Plan Overview

A Data Management Plan created using DMPonline

Title: H2020 ARETE Project

Creator: Eleni Mangina

Principal Investigator: Eleni Mangina

Data Manager: Eleni Mangina

Project Administrator: Na Li

Affiliation: University College Dublin

Funder: European Commission

Template: Horizon 2020 Template

ORCID iD: 0000-0003-3374-0307

Project abstract:

The ARETE project aims to support the pan-European interactive technologies effort both in industry and academia, through the multi-user interactions within AR technologies evaluated in education in both professional and private contexts. The authoring tools used within ARETE and the provision of access of the AR content developed for the broader community of users within the EU, will increase the European innovation capacity in AR. Through systematic application of human-centred design approaches, ARETE will deliver highly usable, useful and desirable AR technologies and contents, leading to a wider uptake and further stimulate their creative usage.

ID: 73139

Start date: 01-11-2019

End date: 30-04-2023

Last modified: 21-04-2023

Grant number / URL: https://www.areteproject.eu/

H2020 ARETE Project - Initial DMP

1. Data summary

Provide a summary of the data addressing the following issues:

- **State the purpose of the data collection/generation**
- **Explain the relation to the objectives of the project**
- **Specify the types and formats of data generated/collected**
- **Specify if existing data is being re-used (if any)**
- **Specify the origin of the data**
- **State the expected size of the data (if known)**
- **Outline the data utility: to whom will it be useful**

Question not answered.

2. FAIR data

2.1 Making data findable, including provisions for metadata:

- **Outline the discoverability of data (metadata provision)**
- **Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?**
- **Outline naming conventions used**
- **Outline the approach towards search keyword**
- **Outline the approach for clear versioning**
- **Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how**

Question not answered.

2.2 Making data openly accessible:

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**
- **Specify how the data will be made available**
- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**
- **Specify where the data and associated metadata, documentation and code are deposited**
- **Specify how access will be provided in case there are any restrictions**

Question not answered.

2.3 Making data interoperable:

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**
- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

Question not answered.

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

- **Specify how the data will be licenced to permit the widest reuse possible**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**
- **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**

Question not answered.

3. Allocation of resources

Explain the allocation of resources, addressing the following issues:

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**
- **Clearly identify responsibilities for data management in your project**
- **Describe costs and potential value of long term preservation**

Question not answered.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

Question not answered.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Question not answered.

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Question not answered.

H2020 ARETE Project - Detailed DMP

1. Data summary

State the purpose of the data collection/generation

Explain the relation to the objectives of the project

Specify the types and formats of data generated/collected

Specify if existing data is being re-used (if any)

Specify the origin of the data

State the expected size of the data (if known)

Outline the data utility: to whom will it be useful

2.1 Making data findable, including provisions for metadata [FAIR data]

Outline the discoverability of data (metadata provision)

Rich metadata will be provided and explain what metadata standard are you going to use (Dublin Core, for example).

Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

A persistent identifier (PI or PID)[1] is a long-lasting reference to a document, file, web page, or other object. The term "persistent identifier" is usually used in the context of digital objects that are accessible over the Internet.
Examples of PIDs include:

- Virtual International Authority File (VIAF)
- ORCID IDs
- International Standard Name Identifier (ISNI)

- International Standard Book Number (ISBN)
- Digital Object Identifier (DOI)

In the ARETE project we will make use of persistent and unique identifiers for uniquely identifying and defining entities. Some of which are mentioned above and will be used e.g DOI's for publications and articles and ORCHID IDs for researcher's unique identity.

[1] https://en.wikipedia.org/wiki/Persistent_identifier#cite_note-1

Outline naming conventions used

Consistent naming conventions will be used in the ARETE project for defining the files, folders and objects.

File Naming Guidelines

In ARETE data files and folders would be labelled and organised in a systematic and consistent way so that they are easy to find for the research team. As there is no recommended way to name files and folders but consistency is the key therefore in ARETE we suggest following Elements of a filename:

- A project acronym
- Content description
- File type information
- Date (YYYY-MM-DD)
- Creator name or initials
- Version number
- Status info, e.g. draft

As operating systems usually default to sorting files alphabetically, it is helpful to define what comes at the start of the file name. Therefore in ARETE, it is more useful to order the files by date, by author, or by subject.

The benefit of consistent naming of data files in ARETE is that it will be easier to identify all files connected to one data collection event (e.g. one interview session). In pilots, the files related to one collection event (e.g. audio tape, its transcription etc that were taken by the interviewee) can be connected by the file name.

Example:

- wav
- txt
- jpg

Folder Structure

Similar to consistent file naming conventions, ARETE is using a meaningful folder structure to make it much easier for partners and contributors to locate relevant documents. In ARETE, shareable content among consortium partners is structured in folders and subfolders using Google Drive which is only accessible to members of the ARETE consortium. All the information relevant to data collection and data processing procedures, as well as any other relevant data management material will be stored in dedicated folders. As a guideline ARETE also restricts the level of folders to three or four deep and to a maximum of not to have more than ten items in each list.

Outline the approach towards search keyword

Making data Interoperable (FAIR) can be achieved by using suitable standards for data and metadata creation. This can be achieved by making use of appropriate and suitable vocabularies (e.g. for providing search keywords), in order to optimize the data re-use by other interested parties. Descriptive metadata will be utilised and will contain information such as the title, author, date, description, location, acquisition information, which will be of utmost importance in search and keyword search in the ARETE project. At the end of the project, the data and tools will be available through a digital repository which will also facilitate keyword search.

There are a number of freely available tools that can be customised to implement the keyword search in ARETE project and some of those are listed ass follow:

- Google Ads Keyword Planner
- Google Suggest

- Bing Ads Keyword Planner

Outline the approach for clear versioning

Managing different versions of ARETE data/ software can be achieved by following the guidelines described below:

- Uniquely identifying different versions of files using a systematic naming convention, such as using version numbers or dates
  - Record the date within the file: 20190902_documentation_for_my_data
  - Include a version number in the file name: Documentation_v2
  - Include information about the status of the file, e.g. "draft" or "final," as long as you don't end up with confusing names like "final2" or "final_revised".
  - Include information about what changes were made, e.g. "cropped" or "normalized".
- Using version control facilities within the software used (e.g apache subversion, git etc)
- Using file-sharing services with incorporated version control
- Designing and using a version control table (example as follow)

Table : Version control naming convention

File name	Changes to file
ARETE_DMP_1.0	Original document
ARETE_DMP_1.1	Minor revisions made
ARETE_DMP_1.2	Further minor revisions
ARETE_DMP_2.0	Substantive changes

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

IEEE P1589 (IEEE Approved Draft Standard for an Augmented Reality Learning Experience Model) [1]

Scope

The proposed Augmented Reality (AR) learning experience model will specify how to represent learning activities and their according workplace reference models in a standardized interchange format in order to lower entry barriers for authoring of learning experience spanning real world interaction using sensors and computer vision, and web applications. ARLEM allows wrapping learning activity data for each content unit produced into a standard format for interchange, so that the data can be transported to and from the participating systems, and from provider to user.

Purpose

This standard for Augmented Reality Learning Experience Models (ARLEM) provides an overarching integrated conceptual model and the corresponding data model specifications for representing activities, learning context and environment (aka 'workplace'), while linking with other data model components needed for AR-enhanced learning activities. The standard distinguishes slow-changing data for environment description from fast-changing data for step-by-step guidance. It defines the required data models and modeling languages and their bindings to XML and JSON. The purpose of this standard is to support the discovery, retrieval, transfer, and execution of AR-enabled learning content, thereby facilitating the creation of repositories and online marketplaces.

The standard supports reuse and repurposing of existing (learning) content in 'mixed' experiences that combine real-world guidance with traditional media such as instructional video material or existing web applications and widgets.

This new IEEE Standard (draft) promises to provide significant boosts in operational efficiency by making information available to employees needing task support in context in real time. To support implementations of AR training systems, this standard proposes an overarching integrated conceptual model that describes interactions between the physical world, the user, and digital information, the context for AR-assisted learning and

other parameters of the environment. It defines two data models and their binding to XML and JSON for representing learning activities (also known as employee tasks and procedures) and the learning environment in which these tasks are performed (also known as the workplace). This interoperability specification and standard is presented in support of an open market where interchangeable component products provide alternatives to monolithic Augmented Reality-assisted learning systems. Moreover, it facilitates the creation of experience repositories and online marketplaces for Augmented Reality-enabled learning content. Specific attention was given to reuse and repurposing existing learning content and catering experiences combining real world learner guidance with the consumption (or production) of traditional contents such as instructional video material or learning apps and widgets.

The consortium anticipates the creation of new standards following the development of the Pilot studies, the ARETE platform and the ARETE digital repository and updates will follow up on the advanced versions of the ARETE DMP (D2.3).

The ARETE architecture is still in its infancy, but the consortium will participate and follow the recently published ETSI GS ARF 003 standard specifying a functional reference architecture for Augmented Reality (AR) solutions published in 2020**[2]** by ETSI Industry Specification Group - Augmented Reality Framework (ISG ARF). It identifies key components and interfaces, and defines generic functionalities required by an AR solution.

[1] https://www.techstreet.com/ieee/standards/ieee-p1589?gateway_code=ieee&vendor_id=6073&product_id=1951311#full

[2] https://www.etsi.org/deliver/etsi_gs/ARF/001_099/003/01.01.01_60/gs_ARF003v010101p.pdf

2.2 Making data openly accessible [FAIR data]

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

In ARETE we will follow EC Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020,[1] and use additional resources and guidelines provided by OpenAIRE.eu and the Digital Curation Centre's DMP online tool.

All data will be suitable for re-use and will be made available at the end of the project in month 36. Data here refers to final raw data, analysis scripts, and final processed data—are stored together with an instruction file (user manual) and code book file (list of variable names and meanings) as a ZIP-file.

[1]Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, March 2017 update: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Specify how the data will be made available

At the end of the project the data will be openly available through a digital repository (ZENODO Community and ARGOS) accessible via ARETE official website: https://www.areteproject.eu/

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

Table : Methods/Software tools needed to access the data per pilot

Pilots	**1) Methods/ Software tools to access data** **2) Documentation about the software needed to access the data included** **3) Is it possible to include the relevant software (e.g. in open source code)**
Pilot 1	The data from the online ´ProfilED´ case-history form is held in PDF format. To open these files, users will need ***Adobe Reader software*** on ***Macintosh OS X*** or ***Windows*** systems. Additionally, PDF files can be opened with Preview on ***Macintosh OS X***. Users who do not have Adobe Reader already installed on their computer, can download the installer and instructions free from the ***Adobe Web site***. Only the anonymized analytical data arising from the information contained in these forms will be available through the digital repository.
Pilot 2	Quantitative data from pilot 2 surveys will be accessed and processed by ***MS Excel*** and ***SPSS*** “Statistical Package for the Social Sciences”. Online teacher surveys will be administered by a suitable online survey service such as Google Forms or SurveyMonkey and exported to Excel files. Qualitative data from focus groups will be analyzed by a qualitative content analysis using ***MAXQDA*** and ***MS Excel***. After analysis, all data will be compiled and presented using MS office products such as MS Word or MS PowerPoint. All of these software tools are professional applications that require respective licenses, which are available for the researchers responsible.
Pilot 3	Behavioural traces will be converted from the raw format (xAPI statements, noSQL triples) to aggregate tables, thereby de-referencing records as necessary to ensure privacy. The resulting aggregate tables will be released as .csv / .tsv or similar. Analysis scripts are written (no code, just analytic procedures) so that others are free to use whatever software they have available to them to analyse the data (e.g., ***R, SPSS, STATA***).

Specify where the data and associated metadata, documentation and code are deposited

In the ARETE project a number of available tools can be considered for depositing data, associated metadata, documentation and code. Some of the famous tools that can be used are listed as follow:

- Git
- Apache Subversion
- Google Drive
- Digital Repository

Git[1] is a distributed version-control system for tracking changes in source code during software development. It is designed for coordinating work among programmers, but it can be used to track changes in any set of files. Its goals include speed, data integrity, and support for distributed, non-linear workflows.

