



D7.3 Data management plan

WP7 Project coordination

By Giulia Inguaggiato, Krishma Labib, Laura Hartman, Guy Widdershoven, Edwin Kanters, Marc van Hoof, Ana Marušić, Ružica Tokalić, Vicko Tomić, Ivan Buljan, Ana Utrobičić, Armin Schmolmüller, Nicole Föger

VIRT2UE Consortium



VIRT2UE has received funding from the European Union's Horizon 2020 research program under the grant agreement N 787580



D7.3 Data management plan

Project details	
Project full name	Virtue based ethics and Integrity of Research: Train-the-Trainer program for Upholding principles and practices of the European Code of Conduct for Research Integrity
Project acronym:	VIRT ² UE
Start date of the project:	01.06.2018
Duration:	36 months
Project number:	787580

Deliverables details	
Work Package	WP7 Project coordination
Description of deliverable	D7.3 Data management plan
Work package leader	VUMC
Responsible for the deliverable	Guy Widdershoven
Submission date	30.11.2018
List of contributors	Giulia Inguaggiato, Krishma Labib, Laura Hartman, Guy Widdershoven, Edwin Kanters, Marc van Hoof, Ana Marušić, Ružica Tokalić, Vicko Tomić, Ivan Buljan, Ana Utrobičić, Armin Schmolmüller, Nicole Föger

Table of contents

History and list of contributors	3
Description of deliverable	3
Background.....	4
1 Data summary	4
1.1. Purpose of data collection/generation	4
1.2. Relation to the objectives of the project.....	5
1.3. Types and formats of data generated/collected.....	5
1.4. Origin of the data	6
1.5. Expected size of data.....	7
1.6. Data utility	7
2 Fair data.....	8
2.1 Making data findable.....	8
2.1.1 Discoverability of data.....	8
2.1.2 Identifiability of data	8
2.1.3 Naming conventions.....	8
2.1.4 Approach towards search keyword.....	8
2.1.5 Approach for clear versioning	9
2.1.6 Metadata	9
2.2 Making data openly accessible.....	9
2.2.1 Ethical concerns related to publication.....	9
2.2.2 Methods/software tools needed to access data.....	10
2.3 Making data interoperable.....	10
2.4 Increase data re-use (through clarifying licenses)	10
2.4.1 Data quality assurance processes	10
3 Allocation of resources.....	10
4 Date security.....	11
4.1 Data recovery, security storage and transfer of sensitive data	11
5 Ethical issues.....	12

History and list of contributors

Version	Date	Description	Revised by
0.1		First draft	Edwin Kanters
0.2		Second draft	Work packages
0.5		Document submitted to the Participant Portal	

List of contributors

Giulia Inaguaggiato, WP3 VIRT²UE
Krishma Labib, WP7, VIRT²UE
Laura Hartman WP7, VIRT²UE
Guy Widdershoven WP7, VIRT²UE
Edwin Kanters WP7, VIRT²UE
Marc van Hoof WP6, VIRT²UE
Ana Marušić WP1, VIRT²UE
Ružica Tokalić WP1, VIRT²UE
Vicko Tomić WP1, VIRT²UE
Ivan Buljan WP1, VIRT²UE
Ana Utrobičić WP1, VIRT²UE
Armin Schmolmüller WP4, VIRT²UE
Nicole Föger WP4, VIRT²UE

Description of deliverable

This deliverable, due in the sixth month of the project, provides a first version of the VIRT²UE data management plan (DMP). The Data Management Plan describes the data management life cycle for the data to be collected, processed and/or generated by the project¹. It includes the following information:

- the handling of research data during & after the end of the project
- what data will be collected, processed and/or generated
- which methodology & standards will be applied
- whether data will be shared/made open access and
- how data will be curated & preserved (including after the end of the project)

The DMP will be updated, where necessary during the VIRT²UE project. This document is based on the Template for the ERC Open Research Data Management Plan (DMP)².

