



XR2 LEARN

D1.2 DATA MANAGEMENT PLAN

Author	Ioannis Chatzigiannakis
Work Package	WP1
Delivery Date	03/04/2023
Due Date	31.03.2023
Classification	Public

Status of deliverable

Action/role	Name	Date (dd.mm.yyyy)
Submitted by	Ioannis Chatzigiannakis	27.04.2023
Responsible (WP leader)	Ioannis Chatzigiannakis	27.04.2023
Approved by (internal reviewer)	Rama Krishnan Gopal Ramasamy Thandapani Enrique Hortal	31.03.2023

Author(s) contact information

Name	Organisation	E-mail
Ioannis Chatzigiannakis	CNIT	ichatz@diag.uniroma1.it

Document reviewers

Name	Organisation	E-mail	Comments
Rama Krishnan Gopal Ramasamy Thandapani	LS	rama.gopal@ls-group.fr	
Enrique Hortal	UM	enrique.hortal@maastricht.university.nl	

Document history

Date	Version	Author	Comments
09.02.2023	v0.1	CNIT	Initial version
15.02.2023	v0.2	CNIT	Introducing main content
14.03.2023	v0.3	CNIT	Updating abbreviations table
03.04.2023	v0.4	CNIT	Addressing QA comments
04.04.2023	v0.5	CNIT	Pending final approval
05.04.2023	v1.0	CNIT	Final version
27.04.2023	v1.1	CNIT	Update document template

Disclaimer

This document is a deliverable of the XR2LEARN project funded by the European Union under Grant Agreement no.101092851. 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, while neither the European Union nor the granting authority can be held responsible for any use of this content.

This document may contain material, which is the copyright of certain XR2Learn consortium parties, and may not be reproduced or copied without permission. All XR2Learn consortium parties have agreed to the full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the XR2Learn consortium as a whole, nor a certain party of the XR2Learn consortium warrants that the information contained in this document is capable of use, nor that use of the information is free from risk and does not accept any liability for loss or damage suffered using this information.

Acknowledgment

This document is a deliverable of the XR2Learn project. This project has received funding from the European Union's Horizon Research and Innovation programme under grant agreement N° 101092851.

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, while neither the European Union nor the granting authority can be held responsible for any use that may be made of the information it contains.

Table of Contents

List of abbreviations	6
Executive Summary	7
1 Introduction	8
2 Data and Research Output Summary	9
3 Guiding Principles	11
3.1 OPEN SCIENCE	11
3.2 INTELLECTUAL PROPERTY RIGHTS	11
3.3 DATA SECURITY	11
3.4 PERSONAL DATA PROTECTION AND ETHICAL ISSUES	12
3.5 DATA SHARING	12
3.6 ALLOCATION OF RESOURCES	13
3.7 OTHER RESEARCH OUTPUTS	13
4 FAIR Data Considerations	14
4.1 MAKING DATA FINDABLE	14
4.2 MAKING DATA OPENLY ACCESSIBLE	14
4.3 ENSURING DATA INTEROPERABILITY	15
4.4 INCREASING DATA REUSABILITY	15
5 DMP Review Process	17

LIST OF ABBREVIATIONS

DMP	Data Management Plan
DOI	Digital Object Identifier
GDPR	General Data Protection Regulation 2016/679
CC BY	Creative Commons Attribution International Public License
FAIR	Findable, Accessible, Interoperable, and Reusable
IP	Intellectual Property
IPR	Intellectual Property
IPA	Intellectual Property
API	Application Programming Interfaces
Partners' names and acronyms	
CNIT	CONSORZIO NAZIONALE INTERUNIVERSITARIO PER LE TELECOMUNICAZIONI
F6S	F6S NETWORK IRELAND LIMITED
MAG	MAGGIOLI SPA
LS	LIGHT AND SHADOWS
SYN	SYNELIXIS SOLUTIONS SA
SUPSI	SCUOLA UNIVERSITARIA PROFESSIONALE DELLA SVIZZERA ITALIANA
UM	UNIVERSITEIT MAASTRICHT
HOU	HELLENIC OPEN UNIVERSITY
EADTU	VERENIGING VAN EUROPEAN DISTANCE TEACHING UNIVERSITIES
EITM	EIT MANUFACTURING SOUTH SRL

EXECUTIVE SUMMARY

The XR2Learn goal is to create an accessible platform that provides training resources, online lessons, guidelines, and tools to its partners.