Apache Subversion[2] (often abbreviated SVN, after its command name svn) is a software versioning and revision control system distributed as open source under the Apache License. Software developers use Subversion to maintain current and historical versions of files such as source code, web pages, and documentation. Its goal is to be a mostly compatible successor to the widely used Concurrent Versions System (CVS).

Google Drive[3] is a file storage and synchronization service developed by Google. Google Drive allows users to store files on their servers, synchronize files across devices, and share files. In addition to a website, Google Drive offers apps with offline capabilities for Windows and macOS computers, and Android and iOS smartphones and tablets. Google Drive encompasses Google Docs, Google Sheets, and Google Slides, which are a part of an office suite that permits collaborative editing of documents, spreadsheets, presentations, drawings, forms, and more. Files created and edited through the office suite are saved in Google Drive.

Digital repository[4] is an online database of digital objects that can store content e.g., text, documents, still images, audio, video, or other digital media formats. Digital objects can consist of digitized content like print or photographs, as well as originally produced digital content like word processor files or social media posts. In addition to storing content, digital repositories provide means for organizing, searching, and retrieving the content contained in the collection.

[1] https://en.wikipedia.org/wiki/Git
[2] https://en.wikipedia.org/wiki/Apache_Subversion
[3] https://en.wikipedia.org/wiki/Google_Drive
[4] https://en.wikipedia.org/wiki/Digital_library

Specify how access will be provided in case there are any restrictions

- ARETE partners will limit access to authorized and identified persons. Limit access to hard copy human subjects data to authorized personnel, and institute a sign-out, sign-in procedure.
- It is envisioned that the documents and codes generated remain restricted and access will be granted only to the authorised users amongst the partners.
- In pilot 3, an extra CSV file will be created in which user identification numbers can be linked over time and datasets. This key file will be stored password-protected and coded on a password-protected computer which is not linked to the safe VU server. Immediately after finalizing data collection and data linking, this key-file will be destroyed.
- At the end of the project final data and tools created will be made publicly available through the ARETE Digital Repository.

2.3 Making data interoperable [FAIR data]

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

To ensure interoperability and to improve data search and citation, the partners will apply data documentation standards and metadata. For all survey quantitative data, CSV format will be used. For all qualitative textual data (interview minutes and other), the PDF format will be used. The following media formats will be used: images, audio, video, and 3D models.

Interoperability of the ARETE data is a priority for this project and the consortium has extensive experience in metadata and vocabularies[1],[2] as fundamental components in the description of the learning resources within all three pilots. Descriptive metadata in educational augmented reality repositories will conform either to Dublin Core or IEEE LOM. The consortium anticipates the area of Augmented reality educational vocabulary will compose a new segment in the metadata vocabularies and standards, especially given the need to link those with the curriculum resources and the provision mechanisms.

As mentioned before (section 2.1.1 and section 2.1.6) that in ARETE "**Standard for an Augmented Reality Learning Experience Model**" and Dublin Core vocabulary is suggested to use for defining data. Using the aforementioned standard and established vocabulary guarantees the interoperability of data. Standards for Augmented Reality Learning Experience Models (ARLEM) provides an integrated conceptual model and the corresponding data model specifications for representing activities, learning context and environment (aka 'workplace'), while linking with other data model components needed for AR-enhanced learning activities. The

standard distinguishes slow-changing data for environment description from fast-changing data for step-by-step guidance. It defines the required data models and modelling languages and their bindings to XML and JSON. The standard supports reuse of existing content (interoperability) in 'mixed' experiences that combine real-world guidance with traditional media such as instructional video material or existing web applications and widgets.

Dublin Core[3] comprises 15 “core” metadata elements. It is one of the simplest and most widely used metadata schema. Built into the Dublin Core standard are definitions of each metadata element that state what kinds of information should be recorded where and how. All elements are optional and repeatable.

Table : Dublin Core defining concepts and element in AR

Name	Data type	Data source (data format)	Metadata details
	Text	Survey data (CSV), interviews minutes (PDF)	Dublin Core Schema; For survey data: variable labels, code labels, and defined missing values (if needed)
AR Learning Experiences	Text, Code	High-level content description data: modelling language files (JSON) packaged in archives (ZIP)	Dublin Core Schema; or direct access via the ARETE digital repository.
Behaviour and impressions data from experience capturing	Text, numeric data	Survey data (CSV), interviews minutes (PDF)	Dublin Core Schema; For survey data: variable labels, code labels, and defined missing values
Stakeholder community events	Text, media data and multimedia	Survey data (CSV), interviews minutes and other feedback (PDF), images (JPEG), audio (WAV), video (MPEG-4)	Dublin Core Schema;

IEEE 1484.12.3-2005[4] - IEEE Standard for Extensible Markup Language (XML) Schema Definition Language Binding for Learning Object Metadata (LOM) defines a World Wide Web Consortium (W3C) Extensible Markup Language (XML) Schema definition language binding of the learning object metadata (LOM) data model defined in IEEE 1484.12.1-2002 Standard for Learning Object Metadata. The purpose of this Standard is to allow the creation of LOM instances in XML. This allows for interoperability and the exchange of LOM XML instances between various systems. This Standard uses the W3C XML Schema definition language to define the syntax and semantics of the XML encodings.

[1] http://lre.eun.org/edrene/publications/756.pdf
[2] https://ieeexplore.ieee.org/document/8346266
[3] https://dublincore.org/

[4] https://standards.ieee.org/standard/1484_12_3-2005.html

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

Standard vocabulary “Dublin core” will be used for all data types that will allow inter-disciplinary interoperability. Dublin Core metadata terms are expressed in RDF vocabularies for use in Linked Data. Creators of non-RDF metadata can use the terms in contexts such as XML, JSON, UML, or relational databases by disregarding both the global identifier and the formal implications of the RDF-specific aspects of term definitions. Such users can take domain, range, subproperty, and subclass relations as usage suggestions and focus on the natural-language text of definitions, usage notes, and examples.

Each term is identified with a Uniform Resource Identifier (URI), a global identifier usable in Linked Data. Term URIs resolve to the (DCMI Metadata Terms) document when selected in a browser or, when referenced programmatically by RDF applications, to one of four RDF schemas. The scope of each RDF schema corresponds to a "DCMI namespace", or set of DCMI metadata terms that are identified using a common base URI, as enumerated in the DCMI Namespace Policy. In Linked Data, the URIs for DCMI namespaces are often declared as prefixes in order to make data, queries, and schemas more concise and readable.

It is worth noting that AR/VR is an emerging field and relevant standards and vocabularies are establishing and emerging as well. As mentioned that the current plan is to use “Dublin Core” but if there will be a need for using other vocabularies and ontologies the mappings will be defined and provided during the later versions of DMPs.

2.4 Increase data re-use (through clarifying licenses) [FAIR data]

Specify how the data will be licenced to permit the widest reuse possible

The ARETE project will share as much data as possible. It is currently understood that the factors which could prevent research data from being shared (ethical, rules of personal data, intellectual property, commercial, privacy-related, and security-related) will not apply in the ARETE project. If some data are to be licensed and released, the project will use Creative Commons BY-NC-ND license[1] - an attribution to allow reproduction and sharing online[2]. Whereas, some parts and tools are under licenses which might be copyrighted due to background IPs involved by the pilot partners.

[1] https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode

[2] Ball, A. (2014). ‘How to License Research Data’. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/how-guides

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

There is no embargo on the data and the tools produced in the ARETE project and all the data, documentation and tools will be made available after the completion of the project.

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

Table : Third party usage of data per pilot

Pilots	Third party usage of the data
Pilot 1	Research data collected in pilot 1 concerning information collected from teachers, parents and children will be re-usable for third parties in an anonymized format in accordance with the FAIR data approach, as agreed by the project consortium.
Pilot 2	Data collected in pilot 2 will be re-usable for third parties in an anonymized format in accordance with the FAIR data approach, as agreed by the project consortium. Additionally, other researchers may refer to ARETE publication/dissemination output and advance respective research.
Pilot 3	All data collected in pilot 3 will be used to realize publications in academic and popular psychological journals, as well as for poster and oral presentation at conferences. Most of the data is suitable for re-use and will be made available after the first publication of the data. This data refers to final raw data, analysis scripts, and final processed data—are stored together with an instruction file (user manual) and code book file (list of variable names and meanings). Due to the longitudinal nature of the pilot, the data for pilot 3 will be stored pseudo-anonymously. All consortium partners have direct interest in making all data available for analysis and publication for the goals and aims mentioned in the project proposal. In theory all data could also be interesting after publication for other researchers outside of the consortium (e.g., behavioral and educational researchers).

Describe data quality assurance processes

In the ARETE project, the data quality will be ensured by different measures. These include validation of the data, replication, comparison with results of similar studies etc. In order to safeguard the transparency, consistency, comparability, completeness and accuracy of the data, ARETE will apply Quality Assurance and Quality Control activities in the form of peer reviews, data summaries, and input data checks.
In ARETE we consider 5 Pillars of Data Quality Management**[1]**:

- The People
- Data Profiling
  - Reviewing data in detail
  - Comparing and contrasting the data to its own metadata
  - Reporting the quality of the data
- Defining Data Quality
- Data Reporting
- Data Repair

Moreover, in ARETE we will define the data quality metrics based on some of the following:

1. Accuracy
2. Consistency
3. Completeness
4. Integrity
5. Timeliness

[1] https://www.datapine.com/blog/data-quality-management-and-metrics/

Specify the length of time for which the data will remain re-usable

As open formats are used for data archiving, the data will remain re-usable until the repository withdraws the data or goes out of business. As defined in the project proposal, the data remains available and reusable up to 5 years after the project ends in Month 42. The project will also archive the research data within Open Research Europe and share with the H2020 ARETE project community in ZENODO. At this stage, it is difficult to estimate the volume of data collected from all three pilots. For example, given the alternative scenarios that has been considered due to the impact of COVID-19, if some video data is required, the volume of data will increase significantly. Currently estimated sizes for each data set are given above (Table NEW: Sharing of Datasets and expected size of data per pilot) with an accurate estimation to be provided at the next update of the DMP (M36). The ARETE Data repository (ARETE Moodle) archiving will be done at University College Dublin (UCD), and the associated costs will be covered by UCD.

3. Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

Estimated costs for data preparation to make it will be presented in the second version of the Data management plan in month 12, but it is envisioned that the expenses consist of additional publication and documentation costs of three months for one full time equivalent and publication costs of the repository. Associated costs for dataset preparation and data management during the project will be covered by the project itself.

We will do a full cost benefit analysis and in order to estimate the exact cost of making data FAIR in ARETE, we will estimate cost (in terms of time, money and effort) in[1],[2]:

- creating and collection of data
- pre-processing and data cleansing
- integration of the data
- analysis of the data
- registration and publication of the data
- peer review

[1] https://op.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1
[2] https://www.rd-alliance.org/how-expensive-fair-compliance-and-how-expensive-it-not-be-fair-compliant-rda-11th-plenary-bof

Clearly identify responsibilities for data management in your project

As mentioned in D 1.2 “POPD - Requirement No. 2”, the appointed Data Protection Officer (DPO) for the entire ARETE project will be the coordinator institution DPO (NUID UCD) and the contact details of the UCD DPO will be made available to all data subjects involved in the research. For host institutions not required to appoint a DPO under the GDPR a detailed data protection policy for the project will be submitted for full ethics approval from the UCD Office of Research Ethics. The project coordinator’s institution (NUID UCD) DPO, will act as an independent advocate for the proper care and use of personal data processed by the consortium. Figure 1 below denotes the structure defined within the ARETE consortium for the management of the data produced within the three ARETE pilots with specific roles and responsibilities for each partner.

Each partner member, as shown in the following Table, (DPO/Data Protection Officer/Data Protection Coordinator/Information Assurance Services Manager) will advise the members of its organisation on how to operate in line with Data Protection Principles. This will include providing members of the organisation within the ARETE Consortium with feedback on their Data Protection Impact Assessment (DPIA), where appropriate and if requested from NUID UCD DPO. DPIAs can be used to identify and mitigate against any data protection related risks arising from the ARETE project, which may affect the organisation or the individuals it engages with. As the

pilots coordinator (WP6), EUN's Data Protection Coordinator will advise on what type of documentation is needed by an organisational unit participating in the pilots to demonstrate GDPR compliance, including Records of Processing Activities (ROPAs) in line with Article 30.

