Ref. Ares(2019)2911846 - 30/04/2019

Horizon 2020 Grant nº 770309
Religious Toleration and Peace
*[Leuven – 29/04/2019]*

# Deliverable 4.4
# Report on the VRE

Contributor(s): Peter Heyrman, Luc Schokkaert, Bram De Ridder, Valérie Adriaens

Institution: KU Leuven, Kadoc

Contact: bram.deridder@kuleuven.be

Related project task: [4.1.]

Type of document: Report

Document reference: D4.4

Version number: 1

Status: [final]

Last revision date: 29/04/2019

Verification date & verifying subject: 30/04/2019

License: /

Other, if applicable: /

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under Grant Agreement n° 770309.

# Table of contents

1. ABSTRACT 3

2. OBJECTIVES AND RESPONSIBILITIES 3

3. ENVISAGED ARCHITECTURE (PROJECT APPLICATION DOSSIER), DEMONSTRATOR, AND FEEDBACK 4

4. REFINING THE ARCHITECTURE AND FURTHER DESIGN (M1-3) 4

5. BUILDING AND TESTING THE VRE (M4-10) 5

6. LAUNCH OF THE VRE (M11-12) 6

## 1. Abstract

One of the main objectives of Retopea WP 4 was the design and building of a Virtual Research Environment (VRE) during the first twelve months of the project by a dedicated team consisting of experts of the IEG in Mainz and KU Leuven. This online infrastructure was constructed as planned. It became fully operational in March 2019.

## 2. Objectives and responsibilities

In order to provide a stable and collaborative online workspace allowing both researchers and the general public to access the information packages and data produced by the project, Retopea created a Virtual Research Environment (VRE), a crucial part of work package 4. Digital Data Management and Curation.

Retopea strived to set up a durable and multifunctional technical environment by implementing trustworthy hard- and software solutions and ensuring the interoperability of the different VRE components.

Although the VRE aims to reach the general public, a special target audience consisted of the groups of teenage students that would create the 'docutubes'. The VRE thus needed to be both well-structured, robust, attractive and easily accessible.

The responsible partners for designing and setting up the VRE were IEG Mainz and KU Leuven (KADOC and Libis).

The main team members involved were:

- Ilenia Laudio (IEG Mainz, Retopea data manager)
- Marco Büchler (IEG Mainz, until January 2019)
- Luc Schokkaert (KADOC KU Leuven)
- Peter Heyrman (KADOC KU Leuven)
- Valérie Adriaens (Libis-KU Leuven)
- Roxanne Wyns (Libis-KU Leuven)
- Joris Lambrechts (Libis-KU Leuven)
- Bram De Ridder (Arts KU Leuven)

The conceptual work on the VRE (design, structure, functional analysis, vocabularies…) was done collectively, the hosting and technical development on the VRE was the responsibility of KU Leuven.

The team-members of IEG Mainz are responsible for the data-management, create and manage the workflow (procedures, meta-data schemes, forms, dialogue with and feedback from the researchers,…).

## 3. Envisaged architecture (project application dossier), demonstrator, and feedback

The envisaged structure/architecture of this VRE provisioned three main components:

1) a collective access database with a simple webform for RETOPEA partners to deliver material produced in WP 2 and 3; and an API to link this material to resources from other databases and publication platforms;

2) a digital repository for data storage and preservation, using Teneo.

3) the publishing tool OMEKA-S

In order to introduce the different partners of the consortium in the functionalities of the VRE and to clarify the procedures and workflow a 'demonstrator' was set up for the Kick-off meeting of the project in Mainz on 24-25/05/2018. http://heron-net.be/retopea/

This meeting offered valuable feedback on the structure of the website (which is being continuously refined), the envisaged architecture, and the functionalities of the VRE. It was thereby also stressed that the VRE is a crucial infrastructure to support the work and objectives of the different (research) teams and to publish their findings, also offering inspiration to the teenage students that would create the docu-tubes.

On the basis of the deliberations in Mainz it was decided that the team would not build a separate collective access database to store the metadata of the clippings and other collections/materials in the VRE, but that these metadata would be incorporated within OMEKA-S. The envisaged web-form to deliver materials to the central data-manager was also considered as unnecessary.

The partners also underlined that it was necessary to offer the materials and metadata in as much languages as possible. These languages and their number were not yet completely clear at that time, but since it has been decided to focus on the languages of the project partners.

Furthermore the partners agreed upon the main formats in which the materials /clippings should be delivered to the data-management officer in Mainz.

The domain name www.retopea.eu was claimed, it was agreed to move the OMEKA S website as soon as possible to this domain.

## 4. Refining the Architecture and further design (M1-3)

On 19/06/2018 the key members of the team responsible for building the VRE as part of WP 4 organised a one-day technical seminar in Leuven (KADOC).