¹ https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

² Based on 'Guidelines on FAIR Data Management in H2020', version 3.0. 26.07.2016

Background

VIRT²UE aims to develop a sustainable train-the-trainer learning programme enabling contextualized ethics and research integrity (ERI) teaching across Europe, focusing on understanding and upholding the principles and practices of the European Code of Conduct for Research Integrity (ECoC). The training programme will be delivered using a blended learning approach, combining online and face-to-face training. This innovative blended learning programme will provide a toolbox of educational resources based on existing ERI education resources and incorporates an eLearning course with face-to-face sessions. VIRT²UE's online component will be delivered through the EC funded platform currently being developed in the EnTIRE project (nr. 741782).

The WPs within VIRT²UE will generate and collect various data. The data include but are not limited to the mapping of scientific virtues, educational resources, training materials and cases related to ERI. During the project, the scope of the data will be decided upon by literature reviews, as well as stakeholders, and will be compiled from existing closed and open data sources. New data will also be produced, mostly in relation to the interpretation of existing data. All non-sensitive data will be made publicly available following FAIR principles.

1 Data summary

1.1. Purpose of data collection/generation

The overall aim of the data collection within the project is to provide evidence and information for the development of a sustainable train-the-trainer blended learning programme and materials for ERI teaching across Europe. For this purpose, each WP is dedicated to the following types of data collection/generation:

- WP1 will conduct a conceptual mapping amongst stakeholders to identify and rank the virtues which are essential for good scientific practice and their relation to the principles and practices of the ECoC.
- WP2 will map existing RI educational materials and training programmes amongst commercial and non-commercial stakeholders and will conduct stakeholder consultations to understand needs and lacunas of existing RI training programs.
- WP3 will evaluate the first versions of the developed face-to-face training using pilot testing.
- WP4 will produce a toolkit of online educational materials and methods to promote blended and active learning based on existing materials, which are updated and modified, as well as new materials (e.g. MOOCs, YouTube videos).
- WP5 will collect and generate data relevant to the organization and dissemination of the training programme, including the compilation of a database of ERI trainers and the development of materials for assisting trainers in implementing the training themselves
- WP6 will be responsible for making the data collected by the other WPs accessible to the public and for evaluating the data based on users' experiences.
- WP7, project coordination, will not collect any data.

1.2. Relation to the objectives of the project

In order to meet the primary objective of this project of developing trainings on ERI for trainers, it is necessary to collect evidence on virtues related to good scientific practice, as well as information about current practices. To develop the trainings, data will be gathered on existing training materials and new materials needed for the training sessions will be generated. To ensure that the quality of the training is high and to improve the trainings, it will also be necessary to collect data on user experiences.

1.3. Types and formats of data generated/collected

For each WP, a short description of the types and formats of data collected is provided below:

- WP1:
 - **Stakeholder focus groups:** Face-to-face focus groups and an online survey from two mixed focus group meetings of 12 persons each will be conducted. This will result in focus group recordings, transcripts, and basic data from a survey of focus groups participants' basic background characteristics.
 - **Literature review:** A scoping review study of scientific virtues addressed in the scientific and grey literature will be performed based on published and unpublished literature which covers an area of ethics and research integrity training. This will result in notes based on the literature review in textual format.
 - **Delphi consultation process:** The Delphi consensus process from 16 panel lists including stakeholders from different domains and different from those participating in focus groups will allow us to produce background characteristics of the participants and an analysis of the results of three rounds of consultation on virtues for the training programme.
- WP2:
 - **Literature review:** Existing scientific and grey literature on RI educational materials and training practices will be consulted. The data collected is mainly textual and is about existing ERI educational materials and training programmes.
 - **Educational materials:** Collected educational materials from different websites and from contacted people. These may be in various formats, including textual data, videos, PowerPoint slides, audio files, etc.
 - **Stakeholder focus groups:** The focus groups with commercial, non-commercial and funder organizations will result in focus group recordings, transcripts, and basic demographic data from a survey (using EU Survey tool) of focus groups participants' basic background characteristics.
- WP3:

- **Pilot testing of training:** Consortium members will take part in the pilot testing of the training and provide verbal feedback, which will be transcribed and stored as a textual document.
- WP4:
 - **Updating existing educational materials:** Textual data and PowerPoint slides on different topics (e.g. data management, publication and authorship, peer review, etc.) and case studies/moral dilemmas, card games, role plays/theater, videos/film scenes, fictional movies and art.
 - **Creation of new materials:** YouTube videos and MOOCs .
- WP5:
 - **Database of ERI trainers:** Institutions will be contacted directly in order to produce a textual file with information on ERI trainers in different institutions in EU member states.
 - **Materials for assisting trainers:** Textual data, which will be generated based on evaluations done by WP6.
- WP6:
 - **Evaluations:** Users of the online platform can provide questions and answers to issues related to the content posted. This functionality will be used to gather information on user experiences, resulting in textual data.
 - **Open source software:** This will be made available in an online repository.

1.4. Origin of the data

Most of the data will be gathered from existing sources (closed and open ones). New data will also be produced (e.g. when consulting stakeholders and creating materials for educational purposes).

A general description of the origin of the data can be found here below:

- WP1:
 - **Stakeholder focus groups:** new data
 - **Literature review:** reusing existing data of information available in the literature and generating new data (i.e. forming new insights based on the literature)
 - **Delphi consultation process:** new data
- WP2:

- **Literature review:** reusing existing data of information available in the literature and generating new data (i.e. forming new insights based on the literature)
- **Educational materials:** existing data
- **Stakeholder consultations:** new data
- WP3:
 - **Pilot testing of training:** new data
- WP4:
 - **Updating existing educational materials:** The data is taken from the following sources:
 - MEDLINE and SCOPUS databases
 - Specialized sites, like ORI (Office for Research Integrity) website (<https://ori.hhs.gov/>) and RRI Tools (<https://www.rri-tools.eu/>)
 - Websites from universities; websites of EU projects, identified in EU project website <http://cordis.europa.eu/>;
 - Data from networks of RE+RI bodies in Europe, such as ENRIO – European Network of Research Integrity Offices (<http://www.enrio.eu/>) and EURECNET – network of national Research Ethics Committees (RECs) associations (<http://www.eurecnet.org/index.html>)
 - Webpages of EU projects addressing RE+RI (such as ENERI (<http://eneri.eu/>), DEFORM (<https://www.deform-h2020.eu/>), PRINTEGER (<https://printeger.eu/>);
 - Public information on RE+RI experts from the above sources
 - **Creation of new materials:** new data
- WP5:
 - **Database of ERI trainers:** new data
 - **Materials for assisting trainers:** new data
- WP6:
 - **Evaluations:** new data

1.5. Expected size of data

At this moment of time, it is not known how much data will be collected/generated. It is expected that the focus group meetings will yield transcripts that are 30.000-40.000 characters (with spaces) per meeting. With regard to creating data, we plan to produce 10 YouTube videos and 6 courses for the MOOC.

1.6. Data utility

The collected data will be relevant for ERI trainers in search for novel and effective materials and training strategies to use in training researchers. In addition, individual researchers who want to find support for research practices will be able to access the online components of the trainings. Moreover, the research community as a whole can benefit from having open access to online teaching material from this innovative blended learning programme designed to foster scientific virtues. To ensure these benefits, VIRT²UE will follow FAIR data principles (section 2).

2 Fair data

FAIR data principles³ will be followed for all data generated. Some of these data will be made publicly available after the end of this project, as detailed in the overview below.

2.1 Making data findable

2.1.1 Discoverability of data

Data can be found in multiple ways. Common search engines can be employed to search through the data, and part of the dataset will also be made available on the platform for offline analysis and use.

2.1.2 Identifiability of data

Two main provisions will be used. The URL naming on the platform is persistent for content. Also, Digital Object Identifiers will be employed for reviewed content to make data entries persistent and easily retrievable and citable to scientists.

2.1.3 Naming conventions

The naming convention used will be based on the work form WPs 1 and 2, since these WPs are responsible for coming up with the conceptual framework for the project.

