28, 2023
Workpackage No.	6

Project coordinator: Brno University of Technology (BUoT)
 WP leader: Pavel Smrz, Brno University of Technology (BUoT)
 Lead author: Pavel Smrz, Brno University of Technology (BUoT)
 Reviewers: Jens Bürger, KU Leuven (KUL)

S + T + ARTS
 SCIENCE + TECHNOLOGY + ARTS



European
 Commission

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069990. This publication (communication) reflects the views only of the author, and the European Commission cannot be held responsible for any use which may be made of the information contained therein.

History of changes

Date	Version	Author	Comment
2022-11-02	0.1	Radka Kavalova	Initial version
2022-11-15	0.2	Radka Kavalova	Integration of the input from the Data Stewardship Wizard questionnaire
2022-12-10	0.3	Pavel Smrz	Update on the information about data sets used by the university consortium partners
2023-01-12	0.4	Radka Kavalova	Open Data strategy and info from the prepared 1 st call for creative coders
2023-02-15	0.5	Pavel Smrz	Further structure clarification and update on the shared data approaches
2023-02-24	0.6	Jens Bürger	Internal review performed
2023-02-24	1.0	Pavel Smrz	Final version reflecting the internal review

Table of Content

History of changes	2
Table of Content	3
1. Abstract	5
2. Documentation of re-used and created datasets	6
3. FAIR data	9
3.1 – Making data findable, including provisions for metadata	9
3.2 – Making data accessible	9
3.3 – Making data interoperable	10
3.4 – Increase data re-use	11
4. Other research outputs and resource allocation	12
5. Data security, ethics, and other issues	13
6. Conclusions	14
Partners Hungry EcoCities	15

List of abbreviations

CC-BY-SA	Creative Commons Attribution-ShareAlike
D	Project Deliverable
DOI	Digital Object Identifier
DMP	Data Management Plan
EC	European Commission
EU	European Union
FAIR	Findable, Accessible, Interoperable, and Re-usable Data
GA	Grant Agreement
HEC	Hungry EcoCities Project
HECLab	Hungry EcoCities Digital Repository
IPR	Intellectual Property Rights
JSON	JavaScript Object Notation
JSON-LD	JSON for Linked Data
M	Month
OA	Open Access
ODbL	Open Data Commons Open Database License
PDF/A	ISO-standardized version of the Portable Document Format
RDF	Resource Description Framework
RFC	Request for Comment
SBE	Simple Binary Encoding
SDK	Software Development Kit
SME	Small or Medium Enterprise
S+T+ARTS	Science, Technology, and Arts Initiative of the EC
T	Project Task
TSV	Tab-Separated Values
WP	Work Package

S + T + ARTS

SCIENCE + TECHNOLOGY + ARTS



European
Commission

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069990. This publication (communication) reflects the views only of the author, and the European Commission cannot be held responsible for any use which may be made of the information contained therein.

1. Abstract

According to the Hungry EcoCities Grant Agreement, this deliverable (in its current – project month 6 – instantiation) details the plan for the data management describing the data management methods, how the project-related data will be generated, processed, collected, stored and documented. One of the key activities of the project is to prepare technologies, organize 2 calls, run the selection process, choose winners, and organize residencies of 10 individual creative technologists in the first call and 10 mini-consortia consisting of an artist and an SME in the second call. The residencies in both the calls will result in artistic experiments that will primarily generate data to be managed within the project. This is the initial version of the Project Data Management Plan, prepared before even the first call for artists of the HEC project was prepared. Consequently, the content corresponds to the initial vision on the data management within the project and its updates – the intermediate one at project month 24 and the final one at the end of the project – will refine the strategy and report the actual approaches and data processes followed in the HEC project.

