
Deliverable No. D1.4

Data Management Plan

Project acronym:
PERFORM

Project Title:
**Participatory Engagement with Scientific and Technological Research
through Performance**

Grant Agreement No: **665826**
**This deliverable is part of a project that has received funding from the
European Union's Horizon 2020 research and innovation programme**

Start date of the project:
1st November 2015

Due date of deliverable: 30/04/2016
Submission date: 30/04/2016
File Name: D1.4 Data Management Plan
Authors: Marina Di Masso and Isabel Ruíz-Mallén (UOC)
Peer-reviewed by: María Heras (UAB), Leonardo Alfonsi (EUSEA)

SUMMARY.....	2
1. INTRODUCTION	3
2. ETHICAL REMARKS.....	3
2.1.1 Legal framework of personal data collection.....	4
2.1.2 Procedures and criteria for research participants' identification and recruitment.....	5
2.1.3 Personal data protection.....	7
3. DATA MANAGEMENT	7
3.1 DATA SETS DESCRIPTION	8
3.1.1 Innovative science education methods based on performing arts (PERSEIA).....	9
3.1.2 Impact assessment of participatory educational processes in students' science learning and engagement (IMPACT)	10
3.2 STANDARDS AND METADATA.....	13
3.3 STORAGE AND ACCESSIBILITY.....	14
3.4 LONG TERM PRESERVATION	17

SUMMARY

This document presents the first version of the PERFORM project Data Management Plan (DMP), which includes the description of the management life cycle for all research data generated by the project. This DMP provides an overview of how the research data will be organized, and how it will be handled during the duration of the PERFORM project and after the project is completed. More specifically, it describes what data will be collected and processed (following what methodology), whether and how these data will be shared and/or made open, and how they will be curated and preserved according to the corresponding ethical requirements.

This DMP is a living document. It shows the current status of reflection within the consortium about the data that the project will be producing. However, during the development of the project internal or external factors may cause changes in data management. Thus the DMP is expected to evolve with the project and will be updated accordingly.

1. INTRODUCTION

As a participant of the Pilot on Open Research Data, the PERFORM project is required to develop a Data Management Plan (DMP) and provide a first version within the first six months of the project. The purpose of the DMP is to provide an analysis of the main elements of the data management policy that will be used by the consortium researchers with regard to all the data sets that will be generated by the project.

In this document the first version of the PERFORM DMP is presented, organized in two main sections. The first section provides an overview of the ethical guidelines and considerations related to our fieldwork and research data management. The second section describes the PERFORM research data sets and provides detailed information on data collection, , data sharing, and data archiving and preservation (including storage and backup).

PERFORM will collect data through three case studies in France, Spain and the UK. Through data collection based on a participatory action research approach, PERFORM will explore the effects of a human-centred education approach based on the use of performing arts in fostering secondary school students' learning about and engagement with STEM (Science, Technology, Engineering, Mathematics). PERFORM will also develop an impact assessment methodology to evaluate the effectiveness of such performance-based approach. According to the described starting point, the PERFORM research data will be organized in two data sets: i) data related to innovative science education methods based on performing arts (PERSEIA) and ii) data related to the impact assessment of the participatory educational process. The justification behind such organization is that we want potential data users to easily understand what kind of information they will find in each data set. Metadata will also orient potential users in the specific content of each data set.

2. ETHICAL REMARKS

PERFORM is working with secondary school students, teachers and researchers, who will be requested to provide personal data during the development of the project. Thus, ethical-related aspects of the research are a basic dimension to be addressed in

the context of data management. As these aspects have been addressed in specific deliverables, this section provides an overview of some relevant ethics remarks related to the research data management (the corresponding deliverables where detailed information can be found are referenced accordingly).

2.1.1 Legal framework of personal data collection

The DMP has been created taking into account several recommendations and guidelines and involving institutions regarding the ethical conduct of research and data / privacy protection:

- The research complies with the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.
- The confirmation copy of the Catalan Data Protection Agency (Catalonian Government) as the competent Institutional National Data Protection Authority is attached in D7.3 POPD. This agency is the correspondent competent authority of both the institution coordinating the research project, the Universitat Oberta de Catalunya (UOC), and the institution that will centralize data collection and storage during the lifetime of the PERFORM project, the Universitat Autònoma de Barcelona (UAB) that is responsible for impact assessment (WP4 leader). The ethical issues of the research will be developed under the supervision of UOC and UAB. The UAB will thus centralise data collection and storage during the lifetime of the research project and the corresponding Ethics Commission for Human and Animal experimentation (CEEAH) – an independent organism appointed by the Council of the UAB, which will supervise the research process. CEEAH members do not have any conflict of interests when making their deliberations- and comply with relevant EU legislations (i.e., the European Charter of Fundamentals Rights). If any unforeseen ethical issues arise during the project, the coordinator will consult with the ethics specialists at the UAB Ethics Commission.