No	Acronym	Contact	Role
1	NUID UCD	gdpr@ucd.ie (Interim DPO: Ulrike Kolch)	DPO
2	CLB	inna@cleverbooks.eu (Inna Armstrong)	DPO
3	WWL	rita@wordsworthlearning.com (Rita Treacy)	DPO
4	SVU	functionarisgegevensbescherming@vu.nl (Hagenauw, D.E.)	DPO
5	ULE	ias@le.ac.uk (Parmjit Singh Gill)	Information Assurance Services Manager
6	EUN	john.stringer@eun.org (John Stringer)	Data Protection Coordinator
7	CNR	rpd@cnr.it (Dott. Giuliano Salberini)	DPO
8	UNW	datenschutz@uni-wuerzburg.de (Behördlicher Datenschutzbeauftragter der Universität Würzburg)	DPO
9	VIC	nora.gurrutxaga@vicomtech.org, (Nora Gurrutxaga)	DPO
10	OBU	BrookesDPO@brookes.ac.uk (Information Compliance Team)	DPO

Table: Consortium partners' DPOs/Data Protection Coordinator/Information Assurance Services Manager

Describe costs and potential value of long term preservation

The relevant costs for the activities of the data management are included within the allocated resources within the ARETE project budget and are associated with the following activities:

- creating and collection of data through the Pilots and the ARETE Moodle digital repository.
- pre-processing and data cleaning, integration of the data and analysis of the data, through the partners involved in the evaluation process of the project
- registration and publication of the data though the utilisation of standard tools (Moodle repository hosted from the coordinator – UCD) and a free of charge research data repository (ZENODO). The long-term preservation for the public research data is guaranteed via the ZENODO Community portal and peer review of the publications through the Open Research Europe, which is the responsibility of Ms Esther Novo (VICOMTECH).

The overall responsibility for data management lies with the project coordinator, Prof. Eleni Mangina, UCD. Supporting the coordinator is a data management team consisting of the Data Controllers for each pilot as shown in the RoPA document of the ARETE project.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

Data collected/ generated during the ARETE project will be dealt with in a safe and secure manner. All the participants from ARETE consortium involved in data collection and generation will be responsible for dealing with the data in a secure manner.

Table: Data security measures per pilot

Pilot	Data Security measures and approaches
Pilot 1	In pilot 1, once data has been transferred to an electronic medium from paper data collection forms and verified, we will destroy original paper forms (e.g., by shredding), unless keeping original paper copies is required (e.g. auditing purposes, by WWL professional standards). WWL will require receipt of completed anonymised NARA II and VERNON spelling forms to double check scoring and in the case of the Vernon test, to obtain a handwriting sample and to analyse spelling error patterns. Once these have been analysed they will be destroyed / shredded.
Pilot 2	Pilot 2 will collect test data via Google Forms or Survey Monkey and export in xls format. For security reasons pilot 2 will export data and save on UCD server. Google Forms responses are stored in a worksheet that can only be accessed through a Google account login[1]. The transmission of data uses SSL to encrypt the data during transport (but not in storage!) The data is as secure as most other systems which take survey data and store it. Access to SurveyMonkey's technology resources is only permitted through secure connectivity (e.g., VPN, SSH) and requires multi-factor authentication. The production password policy requires complexity, expiration, and lockout and disallows reuse. SurveyMonkey grants access on a need to know on the basis of least privilege rules, reviews permissions quarterly, and revokes access immediately after employee termination[2].
Pilot 3	In order to link quantitative and qualitative data of participants over time for pilot 3, an extra CSV file will be created in which user identification numbers can be linked over time and datasets. This key file will be stored password-protected and coded on a password-protected computer which is not linked to the safe VU server. Immediately after finalizing data collection and data linking, this key-file will be destroyed.

[1] https://support.google.com/docs/thread/13156967?hl=en
[2] https://www.surveymonkey.com/mp/legal/security/

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

As mentioned in D1.1 (H Requirement No.1) and D1.2 (Protection of personal data (POPD) Requirement No. 2), the ARETE consortium acknowledges that ethics is essential in fostering responsible research and innovation within this project. The ARETE project is undertaken with total respect for citizen rights (as required by the Lisbon Treaty – European Charter of Fundamental Rights – which came into force 1st December 2009), as well as societal

relevance, fundamental requirements of our conduct[2]. Ethics, being the discipline that provides the tools for systematic reflection on normative claims and questions, plays a central role in all reputable research. **In the ARETE project,** it underpins our research in three ways, by: *enabling better research design*; *translating fundamental commitments into research practice*; and, *enhancing debate and building platforms and guidelines to increase public trust and acceptance.*

The protection of personal data of students and other end users, in particular, involvement of children/minors and other vulnerable groups, is of high importance to the ARETE consortium. Therefore, international and European norms and regulations, as well as best practices, will be applied during each of the three pilots planned. The process and approach to research ethics in the ARETE project follows four main lines of ethical consideration:

1. Recruitment of participants
2. Confidentiality/anonymity/secrecy
3. Consent/assent
4. Measures taken to ensure the avoidance of research misconduct

These aforementioned four lines of ethical considerations are introduced in detail in D1.1: ”H- Requirement No. 1”.

[1] http://www.leru.org/index.php/public/news/leru-publishes-an-agenda-for-ethics-research-in-horizon-2020/.

[2] As noted above our research on security technology will in all phases be guided by the ‘Societal impact checklist for security R&D’ as published in the CIES (2012) Report of the Societal Impact Expert Working Group (http://cies.ie/wp-content/uploads/2014/05/Report-of-the-Societal-Impact-Expert-Working-Group.pdf)

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Table: Other procedures for data management used per partner within the ARETE consortium

Partner(s)	National/funder/sectorial/departmental procedures
NUID UCD	UCD Data Management Plan ● https://libguides.ucd.ie/data/dmp ● https://www.ucd.ie/research/portal/start/implementdatamanagementplan/ UCD Policies & Procedures ● http://www.ucd.ie/gdpr/policiesprocedures/
SVU	For Vrije Universiteit Amsterdam all research data management is in line with the national scientific guidelines of The Netherlands Organization of Scientific Research (NWO). More information about VU-wide data management can be found here: ● https://ub.vu.nl/en/education-research/research-data-services/data-management-for-researchers/index.aspx
UNW	UNW manages any research data in accordance with established national guidelines issued by Deutsche Forschungsgemeinschaft [DFG; German Research Association]. Further information [partly in German language] can be found at: ● https://www.dfg.de/en/research_funding/proposal_review_decision/applicants/research_data/index.html ● https://econtent.hogrefe.com/doi/pdf/10.1026/0033-3042/a000341 ● https://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf
VIC	● https://vicomtech.app.box.com/s/grhrrtb3sfa431q210ckbnddtlwrypzc
CNR	● https://www.itd.cnr.it/page.php?ID=Privacy&FlagSelected=en
OBU/OU	OBU partner will be replaced by Open University (OU) 1st of June 2020. The Open University maintains a portal for research data management: ● http://www.open.ac.uk/library-research-support/research-data-management The research data management policy is accessible here: ● http://www.open.ac.uk/library-research-support/sites/www.open.ac.uk.library-research-support/files/files/Open-University-Research-Data-Management-Policy.pdf
ULE	● https://www2.le.ac.uk/institution/ethics
CLB	● https://www.cleverbooks.eu/gdpr/
WWL	ARETE partner WWL manages any research data in accordance with our Company's GDPR standards as outlined in our Data Protection Policy. The WWL GDPR Policy is available on our website: ● www.wordsworthlearning.com at the following link GDPR Policy
EUN	EUN Partnership has a data coordinator who is responsible for giving advice and guidance to the Office on data protection issues. The EUN data protection coordinator is supported by a member of staff who follows-up specific requests from data subjects covering things such as data rectification, requests for deletion of data and data access requests. Each team in the Office has also appointed a data protection ambassador who is responsible for providing advice and training to their own team members overseen by the data protection coordinator. ● http://www.eun.org/about/partners/code-of-practice

H2020 ARETE Project - Final review DMP

1. Data summary

State the purpose of the data collection/generation

Purpose of data collection/generation per pilot

Pilots	Purpose of Data Collection/ Generation

Pilot 1	Three main aims inform the collection and generation of data in Pilot 1: [1] achieving the pilots' research objectives, as outlined in the previous section, and [2] managing and monitoring the pilots' implementation and [3] informing technical developments in the ARETE WWL app. Anonymised student information has been collected in the form of pre and post intervention standardised tests of reading and spelling competence, in order to compare and evaluate the students' progress with or without the use of the interactive AR content toolkit. Only the students code, month and year of birth and assessment outcomes will be shared outside the school environment for analysis. The anonymised paper NARA II and VERNON test forms will be stored by participating schools. In addition parents of the students in the intervention and control groups have been requested to complete an anonymised online 'ProfilED' Case History form, which is sent directly to WWL for data analysis. The sender's email was not logged nor is there any identifying information on the form. The information gathered in this form is essential for the analysis of Pilot 1 test outcomes as it can identify if certain developmental patterns or comorbid issues enhance or detract from AR intervention outcomes. This data has been securely stored on a server, PC and on an external hard drive (stored in a safe off site). Pilot 1 collects only personal information (name, surname, contact information) from the participating teachers, in the form of consent forms (necessary to obtain the participants' consent to participate in the ARETE research activities). These consent forms have been collected by European Schoolnet (EUN) and stored on EUN's secure servers. All personal data collected within the project activities will be anonymised upon completion of the project, apart from data that need to be kept for the full audit period. In accordance with the data minimization principle, no personal information from the participating students or their parents/legal guardians will be transferred to EUN or to other members of the ARETE consortium for the purpose of Pilot 1. The only personal information collected from students and their parents/legal guardians has been for the purpose of acquiring consent/assent, and these forms have been collected at the level of each participating school. EUN has established a Memorandum of Understanding with each participating school in Pilots 1, in order to ensure that: [1] the heads of the participating schools are appropriately informed about the research carried out in the context of the ARETE project and have provided their agreement for the school to participate in research and [2] that the schools have all the necessary permissions from the parents/legal guardians of the pupils participating in the ARETE research or, in the case such permissions do not exist at the school level, ensure that the schools collect all necessary consent forms from the participating students. This measure is meant to reduce the transfer of sensitive personal information between organisations unless absolutely necessary.

Pilot 2	In Pilot 2, data have been collected mainly for the following purposes: [1] achieving the pilots' research objectives, as outlined in the previous section, and [2] managing and monitoring the pilots' implementation and [3] informing technical developments in the ARETE CLB app. In Pilot 2, anonymized student data have been collected in correspondence with the research objective of testing the efficiency of Augmented Reality for STEM education, to evaluate if the CLB apps can help students improve test scores by up to 33 % and increase the retention rate by up to 100 %. Additionally, pedagogical aspects of the AR use and its impact on teaching and learning will be evaluated. To respond to these claims, the methodology applied included pre-tests and post-tests for students and additional qualitative surveys for teachers. Pilot 2 collected only personal information (name, surname, contact information) from the participating teachers, in the form of consent forms (necessary to obtain the participants' consent to participate in the ARETE research activities). These consent forms have been collected by European Schoolnet (EUN) and stored on EUN's secure servers. All personal data collected within the project activities will be anonymised upon completion of the project, apart from data that need to be kept for the full audit period. As in the case of Pilot 1, and in accordance with the data minimization principle, no personal information from the participating students or their parents/legal guardians will be transferred to EUN or to other members of the ARETE consortium for the purpose of Pilot 2. The only personal information collected from students and their parents/legal guardians will be for the purpose of acquiring consent/assent, and these forms have been collected at the level of each participating school. EUN has established a Memorandum of Understanding with each participating school in Pilots 2 in order to ensure that: ● the heads of the participating schools are appropriately informed about the research carried out in the context of the ARETE project and have provided their agreement for the school to participate in research and ● that the schools have all the necessary permissions from the parents/legal guardians of the pupils participating in the ARETE research or, in the case such permissions do not exist at the school level, ensure that the schools collect all necessary consent forms from the participating students. This measure is meant to reduce the transfer of sensitive personal information between organisations unless absolutely necessary.