The Hungry EcoCities project, as a S+T+ARTS action, stresses its nature as a cascade financing project for artists and, in the second call, small consortia of an artist and an SME. The consortium will follow the same data management approach for each of the 10+10 experiments that will result from the two calls. Nevertheless, the particular collection of datasets that will be used and/or created within and during the project-supported residencies will be unique for each experiment, so the joint approach will be specified in the artistic experiment reports. In general, the project data management follows the Guidelines on the Data Management in Horizon Europe and identifies the resulting data that will be used and made available. This data will allow artists and SMEs winning the project residencies, as well as other parties outside the project consortium, to directly benefit from the project results.

Various kinds of data, metadata and related information have already been and will be generated and gathered along the development, validation, and assessment stages of the project. The document lists all the datasets considered relevant, together with the description of their foreseen management. The datasets include those relevant to specific tools provided by involved academic consortium partners, API specification and protocols, experiment validation datasets, testing and assessment data from experiments, public source code, scientific publications and research and experience data.

2. Documentation of re-used and created datasets

The consortium technology providers – Brno University of Technology, KU Leuven, and Mendel University in Brno – have initially identified several AI and domain-specific tools that will be offered to artists chosen for the project-supported residencies. Some of the tools are based on reference datasets that have been considered for re-use. Depending on the winning experiment proposal, the involved datasets will be refined and documented (in the experiment reports, relevant project deliverables, as well as in the M24 update of this document).

Some of the existing data will need to be harmonized across different sources of the data. Moreover, some experiments may be using data that needs to be made computer readable first. In such cases, we will document the processes leading to providing the data in a relevant machine-readable form and will provide standardised metadata. The data itself will be made available in computer readable form to others through standard repositories.

Most of the data will be open. According to OpenAIRE , open data is “data that is free to access, reuse, repurpose, and redistribute.” The consortium aims to make the public research data resulting from the project accessible as easy as possible and will pay special attention to guarantee the same for the data resulting from the chosen experiments. The primary aim is to maximise the collaboration potential, increase visibility of project results, and shorten the time the artistic project results are adopted.

The M24 and M42 updates of this deliverable will list all relevant datasets used and gathered along the development of experiments and their demonstration. Each dataset will be examined following the template given by the Guidelines on the Data Management in Horizon Europe. The datasets will include API specification and protocols, public source code, scientific publications and experience data.

In accordance to the EU Open Access policy, we will ensure Open Access (OA) to all peer-reviewed scientific publications. Publications arising from the project will be made public preferably through the option of “gold” OA (open access journals or journals that sell subscriptions and also offer the possibility of making individual articles openly accessible via the payment of author processing charges). In other cases, the scientific publications will be deposited in a repository (“green” OA). Sometimes, publishers impose a period of restricted access (embargo period) up to 6 or 12 months.

The metadata, describing the data being published with a necessary context or instructions to be intelligible for other users, will aim at allowing a proper organization,

search, access, and retrieval to the primary data. We will follow the Zenodo scheme set by the OpenAIRE project and record a common (minimum) set of elements describing the public data source and its nature:

Title	Free text
Creator	Last, First and other names
Date	YYYY-MM-DD
Contributors	Acknowledging the HEC consortium or individual artists/parties
Subject	Keywords, a semicolon separated list
Description	Free text
File format	For example, JSON, PDF, TXT, MP4
Resource type	Document, Video, Image, Audio
Persistent identifier	DOI
Access rights	Open Closed Restricted Embargo

The information will be collected in the form of experiment “factsheets” (see also Deliverable D5.2, Action 6). Each experiment will write a factsheet, with an overview of the purpose and ambitions of the art-driven AI component tested and explored in the specific HTE experiment. The factsheets form the basis for the HEC Lab to get access to the open-source codes and platforms for future collaboration. It will be further developed and elaborated in D2.1 and D2.2. The factsheet will be shared publicly, to show the learnings and insights from the experiment, through the Hungry Ecocities project website on starts.eu.

