D.3.4 – Final Report
WP3 – Repository Infrastructure

Version 1.9.1 - FINAL

29 October 2013

Grant Agreement number:	231809
Project acronym:	3D-COFORM
Project title:	Tools and Expertise for 3D Collection Formation
Funding Scheme:	FP7
Project coordinator name, Title and Organisation:	Prof David Arnold, University of Brighton
Tel:	+44 1273 642400
Fax:	+44 1273 642160
E-mail:	D.Arnold@brighton.ac.uk
Project website address:	www.3dcoform.eu

Authors: Martin Doerr, Maria Theodoridou, Christos Georgis, Foundation for Research and Technology, Hellas (FORTH-ICS)

Contributing partner organizations: University of Brighton (UoB)

Fraunhofer Gesellschaft Zur Forderung Der Angewandten Forschung E.V. (FhG-IGD)

Katholieke Universiteit Leuven (KUL)

Eidgenossiche Technische Hochschule Zurich (ETHZ)

Consiglio Nazionale Delle Ricerche – ISTI (CNR-ISTI)

Technische Universitaet Graz (TU Graz)

University of Glasgow (GU-HATII)

University of Bonn (UBonn)

Media Integration and Communication Centre, University of Florence (MICC)

Victoria and Albert Museum (VAM)

Breuckmann

Spheron

Table of Contents

1 Executive Summary

This document presents the work that has been done in the context of WP3 – Repository Infrastructure during the fourth year of activity of the 3D-COFORM project.

WP3 played a central role during the whole duration of the project since it was responsible for the design and the implementation of the 3D-COFORM integrated repository (RI).

Major activities performed and results obtained during the fourth year are:
- ‹ **Maintenance** and **debugging** of the RI server(s) and APIs
- ‹ **Ingestion Tool** v1.4.1
- ‹ **Europeana Extractor**
- ‹ **Long term preservation manager** tool integrated with RIv3
- ‹ **Testing**

All planned partners are contributing to WP3 while some partners exceeded their planned contributions.

The overall organization of the document is as follows. Section 2 gives a brief presentation of the objectives of WP3 for the reporting period. Sections 3-11 present in detail the work done in each of the nine Tasks of WP3 during Year 4 and the results obtained. In particular, in Section 5, as part of the *T3.3 Design and implementation of a metadata repository,* we include the work on the *Ingestion Tool* and the *Europeana Extractor* which were not part of the initial DoW. Section 12 reports on the publications produced so far. Finally, the "*3D-COFORM Query Manager and Metadata Repository Manual*" is presented in detail in Appendix A, and the "*Mapping the 3D-COFORM metadata to the Europeana*" in Appendix B.

2 Introduction and Objectives

In this report, we present the achievements made in WP3 – Repository Infrastructure - during the fourth year of 3D-COFORM. These mainly consist of maintenance and debugging of the RI server(s) and APIs, improved versions of the Ingestion Tool, the design and the implementation of the Europeana Extractor Tool, the integration of the Long term preservation manager with RIv3 and testing.

We present the work that was planned for the fourth year of the 3D-COFORM project and the work that was performed during this fourth year. The presentation is done per task, following the structure of the original Description of Work (DoW). In the task *T3.3 Design and implementation of a metadata repository* we include the work on the *Ingestion Tool* and the *Europeana Extractor* which were not part of the initial DoW.

It must be emphasized that the work in WP3 is a collaborative effort of all the involved partners and it is completed according to the DoW without significant deviations.

3 T 3.1 Scientific and Technical Coordination

The objective of this task is to coordinate and monitor the progress in the development of tools, including identification of deviations and amendments to the work plan. It also ensures the interrelation between work packages.

3.1 Work planned

The technical development was planned to last until Month 42. Afterwards the remaining technical resources would exclusively be dedicated to testing, training, deployment experiments and the final demonstration. The technical effort for the fourth and final year of the project would be focused on serving the practitioners in their real environment and within the established workflows. Thus the development would address these requirements rather than new risky and unstable features. Hence, an analysis for the consolidation of the tools would be performed, in order to identify the more matured and promising features, which would be brought forward until the end of the project. This would be realized by means of re-directing some of the technical resources. Consolidation is the main objective for the final period and this would be strengthened by directly collaborating with practitioners and by proactively supporting the enterprise activities.