Pilot 3	Data collected for the Pilot 3 have been used for the following main purposes: ● to develop the contents of the PBIS-AR application (task 5.1 “Analysis of PBIS Requirements for ARETE” of WP5 “Interactive AR for PBIS”) ● to develop the contents of the PBIS behavioral research lessons (task 5.2 “Analysis of PBIS Requirements for ARETE” of WP5”Interactive AR for PBIS”) ● to investigate the use of AR as a tool for PBIS, evaluating the incremental value of integrating AR contents within PBIS in promoting school-wide and class-wide positive behaviours (task 6.3 of WP6”Pilots’ Implementation Deployment and Evaluation”). All collected data can and will be used to realize publications in academic and popular psychological journals, as well as for poster and oral presentation at conferences. The data will consist of: (1) quantitative data: questionnaire and survey data that will be collected through an online and paper-based survey procedure at one or more time points during the pilot (for specifics, see Annex G “06.2022 - ARETE P3 - Assessment Strategy” of D5.3 “WP5 - D5.3: Analysis of PBIS Requirement for ARETE - Update D5.1”) .
Pilot 4	In pilot 4, data are collected (1) via online surveys, (2) in interviews, (3) from work products, and (4) from system interaction log files. These data are collected to validate the MirageXR toolkit applied in the pilot and to evaluate the teachers’ experiences during the pilot with regards to their user experience (technical perspective) and to their opinions on the potential role of interactive AR toolkits in teaching and learning processes (pedagogical perspective). The data collected in pilot 4 for these purposes will further be used for scientific publications, e.g., in scientific journals and at conferences.

Explain the relation to the objectives of the project

The specific ARETE objectives (specified in the proposal) are listed as follows:

Objective 1: To develop and evaluate the effectiveness of an interactive AR content toolkit

ARETE has ensured that an interactive AR content toolkit is developed for the creation of 3D objects based on AR standards. ARETE toolkit has been designed and has implemented the AR/3D data repositories for storage and retrieval during the lifespan of the project and beyond. ARETE has created standards-compliant AR/3D data infrastructures for educational purposes to ensure applicability, reproducibility, interoperability, accessibility and sustainability.

Objective 2: To apply human-centred interaction design for ARETE ecosystem

ARETE has identified, updated and integrated on an ongoing basis, user-based insights into designing and developing AR content for the pilot studies. The interaction design within ARETE enables different stakeholders to use the AR technology with ease and positive experience for meeting their needs, preferences, and goals, leading to its high adoption and stimulating its creative uses.

Objective 3: To pilot and evaluate the effectiveness of AR interactive technologies

The ARETE ecosystem, which consists of AR emerging technologies, main platform, training platform, mobile app and a multilingual interface has been piloted. Students and EU citizens (i.e. 3000+ in EU member states) participated in four different pilot studies. The ecosystem has been piloted at “γ-phase”, focusing on assessing the impact of the AR content within professional and private contexts (students, teachers, educational technologists). Stakeholders utilised the effectiveness of the ecosystem through the evaluation of specific skill sets and behaviours (STEM; English literacy skills; and behaviour management and self-management skills within the Positive Behaviour Intervention and Support (PBIS) framework).

Objective 4: To communicate, disseminate and exploit the project results

ARETE promotes project awareness and progress details to the wider and targeted markets. In this light, a scientific, societal and economic focused dissemination and market outreach campaign is well-formulated. In this context, the consortium has adapted a three-phase dissemination and market out-reach approach to achieve this objective for take up for beyond the life of the project.

As can be seen in the listed objectives that in ARETE the data has been collected, generated and used during different phases of the project there is a clear need of a comprehensive data management plan. The purpose of data collection and its relation to the objectives of the project is listed as follows:

Table 3: Data collection/generation per pilot in relation to the objectives

Pilots	Data collection and its relation to the objectives
Pilot 1	Anonymised student information has been collected in the form of pre and post intervention standardised tests of reading and spelling competence, in order to compare and evaluate the students' progress with or without the use of the interactive AR content toolkit (***objective 1, 2, 3, 4***). Minimal personal information from the participating pilot teachers will be collected for the purpose of monitoring the pilot's implementation. (***objective 3***)
Pilot 2	During Pilot 2, anonymized student data have been collected in correspondence with the research objective of testing the efficiency of Augmented Reality for STEM education. In detail, the evaluation is designed to evaluate if the CLB apps can help students improve test scores by up to 33 % and increase the retention rate by up to 100 %. Additionally, pedagogical aspects of the AR use and its impact on teaching and learning will be evaluated (***objective 2, 3, 4***). Minimal personal information from the participating pilot teachers will be collected for the purpose of monitoring the pilot's implementation. (***objective 3***)
Pilot 3	Data collected for the Pilot 3 have been used for the following main purposes (***objective 3, 4***): ● to develop the contents of the PBIS-AR application (task 5.1 "Analysis of PBIS Requirements for ARETE" of WP5) ● to develop the contents of the PBIS behavioral research lessons (task 5.2 "Analysis of PBIS Requirements for ARETE" of WP5) ● to investigate the use of AR as a tool for PBIS, evaluating the incremental value of integrating AR contents within PBIS in promoting school-wide and class-wide positive behaviours (task 6.3 "Pilots Deployment." of WP6)
Pilot 4	The data collected in pilot 4 from teachers working with the MirageXR toolkit facilitate conclusions on the usability and usefulness of the toolkit from a technical and pedagogical viewpoint. These results are disseminated, e.g., in scientific publications and conference presentations, to achieve visibility of the research and to stimulate further research in the field.

Specify the types and formats of data generated/collected

The description of the datasets due to the complexity of each Pilot and partners' collaborative involvement along with the distinctive roles in data sharing within each pilot (Data processor, Data collector) is provided in Annex 1. The Record of Processing Activities (RoPA) has been prepared from the consortium to address the needs of the project and will be a live document maintained and updated throughout the duration of the project and beyond its lifetime (up to 5 years after the project ends in Month 42).

In terms of the nature of data, the ARETE project has essentially two classes of data: the first is related to the set of AR objects that will be released as open data. These data are for example video (animations), images (characters) released to support the creation of software tools for learning and teaching processes. The second class of data consists of descriptive (e.g. age and gender), psychological, psychosocial, behavioural, and educational data collected for the purposes of the three pilots. The behaviour data collected will be stored via the xAPI (further details of the implementation will be provided in future updates of the deliverable). Specifically, several psychological and educational standardized tests, questionnaires, interviews, and focus groups will be used for evaluation purposes; behavioural data will be also collected through systematic coding and/or subjective ratings of direct and video-based observations. This second class of data will be released as open datasets files (e.g. csv, xls) and documents files (e.g. .txt, .doc).

Table 4: Types and formats of data generated / collected per pilot

Pilots	Types and formats of data generated / collected
Pilot 1	All students participating in the Pilot will undergo a pre and post intervention assessment on two formal standardized measures (Neale Analysis of Reading Ability revised (NARA II) and Vernon Graded Word Spelling Test). Data regarding reading accuracy, reading comprehension, reading speed and spelling will be collected from the test forms at schools and analysed by the team in Uni Würzburg in pseudonymized format. Students also fill in an online scale to self-assess the motivational effects of the app used. For Pilot 1, Anonymised “ProfilEd” case-history data is collected on every child participating in the Pilot 1 project. This data is collected through an online case history form and is submitted directly to WWL for analysis. WWL will be aware of the students code, however the form cannot be traced back to the sender. The online case-history form “ProfilED” will be used to gather information relative to the project, which is GDPR compliant. Furthermore, all pilot 1 teachers fill in an online survey which seeks to evaluate their experiences with the app during the pilot from a pedagogical perspective. Additionally, the two pilot 1 teacher coordinators are interviewed pre and post intervention to deeper explore their experiences and perceptions of the app-based intervention. Finally, the interaction of the students using the app has been monitored through xAPI specification. To do that, a list of statements has been defined related to specific interactions and then, the application has been prepared so that each time the student interacts, a JSON format statement in the form of “Noun, verb, object” is sent to a learning record store (Learning Locker) deployed in AWS. This makes us have a database with a summary of the interaction of students with the application.

Pilot 2	The following research instruments are used in pilot 2: ● For the target group of students: ○ 1) non-identifying demographic information such as age group and subject; ○ 2) excerpts from a standardized knowledge test for science and mathematics (TIMSS 2015 inventory); ○ 3) Correlates for science/ mathematics knowledge such as attitudes and self-efficacy; ○ 4) A scale on the motivational effects of the app used. ● For the target group of teachers: ○ 1) a qualitative survey including non-identifying demographic information, and pedagogical aspects of and experiences with the intervention and AR implementation; ○ 2) focus groups with teacher coordinators pre and post intervention to deeper explore the teachers' experiences from the pilot study. Student tests are realized in a pen and paper format, teacher data are collected by online surveys and in online focus groups. All data collected are anonymized from the beginning, i.e., no identifying information will be recorded, and no IPs will be saved. In case of paper tests, the piloting schools share pseudonymized test scans with EUN who double check the pseudonymization and then share the scans with UNW for scientific analysis. The data are accessed by qualified ARETE staff only and stored securely. All data collected in pilot 2 and the results of their evaluation can also be used for scientific publications in academic journals, scientific conferences and further dissemination activities. As in pilot 1, the interaction of the students using the apps has also been monitored through xAPI specification, recording JSON format statements in the form "Noun, verb, object" to be able to know how they used the applications and to see how they performed.
Pilot 3	In pilot 3, data refers to final raw data, analysis scripts, and final processed data are stored together with an instruction file (user manual) and code book file (list of variable names and meanings) as a *ZIP-file*. Due to the longitudinal nature of pilot 3, the data will be stored pseudo-anonymously. This means that the data can only be linked with each other through randomly generated user identification numbers. All raw quantitative data will be directly stored digitally as questionnaire and survey outcomes in comma separated values *Excel (csv*) files which are compatible with the basic Text file *(txt) format*. All raw data will be processed and analyzed using statistical programs for which the used analysis scripts will be stored as Text files. The final dataset will also be stored as *CSV* and Text compatible files. All raw qualitative data will directly be stored digitally as audio and/or video files with transcripts as mp3 files and Text files.

Pilot 4	In pilot 4, the survey data are collected pre and post intervention with the online survey tool SurveyMonkey and exported and stored in suitable formats such as SPSS files or .csv files. The survey data are anonymous as no personal data are collected which could be used to identify individuals. The survey includes questions on basic demographics (such as age, country, school type), previous experience with AR and authoring toolkits, teacher attitudes and technology acceptance, information on the learning activity designed and an evaluation of the experiences made with the MirageXR toolkit. In addition, interviews are conducted with volunteers by UNW and UDUR (n=approx. 11). The interviews are organized and recorded by EUN via MS teams. EUN then shares the recordings with UNW and UDUR who create anonymized transcripts. The transcripts are then analysed by a qualitative content analysis. Recordings are deleted after analysis, only anonymized transcripts are stored. In pilot 4, data refers to quantitative data, qualitative data (from interviews), and product data (media asset files from produced AR learning experiences and IEEE P1589-2020 compliant activity and workplace model files) as well as log data gathered from the participating teachers. The 'product data' will be coded against augmentation types to foster investigation of learning designs in use. This descriptive data will be stored as CSV. The participants release their produced AR learning experiences via the ARETE marketplace. The log data is gathered with the help of the eXperience API (xAPI) and stored as triples in a Learning Record Store (Learning Locker) database. This database is deployed in AWS and the collected information is useful to have a summary of how users interacted with the application and draw some conclusions to improve the system.

Specify if existing data is being re-used (if any)

Existing data re-used per pilot

Pilots	Existing Data is being reused
Pilot 1	Selected data from the AHA project[1] are being reused as a historical control group for pilot 1.
Pilot 2	Not applicable for Pilot 2
Pilot 3	No existing (meta) data etc. for pilot 3
Pilot 4	No existing (meta) data etc. is being used for Pilot 4

[1] ADHD Augmented (AHA) project (funded from European Commission, Grant Agreement No. 30-CE-0885096/00-34) https://doi.org/10.1007/s10055-020-00485-z

Specify the origin of the data

A practical approach of collecting input directly from target users has been applied. This allows the collection of specific data unique for the ARETE project's context. Inspired by ethnographic approaches, which involve researchers to assume the role of observers to investigate and collect data from prospective users in situ, a

method known as Contextual Inquiry (CI)[1] has been developed and broadly used. CI enables designers/developers to have a realistic understanding of end-users and their working/learning environments through interviews with and intensive observations of end-users in locations where they typically work/learn. We will apply CI in the early phase of ARETE to inform the design of the ARETE ecosystem.