Were present: Peter Heyrman, Luc Schokkaert, Bram De Ridder, Marco Büchler, Ilenia Laudito, Arpad Welker Roxanne Wyns, Jakub Dadlez

After evaluating the functionalities of OMEKA-S and the feedback offered in Mainz on the website/VRE, the technical team made several important decisions

- The use of the collective access database was definitively dropped. The metadata of the (limited number of) clippings would be stored and structured in Omeka S, which offers various and flexible possibilities to that end.
- The development of web-form standardizing the transfer of information packages of the research teams (and their metadata) was also judged as superfluous.
- In order to determine standards, workflow and specify the procedures, a test case and data-package was developed regarding the Westphalia treaty (1648).
- The structure of the VRE (website/Omeka S) was decided on in general terms. More specific decisions would be taken in the next months, after feedback from the different WP-leaders, in particular WP1 and WP7.
- It was decided that the webpages regarding the project (objectives, consortium, key team members,…; deliverable in M3) would be integrated in the VRE, all this to avoid confusion and to facilitate a coherent communication policy.

It was also acknowledged that this would facilitate the interconnection of the RETOPEA VRE with the OpenAIRE infrastructure.

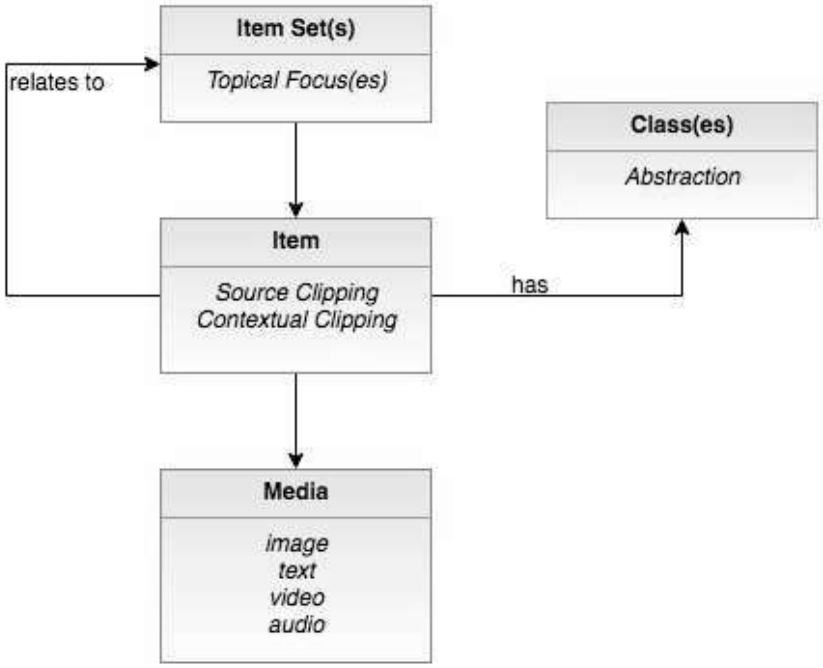
- As for the languages in which the clippings, their metadata and all other information on the website would be published, the meeting suggested that the six languages specified in the grant agreement, would certainly be implemented. Further efforts on this point needed to be decided on by the project coordinator.

## 5. Building and testing the VRE (M4-10)

Technical meetings were held on:

- 14/9/2018
- 5/10/2018
- 6/12/2018
- 13/12/2018

The structure of the clippings was designed as follows:




## 6. Launch of the VRE (M11-12)

On 4/4/2019 a final meeting was held to finalize the last decisions and work on the VRE. The focus thereby was that the RETOPEA website and its public access tools and instruments (OMEKA) 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).

The main deliverables of the project (WP 2, 3 and 4) and the research datasets and materials used to produce them, together with their (administrative, technical, structural,,…) metadata, are 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.

The repository of Teneo will enable us to preserve all the research datasets and materials that are produced during the project in perpetuity. This consists on the one hand the clippings and the docutubes that are described and uploaded into Omeka S, but on the other hand also the administrative documents that are stored on the Seafile Servers of IEG Mainz. Given the different datasets and environments, two different configurations are set out.

In preparation of these configurations, we made also two ingest agreements to secure the long-term preservation of the content centralized in Omeka S (webpages, clippings, docutubes, their descriptive and technical metadata…) in Teneo. The ingest agreements are a description of the procedure and the

generic requirements of the delivered material. This document also contains a number of recommendations regarding the material to be supplied.