3.2 Work performed

The technical development has consistently evolved throughout the whole funding period of the project. The first year was characterized by the achievement of a mutual understanding within the partners and by using existing technologies within an interoperable exchange of information. The second year produced the first versions of the new tools of the project (e.g. RI, IVB, VSL, Sketching Tool, PhotoCloud, etc.), achieving the first system integration of components and the first integration of workflows. Maturing, usability, robustness, and scalability were the main objectives of the third year, for which the practitioners were intensively involved in different activities, such as testing, training, and deployment experiments.

The motivation of the final period was the consolidation of the technology for the 3D documentation of tangible heritage, by means of employing the technical development in professional environments and for professional activities (investigation, reconstruction, learning, replication, interpretation, and sharing). In order to carry out the consolidation of the technical development, the technical partners have conducted an internal evaluation of the tools, prioritizing the most promising features against prototypical or unstable ones. This process was supported by allowing the internal reallocation of resources within the tasks of individual partners (e.g. FORTH for the RI, TU Graz for the RI, or Fraunhofer for the IVB), strengthening the development and creating a high degree of usability and stability.

The results of this process were clearly illustrated in the production of high quality content for the different travelling exhibitions and especially for the 3D-COFORM Reshaping History Exhibition during the summer in Brighton. Additionally, the maturity and consolidation of the tools enable the export of 3D content to Europeana, establishing an automatic workflow, which involves the tools of the project in the production of Europeana compatible metadata and 3D artefacts (first phase export with 11 3D artefacts and second phase export with 37 3D artefacts, and more 3D artefacts in the pipeline).

Different activities were conducted, in order to synchronize, monitor, and control the scientific and technical coordination of the project, as well as to promote the consolidation of the tools and the creation of 3D content:

‹ VSL / IVB Meeting in Graz, January 3 and 4, 2012

‹ Pre-Review Meeting in London, January 18 and 19, 2012

‹ Rehearsal Meeting in Graz, February 21 and 22, 2012

‹ Meeting on 3D Content for the 3D-COFORM Reshaping History Exhibition in Pisa, March 12 and 13, 2012

‹ Meeting on Tools for the 3D-COFORM Reshaping History Exhibition in Brighton, June 27, 2012

‹ Pre-Review Meeting in Darmstadt, October 25 and 26, 2012

Moreover, the daily communication was maintained over email and telephone conferences, thus achieving promptly reactions and avoiding the need for more physical meetings and maximizing the effective development time.

3.3 Deviation from work plan

No deviation from the work plan. The fourth year work plan has been fulfilled according to the DoW.

4 T3.2 Design architecture

4.1 Work planned

During the fourth year the RI was in maintenance and optimization mode.

4.2 Work performed

Since Month 39 (Feb 2012), the RI has been in maintenance and optimization mode. The architecture has proven to be stable and no further changes were necessary, either from requests by partners or arising through development work. Thus, the re-implementation effort in Year 3 has been of benefit. In Month 47 (Oct 2012) version 04 of the RI-API was released (specification, C++ and Java code), which contained only minor enhancements, and no changes to existing functions.

However, two developments did have impact on the internal organization of the RI:

‹ Making the RI stand-alone deployable (see T3.4) for testing and training (see T3.9)
‹ Protocol switch from HTTP to FTP (see T3.4) for file transfer client-location and location-location

These changes are transparent to RI users and RI applications (which only need to be recompiled).

4.2.1 The Organization of the R I

The RI entered into the production phase in Month 38 (Jan 2012) with two instances of the RI, **RICentral** for clean data and **RIDebug** for debugging and software testing purposes. The core software components of each RI instance (see RI architecture diagram in deliverable D3.3) are grouped technically in three parts:

‹ **RI + OR:** The RI webservice offers many functions to RI enabled software on client computers. Client software uses the RI API (C++ or Java version) to connect to this RI webservice. The Object Repository (OR, see T3.4) is the relational database storing the information about Datasets, Files, Groups, Usergroups, Locations etc, and also the uploaded metadata files. The OR is accessed exclusively over the RI webservice. OR+RI are developed and maintained by TU Graz.

‹ **QM + MR:** The Metadata Repository (MR, see T3.3) stores the semantic network in an RDF triple store. It is created from the CIDOC CRM and CRMdig information contained in the uploaded metadata files. The Query Manager webservice (QM, see T3.5) receives SPARQL queries from clients (via the RI API) and dispatches them to the MR, the OR, or the CRI, and returns the query results to the client. QM+MR are developed and maintained by FORTH-ICS.

