Ref. Ares(2018)5608696 - 01/11/2018





This project has received funding from
the European Union's Horizon 2020
Research & Innovation programme
under Grant Agreement no. 770309

D 4.1. Initial Data Management Plan RETOPEA (October 2018)

Contents

1. Aim and purpose / general objectives 2
1.1. Context : RETOPEA 2
1.2. General objective regarding data management and open data 2
1.3. Support 2
1.4. The DMP as work in progress 3
1.5. European collaboration: OpenAire 3
2. Operational premises 3
2.1. Retaining and depositing materials 3
2.2. Submission agreements 4
2.3. Publications 4
2.4. Accessibility 4
3. Data Summary 4
4. Data handling and workflow plan 5
4.1. Stage 1: Creating the clipping 6
4.2. Stage 2: Testing the clipping for readability 6
4.3. Stage 3: Translating the clipping & metadata/keywords 7
4.4. Stage 4: Sending the clipping to the partners for translation review 7
4.5. Stage 5: Sending the clipping to the Research Data Manager 8
5. Costs and Allocations of Resources 8
6. Data Security 8
7. Ethical and Legal Aspects 9

1. Aim and purpose / general objectives

1.1. Context : RETOPEA

RETOPEA brings together eleven partners across Europe to reflect on religious tolerance and peace, with the aim of sharing this reflection with students (age 12-18) and policymakers. The project will create 'clippings' of historical and contemporary source material, which will be shared on an online platform. These clippings will serve to help students reflect on their own dealings with religious differences, a though process that will culminate in the creation of 'docutubes' – short videos created by the teenagers themselves.

The Data Management Plan (DMP) of RETOPEA outlines and explains how its team will handle, collect, process or generate research data during the project. It describes the data management life cycle for all relevant datasets, highlights what methodology and standards will be followed; whether and how these data will be shared and/or made open; and how they will be curated and preserved.

1.2. General objective regarding data management and open data

RETOPEA is committed to the open access of its (scientific and public-oriented) output and the research data and materials produced, collected and used during the project. The consortium conforms to this policy: regarding the digital research data generated during the project ('data'), the beneficiaries must deposit them in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate – free of charge for any user – all the research data produced during the project. These include

- The data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible.
- Other data, including associated metadata, as specified, and within the deadlines laid down in the DMP.

RETOPEA will also provide ample public information and documentation on its digital infrastructure and the tools and instruments used by its researchers, for instance specialized software or software code, algorithms, analysis protocols, etcetera. Where possible, the consortium will also publicly provide the tools and instruments themselves.

1.3. Support

The KU Leuven RDM team supports its researchers in the area of RDM (research data management) by the provision of a customized and free data management plan (DMP) (including guidance to KU Leuven legal requirements and policy guidelines) based on the templates provided by the Digital Curation Centre (UK), and a dedicated helpdesk advising on data storage, metadata and preparing data for sharing.

KU Leuven assists RETOPEA to comply with the ERC Open Research Data Management Plan.

KU Leuven's data policy demands retention of data for a period of at least 5 years after the end of a project, PhD or publication (barring 3rd party agreements). The university also continues to invest in affordable long-term storage and curation. LIAS KU Leuven and its digital assets management system Teneo are the main instruments of RETOPEA for digital preservation. Furthermore Retopea can count on the support of the curation services of LIBIS-KU Leuven.

KU Leuven experts will also assist RETOPEA in checking and complying with the regulations on intellectual property rights, research ethics, privacy (incl. the definition of personal data), and the ICT code of conduct for staff on security and safe storage of data.

1.4. The DMP as work in progress

The consortium promised to elaborate a full but initial data management plan before month 6 of the project, which, effectively, entails this document. This plan will be continuously updated during the project in accordance with the project's reflection on and experiences with data management. The project will use the KU Leuven's DMP Online tool to continuously develop and refine the project's data management plan.

This initial plan was discussed during the RETOPEA six month-meeting on 30 October 2018. A revised/updated versions will be delivered at half-way meeting in Granada, which is currently scheduled for March 2020. A final DMP will be delivered at the end of the project.