Aligned with the development life-cycle of the prototype, data will be collected (by UCD, SVU, EUN, CNR, TOU, led by UDUR) including users' interaction patterns with ICT in today's classrooms, and relevant social and cultural factors influencing the user adoption of new educational technology such as AR. This data will enable us to identify facilitating (e.g. motivated and tech-savvy teachers) and hindering factors (e.g. lack of technical infrastructure) to for the uptake of AR in schools, gaining insights how the ARETE project can best be designed and integrated into existing educative practice to maximize its benefits. Both structured and open observations will be conducted; the former is based on systematic literature reviews to derive a template and checklist for identifying behaviours of interest whereas the latter is grounded in Contextual Inquiry - an ethnographic approach involving in situ observations of and interviews with students and teachers. These complementary approaches allow deep insights to be gained for the construction of ARETE use scenarios. In addition, larger-scale surveys (paper as well as web-based questionnaires administered in countries of partners and beyond) for gathering quantitative data to understand the educational uses of AR technology at different academic levels will be conducted.

Table : Origin of the data per pilot

Pilots	**Origin of the Data**
Pilot 1	Pilot 1 provides the 'ProfilED' case-history form to the parent(s) of every student in the intervention and control groups. The students in the WWL-AR intervention and control groups will all undergo a pre and post standardised assessment on both the NARA II and VERNON test. This gives us an accurate baseline to start from and a point to measure progress post intervention. WWL will not be dealing directly with any student and all standardised tests are administered by school staff onsite.
Pilot 2	For **Pilot 2**, CLB will not collect or store any data from schools (i.e. school name, address, location, any names of teachers or students etc.) EUN will be responsible for recruiting the participating schools/teachers and will be the controller of this data. No personal information about the participating students will be collected by the ARETE consortium from Pilot 2. Evaluation instruments will be administered at the school level and the school will ensure that parents and children have given their consent to participate in the ARETE research activities. The consent forms will remain at the school level, to minimize unnecessary transfer of personal information. Data from mobile app tests will not identify any school in particular or its location, only answers to the test questions in scoring of correct answers vs incorrect (D1.2-POPD Requirement No.2).
Pilot 3	The data will consist of: (1) quantitative data: questionnaire and survey data that will be collected through an online and paper-based survey procedure at one or more time points during the pilot (see Annex G "06.2022 - ARETE P3 - Assessment Strategy" of D5.3 "WP5 - D5.3: Analysis of PBIS Requirement for ARETE - Update D5.1"). Behavioural traces of learner interaction with the system and with the content will be logged by the AR app automatically to the xAPI endpoint (further analysis to be provided in future updates).
Pilot 4	Data is collected via survey, interviews, work products, and from system interaction log files. The participants in pilot 4 are teachers.

[1] Roscoe, R. D., Branaghan, R. J., Cooke, N. J., & Craig, S. D. (2018). Human systems engineering and educational technology. In End-user considerations in educational technology design (pp. 1-34). IGI Global.

State the expected size of the data (if known)

It is unrealistic to gauge the volume of the data at this stage of the project but this section provides an initial estimate regarding the size of the data collected. The information presented in Table 7 below is based on RoPA data set references.

Table : Sharing of Datasets and expected size of data per pilot

Activity	**Data set reference & name**	**Data users (Stakeholders' Groups)**	**File size/data size**	**Access Tools**	**Sharing platform/ method**	**Restricted access/Licence[1]**
Pilot 1	Anonymised tracking timestamps, using app, interactions with AR/3D objects, click counts for exercises, quizzes	Teachers	Normal/MBs	xAPI	Learning Locker	CC-BY-NC
Data collected in student pre & post tests on literacy attainment	Students	Normal/MBs	Paper-based or online questionnaire procedure / commonly used software	Online Open Access repository (ARGOS & ZENODO Project community)	CC-BY-NC	
Data collected through questionnaires	Teachers	Normal/MBs	online questionnaire procedure / commonly used software	Online Open Access repository (ARGOS & ZENODO Project community)	CC-BY-NC	
Pilot 2	Anonymous tracking of time spent within an app, interactions with 3D objects, amount of clicks	Students	Normal/MBs	Learning Locker	Learning Locker	CC-BY-NC
Data collected in student pre, post and retention test on STEM knowledge, motivation ect.	Students	Normal/MBs	Paper-based or online questionnaire procedure / commonly used software	Online Open Access repository (ARGOS & ZENODO Project community)	CC-BY-NC	

Data collected through questionnaires	Teachers	Normal/MBs	online questionnaire procedure / commonly used software	Online Open Access repository (ARGOS & ZENODO Project community)	CC-BY-NC	
Pilot 3	Data collected through questionnaires on the impact of the intervention on students' positive behavioral management and self-management skills	Students and teachers	Normal/MBs	Online and paper-based questionnaire procedure	Online Open Access repository (ARGOS & ZENODO Project community).	CC-BY-NC
Tracking of interactions with AR objects	Students, teachers	Normal/MBs	xAPI	csv files distributed in an Online Open Access repository.	CC-BY-NC	
Pilot 4: ARETE Moodle	AR Learning Experiences (ARLEM)	Students, teachers	Normal/MBs	ARETE Moodle / ARETE market place	Release on ARETE market place (metadata and linked P1589-2020 data file)	CC-BY-NC
Pilot 4: ARETE Moodle	AR Items/assets	Public	MBs/GBs	ARETE Moodle, ARETE market place	Embedded in above	CC-BY-NC
Pilot 4: Survey	Data collected through pre/post questionnaires	Public	Normal/MBs	Statistics software (e.g. R, Excel, SPSS)	Online Open Access repository (ORDO)	CC-BY-NC
Pilot 4: system logs	Tracking of interaction with learning content	Public	MBs/MBs	Statistical software (R, SPSS, Excel); xAPI learning record store (e.g. learning locker)	ORDO	CC-BY-NC
Pilot 3: AR for PBIS	AR Items/assets for PBIS	Public	Normal/MBs	3d model viewer software	ARETE Digital repository	CC-BY-NC

All pilots: Dissemination	Stakeholder events	D2.3 Table 8	Normal/MBs	Commonly used software	Release (metadata available), access to data upon request	CC-BY-NC
All Pilots: Dissemination	Stakeholder Groups' feedback (ARETE platform)	D7.1 Table 5	Normal/MBs	Commonly used software	Release (metadata available), access to data upon request	CC-BY-NC

[1] https://creativecommons.org/licenses/

Outline the data utility: to whom will it be useful

The development of a main platform of the ARETE ecosystem and the associated large number of data items that can be collected, may also trigger a rise in new research directions, e.g. autonomy research based on intelligent data mining and big data methodologies across different domains and algorithms for learning and predictive analysis. Data collected during the pilots will be used to assess the effectiveness of each ARETE solution, compared to traditional teaching solutions, in enhancing students' learning across different subjects, and specifically English literacy, STEM, and behavioural and self-management skills.

Table : Data utility per pilot

Pilots	Data Utility
Pilot 1	Pilot 1 will be analysing the data from the ProfilED Case History forms given to the intervention group, the non-intervention 'control' group and students in the Case history 'control' group (these students are not presenting with any known developmental or learning issues). In conjunction with the expertise of the team in UNW, the data generated from the case history forms and the pre and post NARA II and Vernon test forms will be analysed. Data is collected for every child participating in the Pilot 1 project. The case history form cannot be traced back to the sender and WWL will only have access to a student code and data. Data from assessments of reading accuracy, reading comprehension, reading speed and spelling will be collected from the test forms and analysed by the team in Uni Würzburg.
Pilot 2	The data collected in pilot 2 and their analysis and evaluation are expected to be useful for two target groups in particular. First of all, they provide a basis for relevant and innovative research output such as publications and conference proceedings. After publication, other researchers can follow up on the results and advance further research. This way, the ARETE results contribute to a relevant research objective and pave the way for ongoing educational and HCI research. This way, the second target group of educational addressees can be reached: ultimately, teachers will be supported by an established and validated AR-enhanced approach for teaching STEM, and primary school students will benefit from an innovative and effective new method for teaching and learning, if the piloting can confirm the desired outcomes. This applies to an international audience, i.e., teachers and students from Europe and beyond.
Pilot 3	All consortium partners have direct interest in making all data available for analysis and publication for the goals and aims mentioned in the project proposal. In theory all data could also be interesting *after publication* for, amongst others, researchers outside of the consortium (e.g., behavioral and educational researchers) and educational professionals. The data will be made available for third parties as mentioned below. All data collected could in theory be used towards research purposes and towards practical application purposes.
Pilot 4	Most of the research data in pilot 4 are primarily of interest to researchers, investigating usage patterns and experiences with the AR learning experience authoring tool. The evidence gathered of use, however, informs the wider public, teachers, learners, parents, and policy makers. Data will be deposited for public access after publication, and in deitemised form. The work outputs of the teacher participants are AR learning experiences. We will encourage the teachers to release the produced units to the wider public through the ARETE XR learning experience market place.

2.1 Making data findable, including provisions for metadata [FAIR data]

Outline the discoverability of data (metadata provision)

Project consortium intends to share ARETE data in the publicly accessible ARETE digital repository using descriptive metadata as required/provided by that repository. The repository will assign DOIs for clear identification and citable datasets. Additional metadata of the dataset will be offered within a separate format in a standardised way by using the suitable schema. Files and folders will be organised in a hierarchical and clear structure. Files will be uniquely identifiable and versioned by using a name convention consisting of project name, dataset name, method used, ID, place and date.

Outline the discoverability of data (metadata provision)

IEEE P1589 (IEEE Approved Draft Standard for an Augmented Reality Learning Experience Model: Recently the Standards Board (SASB)[1] of the IEEE standards association, accepted the draft of the P1589[2] working group on the specifications for an 'Augmented Reality Learning Experience Model' (ARLEM). The official final, edited version (punctuation-level edits only) is scheduled to be released on April 27. ARETE DMP will be updated in Month 12 with further details.

IEEE P1589 defines two data models and their binding to XML and JSON for representing learning activities (also known as employee tasks and procedures) and the learning environment in which these tasks are performed (also known as the workplace). The interoperability specification and standard is presented in support of an open market where interchangeable component products provide alternatives to monolithic Augmented Reality-assisted learning systems. Moreover, it facilitates the creation of experience repositories and online marketplaces for Augmented Reality-enabled learning content. Specific attention was given to reuse and repurposing of existing learning content and catering to 'mixed' experiences combining real world learner guidance with the consumption (or production) of traditional contents such as instructional video material or learning apps and widgets. P1589 is further discussed in section 2.1.6. Moreover in ARETE, Dublin Core vocabulary is suggested to use for defining data (metadata provision)

ISO9241[3] focuses on the objective of designing and evaluating systems, products and services for usability is to enable users to achieve goals effectively, efficiently and with satisfaction, taking account of the context of use. This document explains how usability can be interpreted in terms of user performance and satisfaction, and emphasizes that usability is dependent on the specific circumstances in which a system, product or service is used.

Within ARETE, ISO9241 will be interpreted within the context of the three pilots, and each one will be evaluated based on the definition of usability: "the extent to which a system, product or service can be used by specified users to achieve the specified goals with effectiveness, efficiency and satisfaction in a specified context of use". Effectiveness corresponds to the accuracy and completeness with which specified users can achieve specified goals in particular environments. Efficiency calculates the resources expended in relation to the accuracy and completeness of goals achieved and finally, satisfaction measures the comfort and acceptability of the work system to its users and other people affected by its use.

Dublin Core[4] comprises 15 "core" metadata elements. It is one of the simplest and most widely used metadata schema. The name "Dublin" is due to its origin at a 1995 invitational workshop in Dublin, Ohio, nothing to do with Dublin, Ireland unfortunately. Originally developed to describe web resources, Dublin Core has been used to describe a variety of physical and digital resources.