As for the data archiving and preservation, we employ institutional repositories. The project public website, hosted by Ars Electronica, is accessible at <https://starts.eu/what-we-do/residences/hungry-ecocities/>. The site will provide publicly accessible dataset descriptions as a part of the HEC Lab section, including repositories and go-to channels related to the tools and digital insights created in the project. The HEC Lab will be further elaborated in D2.1 and D2.2. Also, the Factsheets on the specific experiment outcomes will be shared on the web.

The next versions of the data management plan will provide details on specific datasets in the form of tables characterizing the content and the nature of the datasets, sharing the following format:

Reference	DSxx – ABC
Name	ABC (acronym) – Full name
Description	Free text

Data format	File format and potential details
Standards and metadata	Link to an established standard or ad-hoc conventions
Sharing and management	License scheme and possible access restrictions
Archiving and preservation	Repository and preservation details

S + T + ARTS

SCIENCE + TECHNOLOGY + ARTS



European
Commission

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069990. This publication (communication) reflects the views only of the author, and the European Commission cannot be held responsible for any use which may be made of the information contained therein.

3. FAIR data

3.1 – Making data findable, including provisions for metadata

As mentioned above, the data offered by the tool providers (project toolbox creators) is and will be available at the specific repositories used by involved parties. The project will also take care of the data findability through specialized search engines that can help you find diverse data on the web, such as:

- OpenAIRE – <https://explore.openaire.eu/search/find/research-outcomes>
- FAIRsharing – <https://fairsharing.org/search?fairsharingRegistry=Database>

or general:

- Google Dataset Search – <https://datasetsearch.research.google.com/>

We may also employ specialized data portals for data from agricultural sciences, engineering and technology, humanities and social sciences, and health sciences, for example:

- ELIXIR – <https://elixir-europe.org/services/tag/data-resources>

If necessary, the provided metadata will characterize additional specialist expertise required to use a particular dataset. In the case of very specific agricultural appliances or other required hardware or software in addition to what is usually available to the users of the dataset, we will make the information on the availability of such necessary tools also easily findable. Last but not least, the metadata will describe and make easily referenceable any national, funder, sectorial, departmental or other policies and procedures for data management used.

3.2 – Making data accessible

The project is working with the philosophy “as open as possible” for the data and other results. All the data will become completely open immediately. Metadata will be openly available including instructions how to get access to the data. Metadata will be available in a form that can be harvested and indexed (managed by the used repository / repositories – see above).

As mentioned above, depending on the selected experiments, the consortium-external artists will take advantage of pre-existing tools and data and we will detail what existing reference data is reused, where and how is it accessible, and what is regarded as reference and non-reference datasets (for example, for experimental results). A special attention will be also paid to harmonization of different sources of existing data for each

individual experiment and, if an experiment will use data resulting from a collaboration between different groups or institutes, whether there is a need for a collaboration agreement that would describe who can have access to what data.

Although most of the data and results will have a non-physical nature, some agri-food experiments resulting from art studio-led residencies can also collect physical samples. In such case, we will describe how will the samples be identified (samples will receive a universally unique persistent identifier – PID), where will the physical samples be stored, and how will we make explicit cross-reference between physical samples and the digital data.

3.3 – Making data interoperable

The primary means of making data interoperable is brought by unification steps. We will start with naming and file organization of datasets, keeping the relationships between data clear in the names. All metadata that is embedded in the names will also be available in the proper metadata, to prevent accidents with, for example, renamed files. If metadata could need to change, embedding it in the file names may require renaming files during the project; and this may have implications for references to those files. To make data further interpretable, we will also use persistent identifiers to refer to data within the project. The technological partners will also guide artists on what encodings/terminologies/vocabularies/ontologies to use in describing the produced data.

As suggested by the metadata tables discussed in the previous chapter, each dataset will detail data types and formats for each particular subset of the data. We will prefer standard data formats widely used by researchers in the respective fields and data formats enabling sharing and long-term archiving. As for the volumes of data we do not expect that the total amount of the data artists and other users would need to work with or produce would require additional measures exceeding the ones mentioned within this document.