1.5. European collaboration: OpenAire

RETOPEA will join the Open Access Infrastructure for Research in Europe (OpenAIRE), the data-packages gathered in its digital repository will be made available on this platform. The data archive of RETOPEA will be made compatible with the OpenAIRE infrastructure by implementing DataCite Metadata Schema v3.1 and the OAI-PMH v2.0 protocol for harvesting dataset metadata. Standard and commonly accepted formats (PDF/A, ODF, JSON, JPEG2000) will be used to archive and share the data. The first contacts for making this possible have been established during the summer of 2018.

2. Operational premises

2.1. Retaining and depositing materials

The RETOPEA partners and their researchers will retain all the research data they have produced and used during the course of their work.

This is of course also the case for the other main deliverables of the project, for instance articles, conference reports, policy briefs, educational packages (Badged Open Courses), the 'docutubes' produced by teenagers etcetera.

The main deliverables of the project (WP 2, 3 and 5) and the research datasets and materials used to produce them, together with their (administrative, technical, structural,,…) metadata, will be centralized by the RETOPEA data manager (IEG Mainz) and durably stored and managed using the long-term digital preservation repository of KU Leuven and its digital assets management system Teneo, supported by the curation services of LIBIS-KU Leuven. This infrastructure guarantees the safe storage and preservation of data with back-ups in a redundant and geographically distributed environment following the rules of disaster recovery. Teneo archives and provides persistent access to born-digital or digitized data objects and datasets and includes metadata, usage policies, upload and access tools. Its scalable and extensible architecture makes it a sustainable and trusted preservation infrastructure.

2.2. Submission agreements

Formal submission agreements will be made for all data packages, specifying the intellectual property rights. The Steering Committee of RETOPEA will handle all ownership issues in accordance to EC guidelines and rules.

2.3. Publications

All peer reviewed scientific publications resulting from RETOPEA will be made open access i.e. free of charge online access for third users through (1) the publisher ("Gold OA"), (2) the academic institutional repositories of the different partner universities ("Green OA", e.g. LIRIAS-KU Leuven) or (3) through the RETOPEA web platform and its public access tools and instruments (OMEKA).

2.4. Accessibility

The RETOPEA website and its public access tools and instruments (OMEKA), will offer quick and easy access to the projects own trusted repository and its asset management system Teneo, respecting property rights and privacy issues, but also to existing digital resources of the partners, for instance the database ODIS, the edition of Early Modern Religious Peace Treaties and the European Peace Treaties Database).

Building and managing the digital infrastructure of RETOPEA is the main objective of WP 4. RETOPEA-partners IEG Mainz and KADOC-KU Leuven will ensure that the project website will be kept online for at least five years after the end of the project, although we can expect that technological support for its main components will proof to be durable.

Given the diversity of the source materials and data-packages used and produced in the project (see above), their accessibility must be evaluated bearing in mind privacy, ownership and copy rights or 3rd party agreements explicitly preventing us from sharing the data involved.

The RETOPEA data manager (WP 4) will coordinate and publicly document this process in close consultation with the research teams involved and the PI/general supervisor.

3. Data Summary

RETOPEA anticipates to work with the following types of data:

Origin	**types**	**Extensions**
Clippings	Text files	.docx, .pdf
	Images	.jpeg
	Videos	.avi, .wmv, .mp4, … (The expectation is that the docutubes will be delivered in different yet common video formats. These will be transformed to the format most suitable for the OMEKAplus environment. Additionaly, we are exploring the options for integrating youtube video's, which also facilitates any copyright issues)
Docutubes	Videos	.avi, .wmv, .mp4, … (The expectation is that the docutubes will be delivered in different yet common video formats. These will be transformed to the

		format most suitable for the OMEKAplus environment)
Research files	Reports	.docx, .pdf
	Deliverables	.docx, .pdf

In quantitative terms, work packages 2 and 3 estimate to produce each around 200 clippings. This number has been determined after consultation with the relevant researchers working in these work packages; the educational specialists involved in the project; and the members of work package 5. The rationale is based on the feasibility of producing the required number of clippings as well as on the amount of information that can be expected to be handled by teenagers age 12-18. A limited but highly relevant selection was preferred over a more extensive but more disparate selection. Determining on further consultations between the different work packages this number might be somewhat reduced or increased in the future, but 200 will remain the most plausible estimate. To this will be added the 25 test docutubes produced by work package 5.

The metadata accompanying these files will be provided in accordance with the Dublin Core standards, which were deemed refined and standardized enough for this type of project.