Built into the Dublin Core standard are definitions of each metadata element that state what kinds of information should be recorded where and how. Associated with many of the data elements are suggested controlled vocabularies.

[1] https://www.sasb.org/governance/standards-board/

[2] https://standards.ieee.org/content/ieee-standards/en/standard/1589-2020.html

[3] https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en

[4] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

Outline the identifiability of data and refer to standard identification mechanism. Do you make use of

persistent and unique identifiers such as Digital Object Identifiers?

A persistent identifier (PI or PID)[1] is a long-lasting reference to a document, file, web page, or other object. The term "persistent identifier" is usually used in the context of digital objects that are accessible over the Internet.
Examples of PIDs include:

- Virtual International Authority File (VIAF)
- ORCID IDs
- International Standard Name Identifier (ISNI)
- International Standard Book Number (ISBN)
- Digital Object Identifier (DOI)

In the ARETE project we will make use of persistent and unique identifiers for uniquely identifying and defining entities. Some of which are mentioned above and will be used e.g DOI's for publications and articles and ORCHID IDs for researcher's unique identity.

[1] https://en.wikipedia.org/wiki/Persistent_identifier#cite_note-1

Outline naming conventions used

Consistent naming conventions will be used in the ARETE project for defining the files, folders and objects.

File Naming Guidelines
In ARETE data files and folders would be labelled and organised in a systematic and consistent way so that they are easy to find for the research team. As there is no recommended way to name files and folders but consistency is the key therefore in ARETE we suggest following Elements of a filename:

- A project acronym
- Content description
- File type information
- Date (YYYY-MM-DD)
- Creator name or initials
- Version number
- Status info, e.g. draft

As operating systems usually default to sorting files alphabetically, it is helpful to define what comes at the start of the file name. Therefore in ARETE, it is more useful to order the files by date, by author, or by subject.
The benefit of consistent naming of data files in ARETE is that it will be easier to identify all files connected to one data collection event (e.g. one interview session). In pilots, the files related to one collection event (e.g. audio tape, its transcription etc that were taken by the interviewee) can be connected by the file name.
Example:

- wav
- txt
- jpg

Folder Structure
Similar to consistent file naming conventions, ARETE is using a meaningful folder structure to make it much easier for partners and contributors to locate relevant documents. In ARETE, shareable content among consortium partners is structured in folders and subfolders using Google Drive which is only accessible to members of the ARETE consortium. All the information relevant to data collection and data processing procedures, as well as any other relevant data management material will be stored in dedicated folders. As a guideline ARETE also restricts the level of folders to three or four deep and to a maximum of not to have more than ten items in each list.

Outline the approach towards search keyword

Making data Interoperable (FAIR) can be achieved by using suitable standards for data and metadata creation. This can be achieved by making use of appropriate and suitable vocabularies (e.g. for providing search keywords), in order to optimize the data re-use by other interested parties. Descriptive metadata will be utilised and will contain information such as the title, author, date, description, location, acquisition information, which will be of utmost importance in search and keyword search in the ARETE project. At the end of the project, the data and tools will be available through a digital repository which will also facilitate keyword search.

There are a number of freely available tools that can be customised to implement the keyword search in ARETE project and some of those are listed ass follow:

- Google Ads Keyword Planner
- Google Suggest
- Bing Ads Keyword Planner

Outline the approach for clear versioning

Managing different versions of ARETE data/ software can be achieved by following the guidelines described below:

- Uniquely identifying different versions of files using a systematic naming convention, such as using version numbers or dates
  - Record the date within the file: 20190902_documentation_for_my_data
  - Include a version number in the file name: Documentation_v2
  - Include information about the status of the file, e.g. "draft" or "final," as long as you don't end up with confusing names like "final2" or "final_revised".
  - Include information about what changes were made, e.g. "cropped" or "normalised".
- Using version control facilities within the software used (e.g apache subversion, git etc)
- Using file-sharing services with incorporated version control
- Designing and using a version control table (example as follow)

Table : Version control naming convention

File name	Changes to file
ARETE_DMP_1.0	Original document
ARETE_DMP_1.1	Minor revisions made
ARETE_DMP_1.2	Further minor revisions
ARETE_DMP_2.0	Substantive changes

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

IEEE P1589 (IEEE Approved Draft Standard for an Augmented Reality Learning Experience Model) [1]

Scope

The proposed Augmented Reality (AR) learning experience model will specify how to represent learning activities and their according workplace reference models in a standardized interchange format in order to lower entry barriers for authoring of learning experience spanning real world interaction using sensors and computer vision, and web applications. ARLEM allows wrapping learning activity data for each content unit produced into a standard format for interchange, so that the data can be transported to and from the participating systems, and from provider to user.

Purpose

This standard for Augmented Reality Learning Experience Models (ARLEM) provides an overarching integrated

conceptual model and the corresponding data model specifications for representing activities, learning context and environment (aka 'workplace'), while linking with other data model components needed for AR-enhanced learning activities. The standard distinguishes slow-changing data for environment description from fast-changing data for step-by-step guidance. It defines the required data models and modeling languages and their bindings to XML and JSON. The purpose of this standard is to support the discovery, retrieval, transfer, and execution of AR-enabled learning content, thereby facilitating the creation of repositories and online marketplaces.

The standard supports reuse and repurposing of existing (learning) content in 'mixed' experiences that combine real-world guidance with traditional media such as instructional video material or existing web applications and widgets.

This new IEEE Standard (draft) promises to provide significant boosts in operational efficiency by making information available to employees needing task support in context in real time. To support implementations of AR training systems, this standard proposes an overarching integrated conceptual model that describes interactions between the physical world, the user, and digital information, the context for AR-assisted learning and other parameters of the environment. It defines two data models and their binding to XML and JSON for representing learning activities (also known as employee tasks and procedures) and the learning environment in which these tasks are performed (also known as the workplace). This interoperability specification and standard is presented in support of an open market where interchangeable component products provide alternatives to monolithic Augmented Reality-assisted learning systems. Moreover, it facilitates the creation of experience repositories and online marketplaces for Augmented Reality-enabled learning content. Specific attention was given to reuse and repurposing existing learning content and catering experiences combining real world learner guidance with the consumption (or production) of traditional contents such as instructional video material or learning apps and widgets.

The consortium anticipates the creation of new standards following the development of the Pilot studies, the ARETE platform and the ARETE digital repository, which has led to the award of the best research project in 2022 CEN/CENELEC for standards and innovation.

[1] https://www.techstreet.com/ieee/standards/ieee-p1589?gateway_code=ieee&vendor_id=6073&product_id=1951311#full

2.2 Making data openly accessible [FAIR data]

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

In ARETE we will follow EC Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020,[1] and use additional resources and guidelines provided by OpenAIRE.eu and the Digital Curation Centre's DMP online tool.

All data will be suitable for re-use and will be made available at the end of the project in month 42. Data here refers to final raw data, analysis scripts, and final processed data—are stored together with an instruction file (user manual) and code book file (list of variable names and meanings) as a ZIP-file.

[1]Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, March 2017 update: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Specify how the data will be made available

At the end of the project the data will be openly available through a digital repository accessible via ARETE official website: https://www.areteproject.eu/

Specify what methods or software tools are needed to access the data? Is documentation about the

software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

Table : Methods/Software tools needed to access the data per pilot

Pilots	**1) Methods/ Software tools to access data** **2) Documentation about the software needed to access the data included** **3) Is it possible to include the relevant software (e.g. in open source code)**
Pilot 1	The data from the online ´ProfilED´ case-history form is held in PDF format. To open these files, users will need ***Adobe Reader software*** on ***Macintosh OS X*** or ***Windows*** systems. Additionally, PDF files can be opened with Preview on ***Macintosh OS X***. Users who do not have Adobe Reader already installed on their computer, can download the installer and instructions free from the ***Adobe Web site***. Only the anonymized analytical data arising from the information contained in these forms will be available through the digital repository. Quantitative data from pilot 1 surveys and from NARA II and Vernon student tests will be accessed and processed by ***MS Excel*** and ***SPSS*** “Statistical Package for the Social Sciences”. Online teacher surveys and the student motivational scale will be administered by SurveyMonkey and exported to SPSS files. Qualitative data from focus groups will be analyzed by a qualitative content analysis using ***MAXQDA*** and ***MS Excel***. After analysis, all data will be compiled and presented using MS office products such as MS Word or MS PowerPoint. All of these software tools are professional applications that require respective licenses, which are available for the researchers responsible.
Pilot 2	Quantitative data from pilot 2 surveys and the student TIMSS knowledge tests will be accessed and processed by ***MS Excel*** and ***SPSS*** “Statistical Package for the Social Sciences”. Online teacher surveys will be administered by SurveyMonkey and exported to SPSS files. Qualitative data from focus groups will be analyzed by a qualitative content analysis using ***MAXQDA*** and ***MS Excel***. After analysis, all data will be compiled and presented using MS office products such as MS Word or MS PowerPoint. All of these software tools are professional applications that require respective licenses, which are available for the researchers responsible.
Pilot 3	Behavioural traces will be converted from the raw format (xAPI statements, noSQL triples) to aggregate tables, thereby de-referencing records as necessary to ensure privacy. The resulting aggregate tables will be released as .csv / .tsv or similar. Analysis scripts are written (no code, just analytic procedures) so that others are free to use whatever software they have available to them to analyse the data (e.g., ***R, SPSS, STATA***).

Pilot 4	Behavioural traces will be converted from the raw format (xAPI statements, noSQL triples) to aggregate tables, thereby de-referencing records as necessary to ensure privacy. The resulting aggregate tables will be released as .csv / .tsv or similar. Analysis scripts are written (no code, just analytic procedures) so that others are free to use whatever software they have available to them to analyse the data (e.g., ***R, SPSS, STATA***). Quantitative data from pilot 4 surveys will be accessed and processed by ***MS Excel*** and ***SPSS*** "Statistical Package for the Social Sciences". Online teacher surveys will be administered by a suitable online survey service such as Google Forms or SurveyMonkey and exported to Excel files. Qualitative data from focus groups will be analyzed by a qualitative content analysis using ***MAXQDA*** and ***MS Excel***. After analysis, all data will be compiled and presented using MS office products such as MS Word or MS PowerPoint. The work products that are made available to the public by the participants that opt-in are contained in zip archives containing **P1589-2020 AR learning experience models**. This can be downloaded and viewed in the **MirageXR** application on iOS, Android, Hololens 1, and Hololens 2 (with a port to ThinkReality A3 in preparation).

Specify where the data and associated metadata, documentation and code are deposited

In the ARETE project a number of available tools can be considered for depositing data, associated metadata, documentation and code. Some of the famous tools that can be used are listed as follow:

- Github
- Apache Subversion
- Google Drive
- Digital Repository

Git[1] is a distributed version-control system for tracking changes in source code during software development. It is designed for coordinating work among programmers, but it can be used to track changes in any set of files. Its goals include speed, data integrity, and support for distributed, non-linear workflows.

Apache Subversion[2] (often abbreviated SVN, after its command name svn) is a software versioning and revision control system distributed as open source under the Apache License. Software developers use Subversion to maintain current and historical versions of files such as source code, web pages, and documentation. Its goal is to be a mostly compatible successor to the widely used Concurrent Versions System (CVS).

Google Drive[3] is a file storage and synchronization service developed by Google. Google Drive allows users to store files on their servers, synchronize files across devices, and share files. In addition to a website, Google Drive offers apps with offline capabilities for Windows and macOS computers, and Android and iOS smartphones and tablets. Google Drive encompasses Google Docs, Google Sheets, and Google Slides, which are a part of an office suite that permits collaborative editing of documents, spreadsheets, presentations, drawings, forms, and more. Files created and edited through the office suite are saved in Google Drive.

Digital repository[4] is an online database of digital objects that can store content e.g., text, documents, still images, audio, video, or other digital media formats. Digital objects can consist of digitized content like print or photographs, as well as originally produced digital content like word processor files or social media posts. In addition to storing content, digital repositories provide means for organizing, searching, and retrieving the content contained in the collection. The Open University maintains Open Research Data Online, ORDO, an online

repository for research data publication.