2.1.4 Approach towards search keyword

Three main approaches towards search keywords will be employed. This only applies to data to be placed on the platform.

- 1) Keywords will be included in the page of the online platform to increase searchability by common search engines. These keywords will be based on the analysis of ERI terminology by the EnTIRE project.
- 2) Users on the platform will be given the opportunity to add tags to content. This folksonomy approach is more flexible and dynamic and ensures that over the longer term, keywords match what people are looking for themselves.

³ M. D. Wilkinson et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3:160018.

- 3) The search of users on the platform will be analyzed to investigate what users are looking for and if they found it. This analysis can be used to tailor the keywords to match what users are searching.

2.1.5 Approach for clear versioning

The Mediawiki software that the online content will be put up on has a versioning system which tracks every modification made on the platform. This overview of modifications (an identification number together with a data) will be made available online. This ensures that when cited, the specific version of the document can be traced back.

2.1.6 Metadata

Since EnTIRE is currently working on developing the platform where most metadata from VIRT²UE will be stored, it is currently not clear which metadata standards will be used. In the last year of the project of VIRT²UE, when it is expected that the platform will be well developed, WP7 will make a plan for the creation and handling of metadata. The approach taken to make this plan will include an investigation of metadata standards present in the RE+RI field. This will consist of a systematic search for current vocabularies and metadata in science and RE+RI, which would be appropriate to annotate the data. Their suitability will be reviewed and these will be employed, adapted and/or extended where necessary.

2.2 Making data openly accessible

In principle all data produced in the project will be made openly available; either on the project webpage or platform that is being developed by EnTIRE, or in another database/repository. The exception is made for material which is theoretically traceable to an individual.

2.2.1 Ethical concerns related to publication

For sensitive data which will not be made publicly available, researchers can contact the relevant WP leader. Contact details and instructions will be present on the project webpage. In order to avoid the risk of participants' personal information being exposed to others, the data collected by the stakeholder consultations will not be fully published. Special measures will be adopted to ensure the protection of privacy and confidentiality:

- 1) Cases and quotes from the focus groups will be pseudonymised and published only after written consent;
- 2) Full transcripts of the focus groups interview will not be published and will only be accessible for authorized study personnel;
- 3) Audio recordings of face-to-face focus groups will be destroyed after they have been transcribed and quality checks have been conducted.

Moreover in order to protect the intellectual property of the authors, grey literature,

such as government reports, will be made publicly only after written consent has been obtained from the author.

2.2.2 Methods/software tools needed to access data

The data can be accessed using a conventional internet browser or an open source (i.e. Python⁴) or closed source software package (Matlab⁵, Mathematica⁶). For data that will be available on the ‘Embassy of Good Science’ platform, it will also be possible to export documents and data for online and offline use. This is with the exception of personal data, which will be contained to secure databases and will not be exported for offline use. All members of the community working on this platform will commit data based on the latest Creative Commons License (4.0) ensuring an open data and open access approach. This adheres to the license requirements of Wikimedia⁷. Documentation which includes the instructions to handle the data will be available on the platform, project website and other databases/repositories used.

2.3 Making data interoperable

In general, data, metadata and vocabularies will be generated (see above, section 2.1.4).

2.4 Increase data re-use (through clarifying licenses)

All the data uploaded on the project webpage, other databases and on the ‘Embassy of Good Science’ platform will be made available through the Creative Commons License structure where applicable. In cases of copyright, data will be linked, and deduced work will be made available under the Creative Commons License where possible. The data will be available from early on in the project. It is estimated that the first data to be available is at the 1 year time point (June 2019). No commercial re-use of the data will be allowed prior to written consent from the consortium lead (WP7).

2.4.1 Data quality assurance processes

An evaluation of the quality of the data uploaded on the platform is part of WP6 (Online platform and evaluation). The community is responsible for curation of data together with the consortium.