Furthermore, in order to comply with national policies, the data protection officers of each participating institution have been informed and have acknowledged their supervisory role for the ethical conduct of our research.

For more information, please see D7.1 and D7.3 POPD.

2.1.2 Procedures and criteria for research participants' identification and recruitment

Participants' involvement in the PERFORM project will be strictly voluntary, and they will have the right to choose not to participate or to withdraw at any point. In all cases, PERFORM will use informed consents to recruit participants (see D7.2 H).

Schools' identification

Case-study coordinators will use purposive sampling to select four secondary schools from a list of schools with which they have previously collaborated or can potentially collaborate based on their previous expertise. The selection criteria will be only based on ensuring socio-economic variability¹ for comparison purposes (i.e., two secondary schools from low socio-economic contexts and two secondary schools from medium socio-economic contexts). Additionally, case-study coordinators will invite another 10 to 16 different secondary schools in each case study (from low and medium socio-economic contexts preferably), randomly selected from existing contacts of case study coordinators and/or the UNESCO's network of associated schools (ASPnet) in each country to ensure their interest in the project.

Students, teachers and researchers' identification

In each school, participation of students will be collectively agreed by the school principal, teachers and the correspondent PERFORM case-study coordinator according to the research purposes in each stage of the project. PERFORM will also work with teachers from selected schools and interested early career researchers from PERFORM universities and organizations and associated research centres. Case-study

¹ From an ethical point of view, being grouped in low and medium socio economic contexts is not discrimination but a frame setting based on research parameters according to the DoA.

coordinators will identify interested teachers and their involvement will be agreed by the school. Case study coordinators will also identify 2-8 early career researchers whose research is related to the topics selected by the students in each school.

Students' recruitment

PERFORM will work directly with secondary school students who are underage and thus considered persons unable to give informed consent to participate in the research as well as vulnerable individuals. In order to meet the EU and national legal and ethical requirements (i.e., EU Charter of Fundamental Rights) of involving underage students in education research and dissemination PERFORM will follow several steps to obtain informed consent, as detailed in D7.2 H:

First (1) an invitation letter including an information sheet and a consent form is sent by each case-study coordinator to the selected secondary school principals, who signs and sends back the consent form. The information sheet includes information on PERFORM project's objective, main activities, benefits and risks, and the PERFORM team commitment of ensuring data confidentiality according to personal data protection national legislation; the consent form includes authorization for the school participation in PERFORM activities as well as for unrestricted permission for images, videos, and recordings of the school and participating underage students to be used in print, video, digital and internet media for PERFORM dissemination purposes. Then, (2) each case study coordinator, in the presence of the teacher, provides an information sheet explaining the project and the consent form to each student in order to obtain their parental consent for their participation in PERFORM activities. This consent form includes authorization for the child/ren participation in PERFORM activities and unrestricted permission for images, videos, and recordings of the participating underage student/s to be used in print, video, digital and internet media for PERFORM dissemination purposes. This parental consent form, based on an approved template, has been modified according to the corresponding national legal and ethical requirements of each case study country. In addition, students are informed that they can decline to participate in the project or withdraw from it at any moment without need of further explanations event if their parents signed the consent form.

Teachers and researchers recruitment

Each case study coordinator provides selected teachers and early career researchers with an information sheet and a consent form in order to obtain their consent to participate in PERFORM activities. Teachers and researchers are asked to return the signed consent form to participate in the project. The involvement of secondary schools and teachers in the project is encouraged through non-economic and/or economic incentives (education material for schools, punctual payments to cover travelling expenses to the schools and/or official recognition for teachers by the corresponding public administration). If needed, teachers and researchers participating in the project might receive a minor symbolic or monetary compensation for the time they spend participating in the project activities.

Drafts of the information letter and the written consent form that is being used are attached in Annex 1 and Annex 2 of D7.2. H.

2.1.3 Personal data protection

The signed consent forms obtained from the participants will be kept by case-study coordinators until three years after the end of the project.

PERFORM will not collect and/or process personal sensitive data (e.g. ethnicity, political opinions, religion, health, finances) since obtaining such type of data will not contribute to achieve PERFORM aims and objectives. The consortium will treat any personal information connected to the audience response data with a high level of consideration, privacy and ethical practice. During the analysis, when transcribing and coding, data will be anonymised.