The aim of the XR2Learn Data Management Plan (DMP) is twofold: firstly, to establish important components that will enable the future utilisation of the data obtained and processed during and after the XR2Learn project, and secondly, to explain the nature of the data, its intended application, as well as its storage and management procedures. Consequently, the DMP will guarantee that the data is searchable, ideally via a Digital Object Identifier (DOI), accessible, comprehensible, and capable of being utilised beyond its original purpose, while also being interoperable to particular quality criteria, as stated in the Horizon Europe Open Research Data pilot. The Consortium will mostly be the primary contenders for utilising the open data. The XR2Learn project will make the data obtained through the assessment of project results using the open calls openly accessible. The anonymity of individuals will be preserved to provide researchers with valuable data while safeguarding personal information. The second objective of the DMP is to clarify the data, its intended use, and how it will be stored and managed.

To summarise, this deliverable contains the first version of the XR2Learn Data Management Plan (DMP). It details how the research data collected or produced by the XR2Learn Consortium will be managed during and after the completion of the project. The DMP outlines the criteria and methods for data collection and generation, and whether and how the data will be shared or made publicly available. Furthermore, it explains how metadata and archiving best practices will be employed to ensure that the data is easily discoverable, accessible, interoperable, and reusable for other potential users. Additionally, the DMP specifies which datasets the consortium intends to preserve and in which format. It also examines legal and ethical concerns arising from the collection and/or processing of personal data by XR2Learn, considering various data collection methods such as sensors and data collected in the magic room, interviews, online surveys, workshops, questionnaires, etc.

1 INTRODUCTION

The XR2Learn project involves extensive research activities that will generate and process a significant amount of data from various sources. This will include both digital (such as software, protocols, and models) and physical outputs (such as new materials). Managing this data and outputs will be complex and require coordinated efforts from all consortium partners to ensure consistency in management guidelines. One of the primary concerns is the protection of the **Intellectual Property Rights** (IPR) of the partners. In that matter, the consortium will take appropriate measures to safeguard its IPR and the one from the consortium partners and the third parties participating. The DMP will establish clear guidelines for the management, dissemination, and exploitation of the research outputs to ensure that IPR is respected and enforced throughout the project's lifecycle and beyond.

Furthermore, some XR2Learn use cases will involve human participation and the collection of basic personal data, such as age, sex, job information, and native language. Protecting this personal data in compliance with the General **Data Protection** Regulation 2016/679 (GDPR) and other applicable regulations is a crucial aspect of data management in the project.

Finally, the datasets and research outputs generated through XR2Learn are expected to be highly valuable to the consortium and third parties due to the advanced technologies involved in the research. Therefore, the project will establish procedures to circulate processed data and research outputs through an accessible, free repository that adheres to the **FAIR principles**¹.

The deliverable includes a summary of the data identified in the project's pilots (Section 4), the guiding directives of the Data Management Plan (DMP) according to the Horizon Europe Data Management Plan template (Section 5), the initial guidelines for implementing FAIR data principles (Section 6) and the DMP updating process (Section 7).

¹ <https://www.go-fair.org/fair-principles/>

2 DATA AND RESEARCH OUTPUT SUMMARY

In this section, the consortium partners have identified datasets and research outputs relevant to the various use cases. The table below provides a brief overview of the identified data and research outputs. It is important to note that this is a preliminary list, and it may be expanded and/or modified throughout the project as the development of XR2Learn enablers and projects funded through open calls are further defined, and as other relevant data sources and research outputs are identified. Updates to this list will be included in future versions of the Data Management Plan (DMP).