[1] https://en.wikipedia.org/wiki/Git
[2] https://en.wikipedia.org/wiki/Apache_Subversion
[3] https://en.wikipedia.org/wiki/Google_Drive
[4] https://en.wikipedia.org/wiki/Digital_library

Specify how access will be provided in case there are any restrictions

- ARETE partners will limit access to authorized and identified persons. Limit access to hard copy human subjects data to authorized personnel, and institute a sign-out, sign-in procedure.
- It is envisioned that the documents and codes generated remain restricted and access will be granted only to the authorised users amongst the partners.
- In pilot 3, an extra CSV file will be created in which user identification numbers can be linked over time and datasets. This key file will be stored password-protected and coded on a password-protected computer which is not linked to the safe VU server. Immediately after finalizing data collection and data linking, this key-file will be destroyed.
- At the end of the project final data and tools created will be made publicly available through the ARETE Digital Repository.

2.3 Making data interoperable [FAIR data]

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

To ensure interoperability and to improve data search and citation, the partners will apply data documentation standards and metadata. For all survey quantitative data, CSV format will be used. For all qualitative textual data (interview minutes and other), the PDF format will be used. The following media formats will be used: images, audio, video, and 3D models.

Interoperability of the ARETE data is a priority for this project and the consortium has extensive experience in metadata and vocabularies[1],[2] as fundamental components in the description of the learning resources within all three pilots. Descriptive metadata in educational augmented reality repositories will conform either to Dublin Core or IEEE LOM. The consortium anticipates the area of Augmented reality educational vocabulary will compose a new segment in the metadata vocabularies and standards, especially given the need to link those with the curriculum resources and the provision mechanisms.

As mentioned before (section 2.1.1 and section 2.1.6) that in ARETE "**Standard for an Augmented Reality Learning Experience Model**" and Dublin Core vocabulary is suggested to use for defining data. Using the aforementioned standard and established vocabulary guarantees the interoperability of data. Standards for Augmented Reality Learning Experience Models (ARLEM) provides an integrated conceptual model and the corresponding data model specifications for representing activities, learning context and environment (aka 'workplace'), while linking with other data model components needed for AR-enhanced learning activities. The standard distinguishes slow-changing data for environment description from fast-changing data for step-by-step guidance. It defines the required data models and modelling languages and their bindings to XML and JSON. The standard supports reuse of existing content (interoperability) in 'mixed' experiences that combine real-world guidance with traditional media such as instructional video material or existing web applications and widgets.

Dublin Core[3] comprises 15 "core" metadata elements. It is one of the simplest and most widely used metadata schema. Built into the Dublin Core standard are definitions of each metadata element that state what kinds of information should be recorded where and how (Table 3). All elements are optional and repeatable.

Table : Dublin Core defining concepts and element in AR

Name	Data type	Data source (data format)	Metadata details
	Text	Survey data (CSV), interviews minutes (PDF)	Dublin Core Schema; For survey data: variable labels, code labels, and defined missing values (if needed)
AR Learning Experiences	Text, Code	High-level content description data: modelling language files (JSON) packaged in archives (ZIP)	Dublin Core Schema; or direct access via the ARETE digital repository.
Behaviour and impressions data from experience capturing	Text, numeric data	Survey data (CSV), interviews minutes (PDF)	Dublin Core Schema; For survey data: variable labels, code labels, and defined missing values
Stakeholder community events	Text, media data and multimedia	Survey data (CSV), interviews minutes and other feedback (PDF), images (JPEG), audio (WAV), video (MPEG-4)	Dublin Core Schema;

IEEE 1484.12.3-2005[4] - IEEE Standard for Extensible Markup Language (XML) Schema Definition Language Binding for Learning Object Metadata (LOM) defines a World Wide Web Consortium (W3C) Extensible Markup Language (XML) Schema definition language binding of the learning object metadata (LOM) data model defined in IEEE 1484.12.1-2002 Standard for Learning Object Metadata. The purpose of this Standard is to allow the creation of LOM instances in XML. This allows for interoperability and the exchange of LOM XML instances between various systems. This Standard uses the W3C XML Schema definition language to define the syntax and semantics of the XML encodings.

Furthermore, we have provided a Moodle plugin to enable content creators and teachers with capabilities to describe in a semi-automatic way their learning resources with LOM-based metadata and make these metadata available to search service providers so that other stakeholders can easily find and retrieve them. It was composed to support reusability and discoverability of 3D/AR and other types of educational resources included in the project's Moodle digital repository. Aiming on utilising previous work on this domain, the code of two existing plugins was modified and enriched to serve the project's needs. The two plugins that will be utilised in ARETE's Moodle digital repository to support the discoverability of learning resources by creating and exposing their metadata to make them available for harvesting. The content in the ARETE repository is particularly relevant to 3D/AR learning activities created through an XR authoring toolkit. Nevertheless, educational content in other formats continues to be supported by the aforementioned plugins. The integration of IEEE-LOM and OAI-PMH standards to a Moodle repository seems to be a feasible way to enhance the development of learning content by utilising relevant already existing resources that can be easily found and retrieved. However, the difficulty of finding service providers that could support the collection of learning resource metadata and be willing to build search engines on top of them suggests the need to consider different approaches.

[1] http://lre.eun.org/edrene/publications/756.pdf

[2] https://ieeexplore.ieee.org/document/8346266

[3] https://dublincore.org/

[4] https://standards.ieee.org/standard/1484_12_3-2005.html

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

Standard vocabulary “Dublin core” will be used for all data types that will allow inter-disciplinary interoperability. Dublin Core metadata terms are expressed in RDF vocabularies for use in Linked Data. Creators of non-RDF metadata can use the terms in contexts such as XML, JSON, UML, or relational databases by disregarding both the global identifier and the formal implications of the RDF-specific aspects of term definitions. Such users can take domain, range, subproperty, and subclass relations as usage suggestions and focus on the natural-language text of definitions, usage notes, and examples.

Each term is identified with a Uniform Resource Identifier (URI), a global identifier usable in Linked Data. Term URIs resolve to the (DCMI Metadata Terms) document when selected in a browser or, when referenced programmatically by RDF applications, to one of four RDF schemas. The scope of each RDF schema corresponds to a "DCMI namespace", or set of DCMI metadata terms that are identified using a common base URI, as enumerated in the DCMI Namespace Policy. In Linked Data, the URIs for DCMI namespaces are often declared as prefixes in order to make data, queries, and schemas more concise and readable.

It is worth noting that AR/VR is an emerging field and relevant standards and vocabularies are establishing and emerging as well. As mentioned that the current plan is to use “Dublin Core” but if there will be a need for using other vocabularies and ontologies the mappings will be defined and provided during the later versions of DMPs.

2.4 Increase data re-use (through clarifying licenses) [FAIR data]

Specify how the data will be licenced to permit the widest reuse possible

The ARETE project will share as much data as possible. It is currently understood that the factors which could prevent research data from being shared (ethical, rules of personal data, intellectual property, commercial, privacy-related, and security-related) will not apply in the ARETE project. If some data are to be licensed and released, the project will use Creative Commons BY-NC-ND license[1] - an attribution to allow reproduction and sharing online[2]. Whereas, some parts and tools are under licenses which might be copyrighted due to background IPs involved by the pilot partners.

[1] https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode

[2] Ball, A. (2014). ‘How to License Research Data’. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/how-guides

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

There is no embargo on the data and the tools produced in the ARETE project and all the data, documentation and tools will be made available after the completion of the project.

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

Table: Third party usage of data per pilot]

Pilots	Third party usage of the data
Pilot 1	Research data collected in pilot 1 concerning information collected from teachers, parents and children will be re-usable for third parties in an anonymized format in accordance with the FAIR data approach, as agreed by the project consortium.
Pilot 2	Data collected in pilot 2 will be re-usable for third parties in an anonymized format in accordance with the FAIR data approach, as agreed by the project consortium. Additionally, other researchers may refer to ARETE publication/dissemination output and advance respective research.
Pilot 3	All data collected in pilot 3 will be used to realize publications in academic and popular psychological journals, as well as for poster and oral presentation at conferences. Most of the data is suitable for re-use and will be made available after the first publication of the data. This data refers to final raw data, analysis scripts, and final processed data—are stored together with an instruction file (user manual) and code book file (list of variable names and meanings). Due to the longitudinal nature of the pilot, the data for pilot 3 will be stored pseudo-anonymously. All consortium partners have direct interest in making all data available for analysis and publication for the goals and aims mentioned in the project proposal. In theory all data could also be interesting after publication for other researchers outside of the consortium (e.g., behavioral and educational researchers).
Pilot 4	All data collected in pilot 4 will be used to realize publications in academic and popular psychological journals, as well as for poster and oral presentation at conferences. Most of the data is suitable for re-use and will be made available after the first publication of the data. This data refers to final raw data, analysis scripts, and final processed data—are stored together with an instruction file (user manual) and code book file (list of variable names and meanings). All consortium partners have direct interest in making all data available for analysis and publication for the goals and aims mentioned in the project proposal. In theory all data could also be interesting after publication for other researchers outside of the consortium (e.g., behavioral and educational researchers).

Describe data quality assurance processes

In the ARETE project, the data quality will be ensured by different measures. These include validation of the data, replication, comparison with results of similar studies etc. In order to safeguard the transparency, consistency, comparability, completeness and accuracy of the data, ARETE will apply Quality Assurance and Quality Control activities in the form of peer reviews, data summaries, and input data checks.

In ARETE we consider 5 Pillars of Data Quality Management[1]:

- The People
- Data Profiling

    - Reviewing data in detail
    - Comparing and contrasting the data to its own metadata
    - Reporting the quality of the data
- Defining Data Quality
- Data Reporting
- Data Repair

Moreover, in ARETE we will define the data quality metrics based on some of the following:

1. Accuracy
2. Consistency
3. Completeness
4. Integrity
5. Timeliness

[1] https://www.datapine.com/blog/data-quality-management-and-metrics/

Specify the length of time for which the data will remain re-usable

As open formats are used for data archiving, the data will remain re-usable until the repository withdraws the data or goes out of business. As defined in the project proposal, the data remains available and reusable up to 5 years after the project ends in Month 42. The project will also archive the research data within Open Research Europe and share with the H2020 ARETE project community in ZENODO. At this stage, it is difficult to estimate the volume of data collected from all four pilots. For example, given the alternative scenarios that have been considered due to the impact of COVID-19, if some video data is required, the volume of data will increase significantly. Currently estimated sizes for each data set are given above (Table 7: Sharing of Datasets and expected size of data per pilot). The ARETE Data repository (ARETE Moodle) archiving will be done at University College Dublin (UCD), and the associated costs will be covered by UCD.

3. Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

Estimated costs for data preparation consist of additional publication and documentation costs of three months for one full time equivalent and publication costs of the repository. Associated costs for dataset preparation and data management during the project will be covered by the project itself.

We will do a full cost benefit analysis and in order to estimate the exact cost of making data FAIR in ARETE by M42, and we will estimate cost (in terms of time, money and effort) in[1],[2]:

- creating and collection of data
- pre-processing and data cleansing
- integration of the data
- analysis of the data
- registration and publication of the data
- peer review

[1] https://op.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1

[2] https://www.rd-alliance.org/how-expensive-fair-compliance-and-how-expensive-it-not-be-fair-compliant-rda-11th-plenary-bof

Clearly identify responsibilities for data management in your project

As mentioned in D 1.2 “POPD - Requirement No. 2”, the appointed Data Protection Officer (DPO) for the entire ARETE project will be the coordinator institution DPO (NUID UCD) and the contact details of the UCD DPO will be made available to all data subjects involved in the research. For host institutions not required to appoint a DPO under the GDPR a detailed data protection policy for the project will be submitted for full ethics approval from the UCD Office of Research Ethics. The project coordinator’s institution (NUID UCD) DPO, will act as an independent advocate for the proper care and use of personal data processed by the consortium. Figure 1 below denotes the structure defined within the ARETE consortium for the management of the data produced within the three ARETE pilots with specific roles and responsibilities for each partner.