No data, other than a count for an anonymous person unwilling to participate, will be kept from people who do not want to take part of the study. PERFORM will keep record of the number of people deciding not to participate to assess whether this type of attrition bias research results.

3. DATA MANAGEMENT

PERFORM will need to collect new data in order to address its research interests and objectives. This will be done mainly through qualitative methods applied in three study cases with secondary schools in France, Spain and the UK.

The generated research data will be organized in two data sets, which are described in this section: data related to innovative science education methods based on performing arts (PERSEIA), and data related to the impact assessment of the participatory educational process (IMPACT). For each data set, we provide the description of methodologies applied for data collection. Also, relevant aspects of data management are addressed: standards and metadata, data storage and sharing, and data archiving and preservation.

3.1 DATA SETS DESCRIPTION

In the following table, the methods, kind of data and samples involved in each data set are described.

Table 1. Overview of the data sets: methods, type of data and samples.

PERSEIAS		
Type of data	Methods	Sample
Primary	Deliberative focus groups	Students from two to four classrooms in each selected school
Primary	Participatory workshops	Students from one classroom in each selected school
IMPACT		
Type of data	Methods	Sample
Secondary	Literature review	Academic articles on science education assessment published at international, peer-reviewed journals
Primary	Participatory workshops	Students from one classroom in each selected school
Primary	Social media analysis	Students' Twitter contributions
Primary	Individual interviews	Students with Twitter contributions, teachers and early career researchers, and key participants in PERSEIA workshops
Primary	Deliberative focus groups	Students participating in PERSEIAs
Primary	Systematic observation	Students participating in PERSEIAs
Primary	Surveys	Students involved in PERSEIAs and a sample of those who are not participating

3.1.1 Innovative science education methods based on performing arts (PERSEIA)

Data collection

The research objectives require qualitative data that are not available from other sources. This data set will contain primary data which will be generated with secondary school students, teachers and early career researchers through qualitative methods:

a) Deliberative focus groups

Deliberative focus groups (DFG) with selected secondary school students and teachers will be organised in each of the three case studies (France, Spain, and UK), each involving two secondary schools from low socio-economic contexts and two secondary schools from medium socio-economic contexts. Between 7 and 12 focus groups will be led by each case study coordinator. The topics for discussion in workshops will be related to the RRI values and dimensions. Data will be stored in an Excel database. The result will be an integrated methodological protocol to develop performance-based activities based on the drama approaches of local partners' expertise. Focus groups will be digitally audio recorded and then copied to an external flash disk. When possible, focus groups will be transcribed in each local language (French, Spanish, and English)². During the transcription process references to other persons will be replaced by pseudonyms.

b) Participatory workshops

Case-study coordinators will conduct participatory workshops with secondary school students, their teachers and early career researchers in each case study with the aim of assisting students in developing their own PERSEIA, following the methodological

²For all methods including recording, non-English transcriptions will be translated into English only where the researcher is fluent in both languages and better able to transcribe in English, or to enable analysis of particular sections of the text. This will avoid unnecessary cost.

protocol generated through DFG. A total of 120 students in each case study (30 students per school, divided in two subgroups) will be actively involved in five participatory workshops to develop their own PERSEIAs and the activities further implementation. The topics to be discussed within these participatory workshops will be: relevant scientific topics addressing societal challenges, critical thinking and self-reflection, interaction between arts and the scientific method, gender issues, and performing skills. Data will be stored in an Excel database. WP2 leader will lead the collection, revision and integration of the individual protocols into a common protocol in order to implement participative workshops with young people, teachers and early career researchers. Workshops will be digitally audio recorded and then copied to an external flash disk. When possible, workshops will be transcribed in each local language (French, Spanish, and English). During the transcription process references to other persons will be replaced by pseudonyms.

3.1.2 Impact assessment of participatory educational processes in students' science learning and engagement (IMPACT)

Developing innovative science education activities linked with applied drama requires accurate and rigorous assessment approaches monitoring and evaluating the effectiveness of these activities in students' engagement and as learning process. Assessments of performance-based approaches are generally based on summative evaluations at the end of the process, although some experiences include experimental designs based on the use of the questionnaires and interviews pre and post-performance. Yet, evaluation methods based on integrative and in-depth approaches covering the whole process of creation and representation of such drama activities have not been systematically developed in the field of science education. Thus, similar research and data do not exist which PERFORM can reuse or which can be compared to.