TABLE 1: IDENTIFIED DATASETS AND RESEARCH OUTPUTS

Type	Details	Format	Estimated size
Project Documents	Including deliverables, and technical documentation regarding development, integration, testing and piloting activities. Those documents originate from the XR2Learn partners. Only the public project document will be publicly available, which might be useful for the research community and the public.	reports (.docx, .pdf, .xls, .csv) photos (.jpeg, .png), project files (.UNITYPROJ, .exe)	MB to GB
Research papers	Research publications will be created by partners, and will be made freely available. Those publications will be useful for the research community focusing on the same research topic.	reports (.pdf)	MB
Sensor and Survey data	Data collected through the monitoring sessions in the magic room, surveys, questionnaires, interviews and focus groups aiming to gather information from a group of individuals on each XR2Learn enabler and project funded through the open calls. Knowledge extracted from this data could become publicly available in anonymized versions.	reports (.docx, pdf, .xls, .csv) dataset (.json, .csv)	MB to some GB

ML models, algorithm, and results	One important goal of XR2Learn is to generate a set of pretrained AI models included in the enablers. Those models will be publicly available with open and standardised protocols in freely accessible repositories.	AI models (.h5, .pb, .pt, .onnx), reports (.docx, pdf, .xls, .csv)	GB
3D computer-aided-design (CAD)	This data will be used to facilitate the creation of the applications. It is anticipated that the data may be sourced from either open repositories or commercial sources. However, the specific means of procurement shall be determined at a later time.	CAD files (.fbx, .obj,.dae) 3DS Max files (.max)	MB to GB
Demonstration material	Referring to any kind of visual or physical aid used to illustrate or showcase a concept, idea, process, or progress. Those data will be generated by the partners during the validation and piloting phases.	presentations (.pptx) videos (.mp4, .mov, .avi) photos (.jpeg, .png)	MB to GB
Dissemination material	Including any kind of content, information, or communication tool that is used to share or distribute research findings, knowledge, or best practices to a wider audience, such as web articles, brochures, videos, presentations, webinars, and social media posts.	N/A	N/A

3. GUIDING PRINCIPLES

3.1 OPEN SCIENCE

Open science is a movement aimed at making research more accessible, transparent, and collaborative by openly sharing research data, methods, and outputs with the scientific community and the public. The XR2Learn consortium believes that open science is essential to promote innovation and progress, increase the impact and reproducibility of research, and facilitate the sharing of research outputs. The consortium aims to make publications and research data available as open and free as possible, while still maintaining necessary privacy and confidentiality, in accordance with the policy of Horizon Europe.

The XR2Learn consortium will deposit scientific publications in a trusted repository under the Creative Commons Attribution International Public License (CC BY) or an equivalent licence. Metadata of the deposited publications will also be created and made open under appropriate licences in line with FAIR principles. Research data will be managed according to FAIR principles, which are outlined in the Data Management Plan (DMP). While participation in Open Research Europe² is encouraged, it is not required that all research data produced from project activities be openly shared. The consortium has the option to protect specific research results from becoming publicly available when appropriate conditions are met. Details regarding which datasets will be made publicly available through open repositories and which will be kept private will be provided in updated versions of the DMP.

3.2 INTELLECTUAL PROPERTY RIGHTS

The XR2Learn Consortium Agreement defines procedures for managing Intellectual Property Rights (IPR) in the project. These procedures cover confidentiality, ownership, legal protection, utilisation, and sublicensing of results. The agreement also includes a comprehensive Intellectual Property Agreement (IPA) that governs IP management throughout the project, including access rights, exploitation, and arrangements with respect to any joint inventions. The IPA also contains a register of background IP included in the project and manages the incorporation of side-ground IP.

The consortium will ensure that there are no IPR impediments to publishing project results on the project website and other appropriate portals. In terms of scientific publications, all XR2Learn partners will comply with the Open Science policy of Horizon Europe and Grant Agreement Annex 5, which requires them to upload scientific peer-reviewed publications related to their results in an open access centralised repository.

3.3 DATA SECURITY

The XR2Learn project recognizes that a significant amount of data will be produced and utilised during the project's lifetime, including data collected from various sources, training and validation data from online public repositories, surveys, interviews,

² <https://open-research-europe.ec.europa.eu/>

terminology dictionaries, textual datasets from virtual assistants, etc. To ensure that this data is protected, the consortium will follow an extensive data protection concept of Information Assurance³ throughout the project. This concept involves the implementation of five main pillars: integrity, availability, authentication, confidentiality, and non-repudiation.

To realise these guidelines, a systematic security analysis will be carried out in all XR2Learn use cases, which will accommodate the nature of the data collected and diagnose any security threats pertaining to the gathering, processing, and sharing of the data. Additionally, several partners have their own policies on data security that they will continue to employ which include data control, data backup, communication security, software security, dynamicity management, robustness management, security monitoring, and physical security measures, among others.