‹ **Locations:** The binary datasets are stored at *RI Locatons*. Locations are computers running a Location webservice. Physically, they reside at the participating institutions, e.g. partners in the

consortium. There can be any number of Locations. At login time, a client computer is associated with a location. When uploading datasets, the file information and metadata are stored in the OR and MR, respectively; but the binary data are directly transferred to the Location (and thus, possibly do not even leave the institution).

‹ **CRI:** Content Retrieval Indices are specialized web services for shape or material search. The CRI interfaces only to the QM. The RI can be deployed without the CRI services as long as the client applications do not wish to execute such services.

Since Month 38 partners have been continuously using RICentral (with two Locations) and RIDebug (with two Locations). A single central point of connection has clear advantages; however, it also became clear that hosting the RI permanently at TU Graz is not a realistic long-term perspective. It was therefore decided that the RI should be made *deployable* so that new RI instances (connecting to their own set of Locations) can be set up anywhere where a suitable server infrastructure is provided. This was accomplished in Month 46 (Sep 2012), see T3.4.

4.3 Deviation from work plan

None.

4.4 Exploitation plans

Ongoing negotiations with other projects for the use of the RI.

5 T 3.3 Design and implemen tation of a metadata repository

5.1 Work planned

Testing and improvement was planned for both the MR functionality and performance, making more efficient use of the advanced features of the BigOWLIM reasoner such as the full-text searching capabilities and query results scoring.

Additionally, a consistency check of the semantic data was planned.

Regarding the Ingestion Tool, we had planned to implement a C++ wrapper (to the Ingestion tool Java-library) and an update mechanism.

By the end of Year 3, the significance of linking 3D-COFORM with the Europeana project triggered the need for a tool that would assist to export the 3D-COFORM models to Europeana. So, although not planned in the DoW, thisimplementation became an objective for Year 4.

5.2 Work performed

5.2.1 BigOWLIM reasoner

The MR functionality implementation is based on the 'Aduna openrdf-sesame' web triple-store enhanced with 'Ontotext BigOWLIM' reasoner features. It handles mainly the ingesting, updating and querying functionality of the semantic network.

During the fourth year the focus was on performance issues for the MR and work on implementing customized rules that are used by BigOWLIM.

In particular, the following tasks were performed:
- ‹ Creation of new reasoning rules applied to MR for query optimization
- ‹ Updating the applied ontologies and rdf schemas for more efficient semantic inferencing
- ‹ Testing the semantic reliability regarding the inference results of our conceptual modeling

5.2.2 QMMR

QMMR WS is a core component of the 3D-COFORM project Repository Infrastructure (RI). It incorporates two of the four basic RI modules: the Metadata Repository (MR) that materializes the graph of the semantic network and the Query Manager (QM) which is responsible for preparing queries and handles the query results. Both are provided in a Web Service form accessible via the RI API. The

other two RI modules are the Object Repository (OR) which contains all the 3D-COFORM datasets and the Content Retrieval Indices (CRI) that contain indexed information about 3D object characteristics.

All components needed for QMMR to run are packaged and delivered together as a Web Service. The details for the administration and usage of the QMMR-WS are presented in Appendix A (3D-COFORM Query Manager and Metadata Repository Manual) and were delivered to the partners on Month 43. The QMMR service is provided via the central FusionForge development repository of 3D-COFORM. In order to support the WP12 –Training (see respective deliverable) the setup procedure of QMMR WS and its respective modules was standardised, including the Metadata-Repository rebuild and Metadata-Repository initialization of types procedures.

In particular, the following tasks were performed:
‹ Creation of a utility (INIT_MR) to initialize the metadata in a new repository. This application loads all the necessary SKOS compatible rdf data containing the AAT Thesaurus and 3D-COFORM defined types
‹ Creation of a utility (MR-rebuild) for rebuilding MR with the rdf data taken from OR, in case of a repository corruption. This functionality guarantees the consistency and completeness of the semantic network
‹ Creation of one central configuration for QMMR, INIT_MR and MR-rebuild applications and D2RServer startup and stop procedures
‹ Update of the coform.properties file with the appropriate parameters used by the QMMR applications
‹ Automatic restart of D2R Server in case of a failure.
‹ Testing and debugging of the software functionality of QMMR
‹ Writing the 3D-COFORM Query Manager and Metadata Repository Manual (Appendix A)

5.2.3 Ingestion Tool

The final implementation included porting to RIv3 and was delivered in November 2011, Month 36. Since then the Ingestion Tool has been in maintenance mode, with several new releases (from original version 1.0, to the current version 1.4.1, with up to 11 minor or major releases) that introduced new functionality such as the update mechanism for events (project, acquisition and process) and entities (persons, legal bodies, devices, places etc), optimization on performance and bug-fixes. Some features that were added to the tool were proposed by the partners that used the tool after its release. These include: a) the infrastructure implementation to enable export of digital objects to Europeana (the user can declare the digital objects that can be exported to Europeana, assigning also the license under which each object will be exported) and b) the support of multipart digital objects (event output data files may be defined as parts of a "startup" data file). The IngestionTool is provided as source code via the central FusionForge development repository of 3D-COFORM.