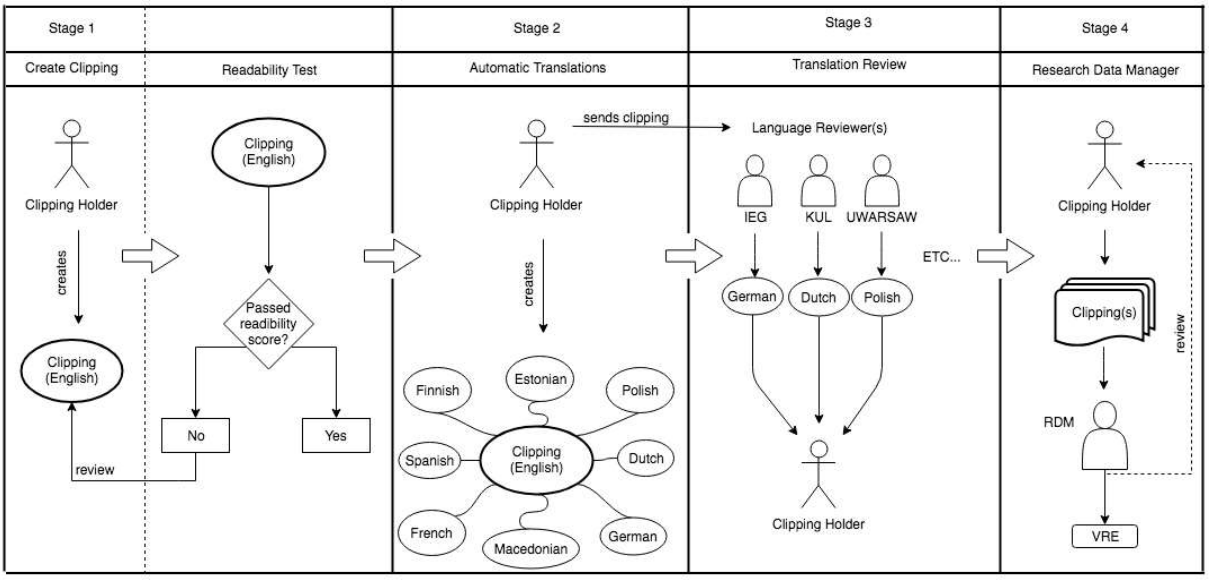
4. Data handling and workflow plan

This section represents the workflow between the work packages with particular attention to creation, translation, review and final upload of clippings. This workflow applies to all written content as well as to metadata keywords.

The structure of the workflow will be as follows:

- Creating the clipping
- Testing the clipping for readability
- Translating the clipping & metadata/keywords
- Sending the clipping to the partners for translation review
- Sending the clipping to the Research Data Manager

This structure can be visualized as follows:



Please consider the following terms when reading this document:

Clipping bundle	A bundle consists of the original clipping and the clippings translations in eight languages.
Clipping Holder	Individual in charge of and responsible for the quality and correctness of the clipping(s).
Language Reviewer	Individual nominated by the designated institution, in charge of and responsible for the quality and correctness of the clipping's translation in the appointed language.

4.1. Stage 1: Creating the clipping

Clippings may consist of different file types and appear in various formats. All the clippings will have a structured metadata vocabulary. Further information about the translation and the metadata structure will be in explained in "stage 2" and at this link respectively.

Clippings with written content are divided as follow: for longer and complex texts a readability test is necessary to match pupil's reading level; for every other written part (i.e. image captions, metadata keywords, etc…) follow the instructions in "stage 2".

Procedure for video documents will be defined later on.

4.2. Stage 2: Testing the clipping for readability

This stage is especially relevant for clippings containing mainly written text. It is recommend to analyze its reading ease and grade level by testing its readability. The readability test and the score results are based on and apply to English written texts only. This way the Clipping Holder will be able to ensure the appropriate understanding level for its target. The project's target group's age will be between 13 and 18 years, which corresponds to children attending classes between the seventh to the twelve grade. The clippings can be tested via this easy-to-use online tool.

The relevant scores for each test in prioritized order are as follows:

	Flesch-Kincaid reading ease score	ARI (automated readability index)	SMOG index	Flesch-Kincaid grade level	Gunning Fog index	Coleman-Liau index
Score	90.0 - 50.0	7 - 12	6-12	6-12	6-12	6-12

If the results of at least four of six tests are in the above range, the clipping can be considered as an understandable text for the target group. If the results extremely differ from the above scores, the Clipping Holder should consider to re-edit the clipping.