On a technical level, the consortium will examine different security measures, such as the use of secure protocols like HTTPS or SSL for data transmission, user authentication and verification methods like login procedures with verified credentials or two-factor authentication, and data encryption for both storage and transmission, if needed. By implementing these measures, the consortium aims to ensure the confidentiality, integrity, and availability of the data throughout the project's lifecycle.

3.4 PERSONAL DATA PROTECTION AND ETHICAL ISSUES

The XR2Learn project recognizes the importance of appropriate handling of personal information that may be collected and processed during its research activities. The consortium will adhere to relevant regulations such as the EU's General Data Protection Regulation (GDPR) 2016/679⁴ and any other applicable regulations at the national level. Additionally, the consortium will analyse ethical issues that may arise during data collection and processing and will take measures to safeguard the right to data privacy. These measures may include data anonymisation or pseudonymisation and explicit informed consent forms, as required by the regulations. The consortium will continually review and update its activities to ensure compliance with the regulatory framework.

3.5 DATA SHARING

In summary, the XR2Learn consortium will implement proper mechanisms for data sharing between partners through file sharing or APIs, depending on the case, to ensure fast, standardised, and secure sharing while complying with data security and personal data protection directives. The consortium will also present all data available for reuse from third parties through an open and free mechanism, leveraging the services provided by OpenAIRE⁵ and following the FAIR principles. The data that will be shared will be uploaded to centralised open access repositories like Zenodo⁶ or through a consortium-hosted one.

³ <https://www.itgovernanceusa.com/information/information-assurance>

⁴ <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

⁵ <https://www.openaire.eu/>

⁶ <https://about.zenodo.org/>

3.6 ALLOCATION OF RESOURCES

The statement explains that the XR2Learn project has taken data management into consideration from the beginning, and that all costs related to data handling according to FAIR (Findable, Accessible, Interoperable, and Reusable) principles, such as data security, processing, and archiving, as well as the development of new open repositories, are covered by the project budget. The statement also mentions that the reimbursement of scientific publications is limited to publication fees incurred in fully open access venues for peer-reviewed scientific publications.

The XR2Learn project has designated CNIT as the beneficiary responsible for data management, and the beneficiary collaborates with technical and pilot partners to draft a detailed data management plan. The plan aims to ensure that appropriate data processing actions are applied to make the dataset compliant with the FAIR principals.

3.7 OTHER RESEARCH OUTPUTS

The XR2Learn consortium plans to extend the data management plan (DMP) guidelines to any other research outputs that may be processed during the project's lifetime, including digital outputs such as software, workflows, protocols, AI models, and physical outputs such as new materials and samples. The four FAIR data principles (Findable, Accessible, Interoperable, Reusable) will be applied to the management of these research outputs, with the goal of facilitating their sharing and reuse.

4 FAIR DATA CONSIDERATIONS

The FAIR data principles (Findable, Accessible, Interoperable, and Reusable) are a set of guidelines that aim to support the management and stewardship of digital assets. These principles were published in 2016 by a consortium of scientists and organisations, and they emphasise the importance of computational systems in enabling the findability, accessibility, interoperability, and reusability of data.

Apart from partner repositories, the consortium, at a later stage of the project lifetime, will select to utilise the services of the suitable external repository such as ELIXIR⁷, CEESDA⁸ and DARIAH⁹ following the Horizon Europe Program Guide guidelines. Furthermore, the consortium plans to utilise open online tools, such as ARGOS¹⁰.

4.1 MAKING DATA FINDABLE

The XR2Learn consortium is committed to applying the FAIR principles to all datasets that will be made available to third parties beyond the consortium. To enable compliance with each of the FAIR principles, the consortium will employ various processes and provisions. These include ensuring that datasets have unique and persistent identifiers, using metadata that conforms to community standards and is machine-readable, making sure that data and metadata are stored in trustworthy repositories, and using open and interoperable formats and standards.

In addition, the consortium will work to enable machine-actionable access to data and metadata, using common vocabularies and ontologies to promote interoperability, and enabling data and metadata to be easily discovered and cited through the use of proper citation practices. Finally, the consortium will strive to ensure that data and metadata are reusable, through the use of clear and standardised licences and terms of use, and through the provision of appropriate documentation and support for users.