Data collection

This data set will include primary data generated with secondary school students, teachers and early career researchers using several methods, as well as secondary data:

a) Literature review

Secondary data will be collected through a literature review on transdisciplinary assessment frameworks (e.g. educational psychology, science communication, sociology, performance-based approaches) in order to generate a set of expert-based assessment indicators. Data extraction and analysis will be carried out by three coordinated reviewers, according to the keywords and categories of variables previously defined and stored in an Excel database.

b) Participatory workshops

The literature review will be complemented by the implementation of a participatory workshop in each of the four secondary schools involved in the three case studies. The workshop will be developed prior to the design of the assessment so as to include the criteria and indicators that students consider important to the assessment of the impact of the project. Such approach will contribute to students' active engagement in the evaluation process, while including their voices and values, and thus, complementing the expert-based indicators. UAB will develop the workshop's design and each case study coordinator will implement it in the correspondent country (except in Spain, implemented by the UAB). Data will be stored in an Excel database. Workshops will be digitally audio recorded and used to complement the outputs of the workshop (students' identification of learning and motivational aspects of science educational activities and students' prioritization of learning dimensions). After recorded, the audio file will be copied to an external flash disk.

c) Social media analysis

A social media analysis tool will be specifically developed within the project which will pull tweets related to the different PERFORM activities from the social media Application-Programming Interface (API). The tool will categorise responses according

to specific impact categories derived from an online ethnography results and a communication expert engaged in the project as an Advisory Board member.

d) Individual interviews

Individual interviews will be conducted with different groups: (i) students who have contributed their views on Twitter, (ii) teachers and early career researchers, about their perceptions on students' attitudinal changes and skill improvement at school, and (iii) key participants (i.e. selected students, teachers and early career researchers) of PERSEIA workshops. Interviews will be digitally audio recorded and then copied to an external flash disk. Interviews will be transcribed in each local language (French, Spanish, and English) and references to other persons will be replaced by pseudonyms during the transcription process.

e) Deliberative focus groups

Deliberative focus groups will be conducted with students participating in PERSEIAs in the three case studies to explore and reflect about their learning and engagement in the process. Data will be stored in an Excel database. Data will be also recorded in video and audio. Audio and video files will then be copied to an external flash disk. When possible, DFG will be transcribed in each local language (French, Spanish, and English). References to other persons will be replaced by pseudonyms during the transcription process.

f) Systematic observation

Systematic observation will be conducted during selected PERSEIA workshops. Data will be recorded in video and audio, and then copied to an external flash disk.

g) Surveys

Surveys will be conducted before and after the PERSEIA workshops, in each case study with students involved in PERSEIAs and a sample of those who are not participating. A file needs to be created with the identifications and anonymous codes of survey participants. This will be necessary to match questionnaire submissions with research

performance data. Data will be stored in a digital database. Survey variables will be compiled in an Excel file.

3.2 STANDARDS AND METADATA

Standards

Both data sets will be self-explanatory. Thus the use of standards is not foreseen.

File formats & data processing

We will use Microsoft Word and Excel for text based documents, MP3 or WAV for audio files, and Quicktime Movie or Windows Media Video for video files. These file formats have been chosen because they are accepted standards and in widespread use. Files will be converted to open file formats where possible for long term storage.

Data will be coded through qualitative content analysis (most probably using Atlas.ti; the most appropriate software for a comparative multi-language study has not yet been identified).

Backup

Quantitative and qualitative data will be backed up in an external disk and secured by the corresponding WP and/or task leader on a regular basis, and metadata will include clear labelling of versions and dates.

Confidentiality

In the three study cases we will associate the names of the participants with anonymized IDs. Personal information (e.g. name, age) will be compiled in an Excel file associated to those anonymized IDs, and managed according to ethical requirements.

Due to the personal nature of the interviews, workshops and focus groups, recordings and transcripts will not be made public at any point. Furthermore, any publication resulting from participation in the study will not identify subjects by name.

Metadata: preparing for release

Data need to be prepared once it becomes available for submission to the open repository (see section 3.2). Metadata to be published with the data sets include:

- A description of the PERFORM project that produced the data, its funding agency, as well as a suggestion for citing the data set itself (its authors).
- A description of the methodology that has produced the data.
- A basic description information of the data sets: name, reference, author(s), file names, keywords.

3.3 STORAGE AND ACCESSIBILITY

Storing

Files containing anonymized data will be stored on the respective researchers office computers that are password protected. These computers are connected to the internet. However, in case of unauthorized access, research participants identity is protected because any direct references to personal names, teams or organizations has been replaced with anonymous codes or pseudonyms. Transcripts will be stored on researchers' computer under password protection. Additionally, data such as survey submission and transcripts needs to be backed up by each responsible partner individually on external disks.