The ingestion tool is also available in the form of Java library: as a good-practice library of software components for metadata creation that 3D-COFORM tool providers can employ to integrate their tools

more consistently and efficiently. In the first days of Month 38 the C++ wrapper (to the Java-library) was introduced, for users that are using C++ development platforms.

5.2.4 Europeana Extractor

The Europeana Semantic Elements (ESE) consist of the 15 original Dublin Core (DC) metadata elements, a subset of the DC terms and a set of thirteen elements which were created to meet the needs of Europeana.

A mapping was created for ESE v3.4.1 to appropriate metadata of the 3D-COFORM Metadata repository. The detailed mapping is presented in Appendix B (*Mapping the 3D-COFORM metadata to the Europeana ESE*). Additionally, we extended the 3D model metadata with two elements to meet the needs of Europeana not foreseen in the MR (permission to export, rights) and these changes were incorporated into the Ingestion Tool.

In order to export the mapped metadata to Europeana, a java command-line application was implemented, which extracts from the RI 3D Models marked as "exportable to Europeana" and creates according to the mapping, ESE conformant xml files with the appropriate information.

The xml file has the form:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<metadata xmlns="http://www.europeana.eu/schemas/ese/"
xmlns:europeana="http://www.europeana.eu/schemas/ese/"
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <record>
        <dc:identifier>uuid:49175a55-e736-4d9d-8874-cfcc22b66aec</dc:identifier>
        <dc:title>3D model of Loaves and Fishes</dc:title>
        <dc:creator>Cultural Informatics Research Group, University of Brighton</dc:creator>
        <dc:contributor>Cultural Informatics Research Group, University of
                Brighton</dc:contributor>
        <dc:contributor>Karina Rodriguez Echavarria</dc:contributor>
        <dc:language>en</dc:language>
        <dc:format>model/wrl</dc:format>
        <dc:source>http://www.publicsculpturesofsussex.co.uk/object?id=182</dc:source>
        <dc:rights>Public Monuments and Sculpture Association</dc:rights>
        <dcterms:created>2012-02-14</dcterms:created>
        <dc:subject>Loaves and Fishes</dc:subject>
        <dc:subject>Refers to Christ feeding the 5000 with but five loaves and two fish.
                (Matthew 14:13-21, Mark 6:31-44, Luke 9:10-17 and John 6:5-15)</dc:subject>
        <dcterms:spatial>long -0.141888856887817 lat 50.826027910088506</dcterms:spatial>
        <dc:type>3D</dc:type>
        <dc:relation>http://www.publicsculpturesofsussex.co.uk/object?id=182</dc:relation>
        <dcterms:provenance>Public Monuments and Sculpture Association</dcterms:provenance>
        <europeana:object>http://ORCentral.cgv.tugraz.at/COFORM/ObjectRepositoryData.svc
                /203e7f03-c148-42b5-96ad-d5f1680d40a8/Thumbnail/Original</europeana:object>
        <europeana:provider>3D-COFORM consortium</europeana:provider>
        <europeana:dataProvider>Public Monuments and Sculpture
                Association</europeana:dataProvider>
        <europeana:rights>http://creativecommons.org/licenses/by-sa/3.0/</europeana:rights>
        <europeana:type>3D</europeana:type>
        <europeana:UGC>true</europeana:UGC>
        <europeana:isShownAt>http://www.3d-coform.eu/3Dmodels/uuid:49175a55-e736-4d9d-8874-
                cfcc22b66aec</europeana:isShownAt>
    </record>
</metadata>
```

Initially a set of 11 models from the Share Your Heritage project of the University of Brighton were exported to Europeana. Figure 1 displays the example of the Loaves and Fishes 3D Model as it is presented by the Europeana site.

Figure 1: Loaves and Fishes 3D Model as presented by the Europeana site

The second set of exported data included:
- ˝ ISTI-CNR: 9 models
    - . Temple of Luni - composed of 3 meshes
    - . Male stele statue - 1 mesh
    - . Pisa Gryphon - 4 meshes
    - . Bargello - 1 mesh
- ˝ David Arnold's Personal Collection: 8 models
    - . 4 Egyptian ushabti
    - . Ivory wallet
    - . Ganesh statuette
    - . Egyptian Alabaster Cup
    - . Shao Lao ivory figurine
- ˝ V & A collection: 22 models