Related to data interoperability is also the question of data acquisition. At least some of the field data will be acquired using specialized measurement equipment. In particular, this is the case of the data collected with the help of the technological tools provided by the Mendel University in Brno. In such cases, the measurement equipment used will be fully documented in the metadata and, if experimental, will become a part of the experiment description.

We do not expect there will be a data integration tool that could handle and combine all the data types creative coders will be dealing with in the project. Nevertheless, if a data integration tool will be created or used, it will be documented in the experiment reports.

3.4 – Allowing data re-use

As stated already in Section 3.2, all the data resulting from the experiments will become open immediately. We will be archiving data (using so-called cold storage) for long term preservation already during the project. The data are expected to be still accessible after minimum 5 years.

To validate the integrity of the results, the following will be done:

- We will run a subset of our processes several times across the different compute infrastructures.
- We will be instrumenting the tools into pipelines and workflows using automated tools.

4. Other research outputs and resource allocation

We use Data Stewardship Wizard for planning our data management and creating this DMP. The management and planning of other research outputs will follow the same principles as described for the data and datasets. Even for the other research outputs, we benefit from data stewardship guidance (e.g., FAIR principles, openness, or security) and it is reflected in our plans.

FAIR is a central part of our data management; it is considered at every decision in our data management plan. We use the FAIR data process ourselves to make our use of the data as efficient as possible.

We will be archiving data (using so-called 'cold storage') for long term preservation after the project but also already during the project. The costs for archiving data will be paid out of project and institutional budgets. The minimum lifetime of the archive is 5 years. The archival period can be extended – library or archive staff will decide. The decision whether or not to extend the renewal be based on available budget. Data formats of data in cold storage will be upgraded if they become obsolete. Archived data will be migrated regularly to more modern storage media (e.g., newer tapes).

We have a reserved budget for the time and effort it will take to prepare the data for publication. In the case additional specialist expertise is required to deal with data issues, we have such trained support staff available.

5. Data security, ethics, and other issues

The archives will be stored in several locations to protect the data against disasters. The archives are protected against loss or theft. It is clear who has physical access to the institutional archives.

We do not expect any experiment will collect any personal data. Nevertheless, if it is the case, the project will specify what legal basis is used to collect and process the personal data, whether the person/the team behind the experiment properly asked the data subjects for their consent, whether we have a legitimate interest: data subjects all expect us to do this data processing, or whether we the collection has a public interest – it is done for the benefit of society.

We use the Data Stewardship Wizard with its Common DSW Knowledge Model (ID: dsw:root:2.4.4) knowledge model to make our DMP. More specifically, we use the <https://vutbr.ds-wizard.org> DSW instance where the project has direct URL: <https://vutbr.ds-wizard.org/projects/b992d2bd-51b8-46f1-bdcf-ca00f2887830>.

S + T + ARTS

SCIENCE + TECHNOLOGY + ARTS



European
Commission

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069990. This publication (communication) reflects the views only of the author, and the European Commission cannot be held responsible for any use which may be made of the information contained therein.

6. Conclusions

This document provides an initial version of the project data management plan. It would not be possible to specify all potential data that is currently available to the consortium partners and could be used by selected residency artists. Rather, we specified general principles and methodology of the data management. Details on particular datasets will be given (for the first set of 10 experiments) in the M24 update of the plan. Its final version will then document all datasets reused and, also, all data collected during the project experiments themselves.

S + T + ARTS

SCIENCE + TECHNOLOGY + ARTS



European
Commission

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069990. This publication (communication) reflects the views only of the author, and the European Commission cannot be held responsible for any use which may be made of the information contained therein.

Partners Hungry EcoCities



Horizon Europe Research and Innovation Action – This Hungry EcoCities project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement 101069990. It is part of the S+T+ARTS programme. S+T+ARTS is an initiative of the European Commission to bring out new forms of innovation at the nexus of arts, science and technology.