Sensitive data

As stated in the ethics remarks section, PERFORM will not collect and/or process personal sensitive data (e.g. ethnicity, political opinions, religion, health, finances) since obtaining such type of data will not contribute to achieve PERFORM aims and objectives. In the specific case of social media analysis, PERFORM will not gather private personal information, including personal sensitive data, or using secretive methods to gain access to social media data. The consortium will treat any personal information connected to the audience response data with a high level of consideration, privacy and ethical practice. For example, PERFORM will anonymise social media quotations included in the research. There will be no public access to participants' personal data (i.e. name, age, gender). See D7.1 for further detailed information on the ethical considerations and requirements regarding collection and storage of data.

Sharing

The most appropriate means of sharing the data generated through the project among the consortium will be online, through the project's intranet (a private space for the consortium, linked to the project's webpage). Thus shared data will be stored in electronic databases only accessible to team members. As for external use of the data, an embargo of 24 months will be established for those data to be published through scientific papers. Also, raw data will be subject to an embargo of 36 months.

Open access

PERFORM partners will provide the project manager with resultant publishing data bases that will be stored in Zenodo (<http://www.zenodo.org>). Zenodo is a multi-disciplinary (non-specific) EU scale repository recommended by the EU OpenAIRE initiative, hosted at CERN. It allows to store the different types of data that the PERFORM project will generate (e.g. plain text, structured text, raw data, audio visual data, images), both open and embargoed. Its 2GB capacity is sufficient for the expected data to be generated. Service is free of charge while offering features such as automatic DOI assignation or direct Dropbox. Files can be published as either private or public, allowing different access levels. The repository can receive data sets, publications but also presentations or lessons among other material to be shared.

Lifecycle of collected data

The following table provides an overview of the lifecycle of the data collected by PERFORM, and the files foreseen to be published via the open repository:

Table 2. Overview of the lifecycle of PERFORM collected data.

Method	Foreseeable period(s) of collection	Date for deletion	Permanent storage in Open Access?
Deliberative focus groups audio files	November 2015 – March 2017 July 2016 – October 2018	5 years after end of project	NO
Deliberative focus groups video files	November 2015 – March 2017 July 2016 – October 2018	5 years after end of project	NO
Deliberative focus groups transcripts	November 2015 – March 2017	None	NO

	July 2016 – October 2018		
Participatory workshops audio files	November 2015 – July 2016 September 2016 – October 2018	5 years after end of project	NO
Participatory workshops transcripts	November 2015 – July 2016 September 2016 – October 2018	None	NO
Literature review database, keywords and categories	November 2015 – July 2016	None	YES
Interviews audio files	May 2016 – October 2018 July 2016 – October 2018	5 years after end of project	NO
Interviews transcripts	May 2016 – October 2018 July 2016 – October 2018	None	NO
Survey variables	May 2016 – October 2018	None	YES
Survey data base	May 2016 – October 2018	None	NO
Survey IDs and names	May 2016 – October 2018	5 years after project ends	NO

Intellectual property licence

Deposited data will be attached to a Creative Commons Licence, as a straightforward and effective way of enabling third parties to access and use research data recommended in the Guidelines on Open Access to Scientific Publications and Research Data in H2020 (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf). More specifically, deposited data will be attached to a CC-BY licence, which allows re-distribution and re-use of the licenced work on the condition that the creator is appropriately credited (<https://creativecommons.org/licenses/by/4.0/>).

Ownership

Ownership of the data generated by the project is regulated by the Consortium Agreement, section 8. Online and archival sources will be cited and clearly

acknowledged in the database and research outputs. Permission will be sought from secondary sources to share the findings of the research on public websites.

Release of publications & data

Consortium members will have the right to benefit from their work in the form of publishing its results. The data needed to reproduce the reported results will be made available latest with the publication of the article.

Project deliverables, insofar as they are public, will be available through the project website.

Articles will published in open access directly through an open access scheme of the chosen publisher.

3.4 LONG TERM PRESERVATION

Long term data conservation will be ensured through the storage of data in Zenodo data repository (see previous section). At this stage of the project, no additional funding or training resources are expected. However, we have to consider the possibility of having unexpected needs (e.g. extra storage space, security copies). Thus, in future revisions of this plan this funding issue could be updated. Raw data (paper and video) will be kept in a locked file cabinet at UOC accessible only to the PI and team members in order to protect the confidentiality of subjects. Research data will be retained for a period of 5 years after the project ends, after which they will be destroyed.

Signed consent forms obtained from the participants will be kept by case-study coordinators until three years after the end of the project (as stated in D7.2 H). After that period, they will be destroyed.