3 Allocation of resources

In this project the data is FAIR by design. No additional costs are associated with FAIR data management. As the availability of the data can be expected to be valuable to many stakeholders, a plan will be created by WP7 to ensure long term preservation and to

⁴ <https://www.python.org/>

⁵ <https://www.mathworks.com/products/matlab.html>

⁶ <https://www.wolfram.com/mathematica/>

⁷ https://wikimediafoundation.org/wiki/Resolution:Licensing_policy

distribute the long-term upkeep costs amongst the stakeholders. Data management on the project level is initially the primary responsibility of project co-ordination (WP7), but leader beneficiary of each work package will be responsible for data they have collected themselves.

4 Date security

4.1 Data recovery, security storage and transfer of sensitive data

During the project:

During the project (01-06-2018 until 10-06-2021), all data will be stored on DropBox Business as a centralized backup and on the local servers of the respective universities of the WPs. DropBox Business has been chosen because it allows for secure data storage, with good back up capacity. VIRT²UE has made an agreement with DropBox to have all data stored on EU servers only, to ensure data security. The GDPR compliance of Dropbox can be found at

https://aem.dropbox.com/cms/content/dam/dropbox/www/en-us/security/dropbox_gdpr_compliance_journey.pdf

After the project:

All data will be made openly accessible as much as possible at the end of the project. The data will be made accessible either through an open repository or through the 'Embassy of Good Science', the EnTIRE platform that is currently being developed by the EU funded project EnTIRE.

EnTIRE's ICT partner, gesinn.it (nr.2 GI) (expert for automating information and knowledge management based in Germany) will be responsible for data security on the EnTIRE platform.

Regardless of which platform the data are stored in, standard measures such as data backup and good computer security practices will be used to ensure data security. In order to guarantee data security, user authentication (SHA-512) and SSL (HTTPS) will be used. In case of a data breach, affected users of the community will be contacted, the server will be taken offline and leaks in the platform will be solidified within 48 hours. Any breach of the integrity of the IT system will be reported to the responsible national and/or international authorities, in line with the EU's General Data Protection Regulation (GDPR)⁸. VIRT²UE will adhere to all the rules laid out by the GDPR.

Sensitive Data

Transfer of sensitive data will be made by establishing a secure connection (SSL), using Dropbox Business with servers located in the EU (with two-stage authentication, SSL and AES-256 bit encryption conforming to the latest standard). Additionally, sensitive data resulting from WP2's stakeholder consultations will be stored for 10 years in the KU Leuven stored J-drive accessible only by people involved in the VIRT²UE project. Other WPs will similarly store data in their institutional servers for 10 years, ensuring that access is restricted to VIRT²UE members. Data may be used for further research

⁸ <https://eugdpr.org/>

within the scope of VIRT²UE and, when applicable, will adhere to the informed consent obtained.

5 Ethical issues

Most of the data collected within the project comes from the public domain. However, the project involves research with human participants (questionnaire, focus groups, Delphi consensus process). Participants will not be exposed to the risk of physical injury, financial, social or legal harm, and potential psychological risks will not exceed the daily life standard. Privacy and confidentiality of research participants and of the members of the community on the platform will be protected. Before publishing information on the project webpage, on other repositories or on the EnTIRE platform, confidentiality and privacy issues will always be addressed. If necessary, informed consent will be obtained. No personal data will be placed online; all online material will consist of aggregated data.

We are not aware of and do not expect any potentially critical ethical implications of the research results such as the protection of dignity, autonomy, integrity and privacy of persons, biodiversity, protection of the environment, sustainability or animal welfare. The research does NOT include research activity aimed at human cloning, intended to modify the genetic heritage of human beings, to create human embryos, or involving the use of human embryos or embryonic stem cells. This research does not include any security sensitive issues.

Ethical standards and guidelines of Horizon2020 will be rigorously applied, regardless of the country in which the research is carried out. Ethical approval from the respective country that the research is carried out in will be sought when conducting research with human participants. We have obtained ethics approval or waivers for researcher being conducted by WP1 and WP2 from Croatia and Belgium. For WP6, an addendum has been written for ethics approval obtained for EnTIRE's WP2, led by Dr. Natalie Evans.