4.2 MAKING DATA OPENLY ACCESSIBLE

The XR2Learn consortium plans to make research data and metadata collected, produced, and processed during the project, as well as research outputs generated during the project, publicly available in an online repository. For the code-related outputs, the metadata of the datasets and code-examples on how to use the datasets, the GitHub repository will be used, as defined in D1.1.

The datasets generated from the XR2Learn research activities will be publicly available, except for cases where partners hold Intellectual Property Rights to the datasets or where the datasets contain personal information protected by GDPR and other relevant regulations. In the latter case, after careful review and appropriate anonymisation and security checks, processed versions of the datasets will be made available for third-party reuse.

⁷ <https://elixir-europe.org/>

⁸ <https://www.cessda.eu/>

⁹ <https://www.dariah.eu/>

¹⁰ <https://argos.openaire.eu/home>

If datasets collected during XR2Learn research activities contain personal information protected by GDPR or other relevant directives, the consortium will conduct a careful inspection to determine whether they can be made reusable by third parties. If deemed feasible, the datasets will be published after undergoing proper anonymisation, aggregation, and security checks. Tools and documentation for accessing the datasets will be available on the consortium's open data repository or in-house project repository, along with appropriate licensing under the Creative Commons Attribution International Public License (CC BY), Creative Commons Public Domain Dedication (CC 0), or an equivalent licence, following the principle of being "as open as possible as closed as necessary".

4.3 ENSURING DATA INTEROPERABILITY

Choosing appropriate metadata standards and vocabularies is essential to ensure data interoperability when publishing datasets. Therefore, since the XR2Learn research data comes from various fields, the consortium's efforts to support interoperability will depend on the metadata that will categorise them. The upcoming versions of the DMP will specify the particular details on the standards, formats, types, and vocabularies that will be utilised for the metadata datasets.

Additionally, it is important that the metadata standards and vocabularies align with the repository service chosen by the consortium. If Zenodo is chosen as the repository, the metadata must adhere to DataCite's Metadata Schema¹¹ and include recommended terms. Other standard formats, such as Dublin Core¹² or MARCXML¹³, can also be used to export the metadata.

4.4 INCREASING DATA REUSABILITY

The data collected during the XR2Learn research activities span across various cutting-edge research fields, making them highly valuable to scientists in those fields. As a result, the consortium intends to publish the datasets they produce along with licences that adhere to the "as open as possible as closed as necessary" principle, as specified in the Grant Agreement, to encourage their reuse within the scientific and research community.

Considering the diverse concerns and commitments of various research activities and consortium partners, it may not be appropriate to make all datasets available under the same licence. The commonly used Open Access licences for publishing XR2Learn datasets may include Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or an equivalent licence, following the principle of being "as open as possible as closed as necessary". The decision regarding the type of Open Access licence to be used for a particular dataset will be made by the partner who generates or gathers the data, unless it conflicts with the conditions specified in the Grant Agreement, which states "providing open access" to the outputs generated. In particular, if providing open access to some or all of the

¹¹ <https://schema.datacite.org/>

¹² <https://www.dublincore.org/>

¹³ <https://www.loc.gov/marc/marcxml.html>

data would go against the beneficiary's legitimate interests, including commercial exploitation, or would contradict other constraints such as EU competitive interests or the beneficiary's obligations under this Agreement, then the beneficiary may choose not to provide open access to that data. In such cases, the reasons for not providing open access must be clearly stated in the DMP.

Although it is currently not expected, partners have the option to delay the open access publication of certain datasets for a period of time in order to permit publication or to pursue patents or other Intellectual Property Rights. If such a situation arises, future versions of the DMP will provide specific information on the dataset, reasons for the embargo, and the length of the embargo.

5 DMP REVIEW PROCESS

This deliverable represents the initial version of the Data Management Plan for XR2Learn, which will serve as the main reference document for the project's data policy. At this early stage, this deliverable provides preparatory information regarding the data that will be collected, produced, and processed through the project's research activities and use cases, as well as how this data will be made accessible to third parties.

As the project progresses, the Data Management Plan (DMP) will be subject to revision to accommodate evolving requirements and conditions. This is because the DMP should consider a range of potential scenarios that may arise, including the production of new datasets, modifications to existing dataset metadata, adaptations to data access policies or intellectual property rights (IPRs) of the consortium or some of its partners, and unexpected changes in consortium composition and other external factors.