Each partner member, as shown in Table 4, (DPO/Data Protection Officer/Data Protection Coordinator/Information Assurance Services Manager) will advise the members of its organisation on how to operate in line with Data Protection Principles. This will include providing members of the organisation within the ARETE Consortium with feedback on their Data Protection Impact Assessment (DPIA), where appropriate and if requested from NUID UCD DPO. DPIAs can be used to identify and mitigate against any data protection related risks arising from the ARETE project, which may affect the organisation or the individuals it engages with. As the pilots coordinator (WP6), EUN’s Data Protection Coordinator will advise on what type of documentation is needed by an organisational unit participating in the pilots to demonstrate GDPR compliance, including Records of Processing Activities (ROPAs) in line with Article 30.

Table : Consortium partners’ DPOs/Data Protection Coordinator/Information Assurance Services Manager

No	Acronym	Contact	Role
1	NUID NUID NUID UCD	gdpr@ucd.ie (Interim DPO: Ulrike Kolch)	DPO
2	CLB	inna@cleverbooks.eu (Inna Armstrong)	DPO
3	WWL	rita@wordsworthlearning.com (Rita Treacy)	DPO
4	SVU	functionarisgegevensbescherming@vu.nl (Hagenauw, D.E.)	DPO
5	ULE	ias@le.ac.uk (Parmjit Singh Gill)	Information Assurance Services Manager
6	EUN	privacy@eun.org (John Stringer)	Data Protection Coordinator
7	CNR	rpd@cnr.it (Dott. Raffaele Conte)	DPO
8	UNW	datenschutz@uni-wuerzburg.de (Behördlicher Datenschutzbeauftragter der Universität Würzburg)	DPO
9	VIC	pi@vicomtech.org (Nora Gurrutxaga)	DPO
10	OBU	BrookesDPO@brookes.ac.uk (Information Compliance Team)	DPO
11	TOU	data-protection@open.ac.uk Post: The Data Protection Officer, PO Box 497, The Open University, Walton Hall, Milton Keynes MK7 6AT.	DPO
12	UDUR	info.access@durham.ac.uk (Kristina Holt)	DPO and Head of Information Governance

Describe costs and potential value of long term preservation

The relevant costs for the activities of the data management are included within the allocated resources within the ARETE project budget and are associated with the following activities:

- creating and collection of data through the Pilots and the ARETE Moodle digital repository.
- pre-processing and data cleaning, integration of the data and analysis of the data, through the partners involved in the evaluation process of the project
- registration and publication of the data through the utilisation of standard tools (Moodle repository hosted from the coordinator – UCD) and a free of charge research data repository (ZENODO). The long-term preservation for the public research data is guaranteed via the ZENODO Community portal and peer review of the publications through the Open Research Europe[1], which is the responsibility of Ms Esther Novo (VICOMTECH).

The Pilot managers decides what data will be kept and for how long. The overall responsibility for data management lies with the project coordinator, Prof. Eleni Mangina, UCD. Supporting the coordinator is a data management team consisting of the Data Controllers for each pilot as shown in the RoPA document of the ARETE project.

[1] https://open-research-europe.ec.europa.eu/

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

Data collected/ generated during the ARETE project will be dealt with in a safe and secure manner. All the participants from ARETE consortium involved in data collection and generation will be responsible for dealing with the data in a secure manner. The data is safely stored in certified repositories for long term preservation and curation.

Table : Data security measures per pilot

Pilot	Data Security measures and approaches

Pilot 1	Student test data are collected by the teachers and the paper tests remain with the teachers who are instructed to destroy them. Pseudonymized scans are sent to EUN via a safe data upload server and then, after double checking the pseudonymization, shared with UNW for analysis via a safe data upload server. At UNW, all data is stored on a safe and password protected UNW server. EUN will delete all data upon successful delivery to UNW. For the collection of ProfilEd data, WWL uses an interactive webform (***HTML format***) for gathering personal, clinical and educational data from a parent/guardian. They will enter information about themselves and their child, which is retained on our secure server for analysis and processing. The webform code is embedded on the WWL website (***HTTPS***) and there are several field-entry types such as multiple choice, drop-down menus, checkboxes and radio buttons. There are also short or long-text entry fields for questions that require detailed answers. When a parent clicks the “Submit” button, WWL will receive email notification(s) containing the webform. The data is protected on a secure website that is backed up on our server every 24 hours, and also securely backed up to an external hard drive that is stored offsite in a fireproof safe. The interviews with teacher coordinators are recorded by EUN and transferred to UNW for analysis. They are deleted after transcription and analysis and processed in fully anonymized format. Data from SurveyMonkey (student motivation scale, teacher survey) are password protected and accessed by qualified UNW staff only. They are exported from SurveyMonkey in SPSS format and stored on a safe and password protected UNW server, pseudonymized by non-identifiable IDs. The data are shared open source after analysis and publication in fully anonymized format only. Access to SurveyMonkey’s technology resources is only permitted through secure connectivity (e.g., VPN, SSH) and requires multi-factor authentication. The production password policy requires complexity, expiration, and lockout and disallows reuse. SurveyMonkey grants access on a need to know on the basis of least privilege rules, reviews permissions quarterly, and revokes access immediately after employee termination[1]. Finally, The Learning Record Store deployed by VIC and hosted in AWS uses secure http and role-based access control to protect users and their data.

Pilot 2	Student paper test forms are retained at school where teachers are instructed to destroy them after the end of the project; only scans with an ID code are shared with EUN who then double check the pseudonymization before sharing the scans with UNW via a safe data upload server. At UNW, all data are stored on a safe and password protected UNW server. The teacher coordinator interviews are recorded by EUN and transferred to UNW for analysis. They are deleted after transcription and analysis and processed in fully anonymized format. Data from SurveyMonkey (teacher surveys) are password protected and accessed by qualified UNW staff only. They are exported from SurveyMonkey in SPSS format and stored on a safe and password protected UNW server, pseudonymized by non-identifiable IDs. The data are shared open source after analysis and publication in fully anonymized format only. Access to SurveyMonkey's technology resources is only permitted through secure connectivity (e.g., VPN, SSH) and requires multi-factor authentication. The production password policy requires complexity, expiration, and lockout and disallows reuse. SurveyMonkey grants access on a need to know on the basis of least privilege rules, reviews permissions quarterly, and revokes access immediately after employee termination[2]. Finally, The Learning Record Store deployed by VIC and hosted in AWS uses secure http and role-based access control to protect users and their data.
Pilot 3	In order to link quantitative data of participants over time for pilot 3, an extra CSV file will be created in which user identification numbers can be linked over time and datasets. This key file will be stored password-protected and coded on a password-protected computer which is not linked to the safe VU server. Immediately after finalizing data collection and data linking, this key-file will be destroyed.
Pilot 4	Data collected will be de-itemised at source at the collecting partner. Any data shared will be anonymous. Data sets are not linked. The ARETE hosted Moodle (hosted at UCD) provides state-of-the-art privacy and data security. The web server is using secure http (https). The ARETE marketplace (hosted from UCD) uses moderated self-registration as a safeguard for spam protection. The Learning Record Store (deployed by VIC and hosted in AWS) uses secure http. All three systems use role-based access control to protect users and their data. Research data are exported from these systems as described above. Other than the other three pilots, pilot 4 does not deal with a vulnerable target group (but with teachers).

[1] https://www.surveymonkey.com/mp/legal/security/

[2] https://www.surveymonkey.com/mp/legal/security/

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

As mentioned in D1.1 (H Requirement No.1) and D1.2 (Protection of personal data (POPD) Requirement No. 2), the ARETE consortium acknowledges that ethics is essential in fostering responsible research and innovation within this project. It has been noted that:

*The Horizon 2020 programme has ambitious aims that not only presuppose empirical claims, e.g. with respect to natural resources, available technologies or existing infrastructures, but also set specific goals, e.g. regarding sustainability, health and well-being, food production, the bio-based economy and resource efficiency. These goals show that the research themes are not ends in themselves, but means to realising European commitments to human rights and values. These rights and values presuppose normative claims about how humans should treat one another, what makes for good society and how responsible governments and businesses should behave**[1]**.*

The ARETE project is undertaken with total respect for citizen rights (as required by the Lisbon Treaty – European Charter of Fundamental Rights – which came into force 1st December 2009), as well as societal relevance, fundamental requirements of our conduct[2]. Ethics, being the discipline that provides the tools for systematic reflection on normative claims and questions, plays a central role in all reputable research.

In the ARETE project, it underpins our research in three ways, by: *enabling better research design*; *translating fundamental commitments into research practice*; and, *enhancing debate and building platforms and guidelines to increase public trust and acceptance.*

Ethical considerations

The protection of personal data of students and other end users, in particular, involvement of children/minors and other vulnerable groups, is of high importance to the ARETE consortium. Therefore, international and European norms and regulations, as well as best practices, will be applied during each of the three pilots planned. The process and approach to research ethics in the ARETE project follows four main lines of ethical consideration:

1. Recruitment of participants
2. Confidentiality/anonymity/secrecy
3. Consent/assent
4. Measures taken to ensure the avoidance of research misconduct

These aforementioned four lines of ethical considerations are introduced in detail in D1.1: ”H- Requirement No. 1”.

Furthermore, there are no ethical or legal issues than can have an impact on data sharing. We have data sharing agreements within the consortium where applicable, and there is informed consent for data sharing and long term preservation, information provided to the stakeholders, through the information sheets and consent forms.

[1] http://www.leru.org/index.php/public/news/leru-publishes-an-agenda-for-ethics-research-in-horizon-2020/.

[2] As noted above our research on security technology will in all phases be guided by the ‘Societal impact checklist for security R&D’ as published in the CIES (2012) Report of the Societal Impact Expert Working Group (http://cies.ie/wp-content/uploads/2014/05/Report-of-the-Societal-Impact-Expert-Working-Group.pdf)

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

(For the full version of the DMP including images and diagrams, please refer to the ARETE H2020 community at ZENODO: https://zenodo.org/record/7271431#.Y32bbezP2w0)

Table: Other procedures for data management used per partner within the ARETE consortium

Partner(s)	National/funder/sectorial/departmental procedures
NUID UCD	UCD Data Management Plan ● https://libguides.ucd.ie/data/dmp ● https://www.ucd.ie/research/portal/start/implementdatamanagementplan/ UCD Policies & Procedures ● http://www.ucd.ie/gdpr/policiesprocedures/
SVU	For Vrije Universiteit Amsterdam all research data management is in line with the national scientific guidelines of The Netherlands Organization of Scientific Research (NWO). More information about VU-wide data management can be found here: ● https://ub.vu.nl/en/education-research/research-data-services/data-management-for-researchers/index.aspx
UNW	UNW manages any research data in accordance with established national guidelines issued by Deutsche Forschungsgemeinschaft [DFG; German Research Association]. Further information [partly in German language] can be found at: ● https://www.dfg.de/en/research_funding/proposal_review_decision/applicants/research_data/index.html ● https://econtent.hogrefe.com/doi/pdf/10.1026/0033-3042/a000341 ● https://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf
VIC	● https://vicomtech.app.box.com/s/grhrrtb3sfa431q210ckbnddtlwrypzc
CNR	● https://www.itd.cnr.it/page.php?ID=Privacy&FlagSelected=en
OBU	OBU partner will be replaced by Open University (OU) 1st of June 2020.
ULE	● https://www2.le.ac.uk/institution/ethics
CLB	● https://www.cleverbooks.eu/gdpr/
WWL	ARETE partner WWL manages any research data in accordance with our Company's GDPR standards as outlined in our Data Protection Policy. The WWL GDPR Policy is available on our website: ● www.wordsworthlearning.com at the following link GDPR Policy
EUN	EUN Partnership has a data coordinator who is responsible for giving advice and guidance to the Office on data protection issues. The EUN data protection coordinator is supported by a member of staff who follows-up specific requests from data subjects covering things such as data rectification, requests for deletion of data and data access requests. Each team in the Office has also appointed a data protection ambassador who is responsible for providing advice and training to their own team members overseen by the data protection coordinator. ● http://www.eun.org/about/partners/code-of-practice
TOU	The Open University maintains a portal for research data management: ● http://www.open.ac.uk/library-research-support/research-data-management The research data management policy is accessible here: ● http://www.open.ac.uk/library-research-support/sites/www.open.ac.uk.library-research-support/files/files/Open-University-Research-Data-Management-Policy.pdf
UDUR	The University of Durham data management procedures are outlined in: https://libguides.durham.ac.uk/open_research/rdm