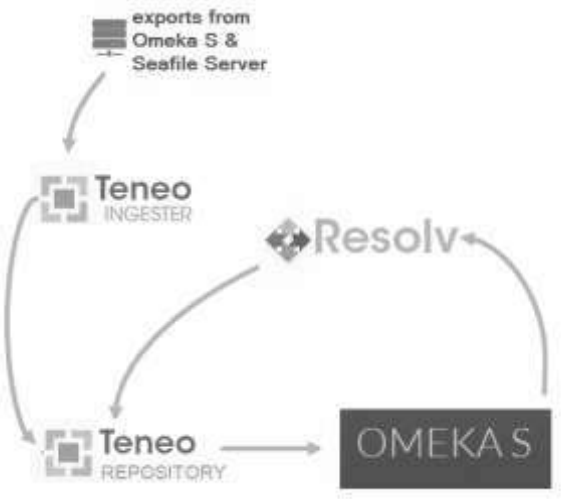



Figure 1: Teneo infrastructure

The ingest workflow starts with an export of the content, one of the Seafile Servers and one of the Omeka S. These two exports will be subsequently delivered to the Teneo Ingester. The Teneo Ingester is an integrated environment for uploading and ingesting digital files and metadata. It prepares the digital files (exports) for submission to Teneo during different phases:

1) Collection: The first task of the Ingester is to gather all the files from a directory, in this case to gather the export from a certain directory, folders and subfolders.

2) Preprocessing: During this phase different operations will be performed on individual file objects, like virus scanning, format identification, calculating and checking checksums, checking file names or file types … To prevent corrupted files from being inserted, the Ingester will check the files with programs like DROID. DROID searches the file format in the PRONOM database. This database contains a more detailed description of the file format. The integrity is controlled by checking if the file format matches the extension and the authenticity is checked by calculating an MD5 checksum.

3) Pre-ingest: When the files are controlled and validated, the ingester will combine multiple files into one object or Intellectual Entity (IE) and add metadata to the files and object. Since the metadata is stored in the Dublin Core (DC) format in Teneo, the metadata also needs to be supplied in this format.

4) Ingest: Afterwards the content is loaded into the repository of Teneo. The loaded content is packaged as a SIP (Submission Information Package), namely a package of files and associated metadata that is

suitable for being uploaded into a digital archive. The SIP is stored in a METS XML file. The METS XML contains all the information (files and metadata) regarding the Intellectual Entity and is based on the PREMIS reference model.

The repository of Teneo is based on the ISO-recognized Open Archival Information System (OAIS). The OAIS safeguards the authenticity, unicity and integrity of data and fulfils all requirements that must be present in a digital preservation system, being: (1) archival storage, (2) data management, (3) administration, (4) preservation planning and (5) access.

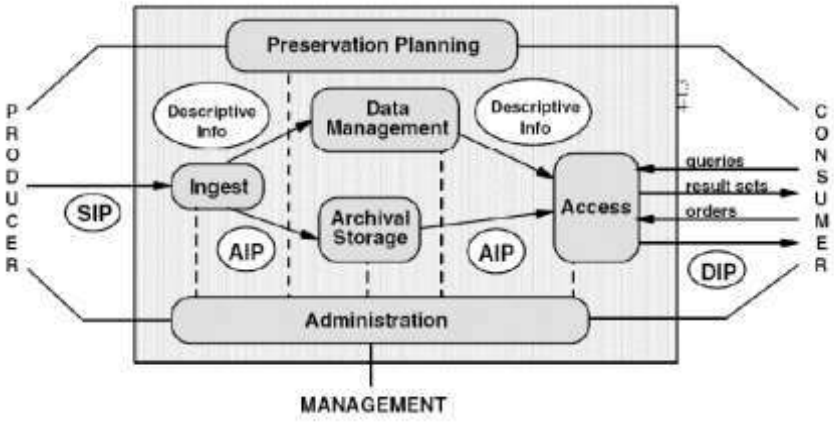



Figure 2: OAIS standard

OAIS defines three types of information packages that needs to be managed by a preservation system:

- SIP – Submission Information Package;
- AIP – Archival Information Package;
- DIP – Dissemination Information Package.

When the Ingester has loaded the content into Teneo, the SIP is sent to the Working Area of the repository at which point the SIP becomes an archival information package (AIP). The working area supports different processes:

- virus checks, format identification, technical metadata extraction and validation to ensure the integrity of the files and metadata;
- curatorial, accession processes that monitor the quality of the material;
- enrichment processes that convert information into a stable format.

Approved content is moved to the permanent repository, where the information is stored in perpetuity. Besides the permanent repository, Teneo provides an operational repository that enables searching, indexing, and quick access. The operational repository is where preservation actions take place. Dissemination information packages (DIP) are created in the access module. These packages are used for the dissemination of information stored in the permanent repository.

When the files and metadata are stored in the permanent repository, the access rights for the ingested files are set in Teneo, this way files can be protected and only accessible with a password.

Every object or dataset receives a unique identifier in Teneo, which ensures the provenance of the digital resource. The service of Resolv (API) uses these identifiers to create a persistent URI to ensure the long-term access. This link will be uploaded in Omeka S.

Teneo archives and provides persistent access to the datasets and includes metadata, usage policies, upload and access tools. With this system, large amounts of digital data - including video, audio and text content - can be stored and managed. Its scalable and extensible architecture makes it a sustainable and trusted preservation infrastructure.

When the content in the Omeka S environment is representative, we will plan a test ingest to be sure that the objects are properly ingested into the repository. When the project is finished, all the objects (files and metadata) will be exported and uploaded at once in Teneo.