Here is a brief example on how the scores can vary by simply splitting a long sentence in two shorter ones:

> *The Peace of Augsburg was a treaty signed between Charles V, Holy Roman Emperor, and the forces of the Schmalkaldic League on September 25, 1555 at the city of Augsburg in Germany.*
>
> *The Peace of Augsburg was a treaty signed between Charles V, Holy Roman Emperor, and the forces of the Schmalkaldic League. The treaty was signed on September 25, 1555 at the city of Augsburg in Germany.*

	Flesch-Kincaid reading ease score	ARI (automated readability index)	SMOG index	Flesch-Kincaid grade level	Gunning Fog index	Coleman-Liau index
Score	90.0 - 50.0	7 - 12	6-12	6-12	6-12	6-12
Example 1	44.8	15	11.6	15	12.8	9.8
Example 2	61.7	8.1	8.3	9	7.2	9.9

4.3. Stage 3: Translating the clipping & metadata/keywords

The clippings have to be translated in eight different languages: German, French, Spanish, Polish, Dutch, Macedonian, Estonian and Finnish. To accelerate the clipping's translations from the English source to the above mentioned languages, the following tools are recommended:

For translations from English to German, French, Spanish, Polish, Dutch it is recommended to use DeepL.

For translations from English to Macedonian, Finnish and Estonian, please evaluate these tools besides Google Translate:

Macedonian	→ Tartu University Translator → Yandex
Finnish	→ Matrix
Estonian	→ Tilde

Likewise all metadata keywords are translated in the mentioned eight languages with Babelnet.

4.4. Stage 4: Sending the clipping to the partners for translation review

Subsequently the automated translations, the clippings presumably require a minor grammatical and semantic revision. Each institution nominated a Language Reviewer, who will be responsible for the quality and correctness of the translated clipping. The Clipping Holder sends the automatically translated clipping to the Language Reviewer. The Language Reviewers will afterwards send the examined and reviewed version of the clipping back to the Clipping Holder.

Language	Partner	Contact
German	IEG	juergens@ieg-mainz.de
French	Le Foyer	ann.trappers@foyer.be
Spanish	UGR	earigita@ugr.es
Polish	Uniwarsaw	michal.choptiany@al.uw.edu.pl
Dutch	KUL	bram.deridder@kuleuven.be
Macedonian	MLIMC/MCIC	georgieva@fzf.ukim.edu.mk
Estonian	Utartu	riho.altnurme@ut.ee
Finnish	HELUNI	arpad.welker@gmail.com

4.5. Stage 5: Sending the clipping to the Research Data Manager

The Clipping Holder sends the clipping bundle to the Research Data Manager (RDM - see contacts below), who will go back over and re-examine the clipping bundle, before inserting and integrating the data in the VRE. If the clipping bundle presents imprecisions, the RDM will send the clipping bundle back to the Clipping Holder for reexamination.

Laudito, Ilenia	IEG	laudito@ieg-mainz.de
Büchler, Marco	IEG	buechler@ieg-mainz.de

5. Costs and Allocations of Resources

The costs of the data management practices and the allocation of resources are covered by the Grant Agreement. No changes are to be expected here.

6. Data Security

Teneo is an e-depot with a secured storage system that centrally gathers and stores files and metadata. Teneo complies with all international standards for digital preservation (such as OAIS and Premis). From the start of the data entry, its automatic processes immediately support the validation of the content.

From there, long term preservation is guaranteed through the distinct instruments linked to Teneo. These include the following:

- Persistent identification: each object in the e-depot is provided with a unique code or Persistent Identifier (PID). These URI's are used to keep objects accessible beyond the end date of the project.
- Scalability: Teneo has a distributed architecture that allows for the preservation of collections of all sizes. For Retopea, the option has been chosen to start with a small scale system that will be expanded in function of the eventual needs.
- Open architecture: Teneo has a large set of API's that allow the integration of the collections within other systems. The built-in 'storage abstraction layer' ensures that connections with other e-storage facilities is possible.
- A secure physical environment: Libis offers different types of storage infrastructure. The files archived on Teneo are redundantly saved, on different servers in different physical locations.

7. Ethical and Legal Aspects

The ethical and legal issues are by and large covered by the grant agreement, combined with the procedures outlined in the ethical deliverables forwarded to the Commission in the last six months (D 8.1-8.11).