    - . Mr Punch, Wooden mask, Sword, Gypsum bust, Morion, Bronze vessel, Areca leaf hat, Betel nut crusher, Indian stone carving, Indian religious sculpture, Head of Buddha, Ivory oxen, Walnut monkey, Madonna and child (2), Alhambra panel, Stone corbel, Wooden figure with two children, Porcelain fountain (4 pieces)

5.3 Deviation from work plan

There was no deviation from the work plan. The fourth year work plan has been fulfilled according to the DoW.

5.4 Future Plans

Regarding the linking with Europeana, the export and submission tools will be maintained and any content provider will be assisted.

6 T 3.4 Design and implementation of an object repository

6.1 Work planned

The OR was completed by the end of Year 3 and there were no plans for changes. The only larger addition that was necessary was the integration of content retrieval indices (**CRI**).

Concerning the RI API Spec, only minor adjustments were envisaged. Minor API function additions were planned for convenience, but the main focus was on optimizations and bug hunting. A "bugfix becomes unit test" approach was adopted where any reported bug was fixed as soon as possible, and a respective unit test was written that tests exactly this case.

6.2 Work performed

The maintenance work in Year 4 on the RI at TU Graz focused on three tasks described in the following:

‹ OR optimization, bug fixing of server code, client API (C++ and Java code), file transfer
‹ Making the RI deployable
‹ Improvements of the collection management software, CorExplorer and MetadataGenerator

6.2.1 OR Optimization, bug fixing, file transfer

The work was comprised of the following tasks:

‹ Adjust the Microsoft IIS 7.5 Webserver to obtain better performance (buffer sizes, timeouts, …)
‹ Find and fix performance bottlenecks in the RI server, in the RI Client API, and communication overheads, e.g. by query batching: instead of asking for the data of each file in a group individually, all file data of a group can be retrieved now using a single function call.
‹ Configure timeouts to allow for the upload of larger files
‹ Introduction of *upload tickets* to fix the problem of sessions expiring during the upload of very large files. Uploads require no login anymore; instead, a valid upload ticket is sufficient.
‹ Adding support for the use of FTP (File Transfer Protocol) for the ingestion of very large files (>2 GB). So far, the file upload using the HTTP protocol worked quite well for files up to 500 MB. However, since files are assembled in the main memory of the receiving machine, the process limit of 1 GB becomes an issue. The FTP protocol also allows resuming file transfer after interruption.
  - The FTP protocol is supported in addition to the HTTP protocol (to remain compatible)
  - FTP is used both for the communication client , Location and Location , Location (for so-called *replica creation*)

16

- No client application code is affected by the protocol change; it is internal to the RI server/RI client API. Only a recompile is needed, linking against the updated libraries
‹ Support for the setup of new Locations "Heraklion" and "HeraklionDebug" at FORTH, and for new Locations at Brighton and Bonn
‹ Support for adding new Mimetypes
‹ Minor Bug fixes at OR-Server, OR-Location and Java API

6.2.2 Making the RI deployable

As explained above in T3.2, the requirement arose for the possibility to let a professional provider host the RI. Furthermore, for a training course at PIN (see T3.9), all RI server components, as well as a Location, had to run on a single computer (in fact a strong laptop).

Making the RI deployable implied revising the code base of RI webservice, Location webservice, and client API (C++ and Java) to remove all hard-coded URLs or references to specific computers, and making all aspects of the RI configurable via config files. However, the RI is server software, which is genuinely different from usual desktop software; the installation is not done simply by running a setup program. Instead, the RI requires a server infrastructure with correctly configured components. As explained, the deployable RI is composed of two parts (OR+RI and QM+MR) running best on two computers, or at least on two Virtual Machines (VMs). The requirements are as follows:

‹ **Computer**: Intel Quad-Core or higher, 8 GB RAM, fast HD >100 GB, VMware player for Windows
‹ **OR+RI**: (VM with) Microsoft Windows Server 2008 R2 with ASP.NET 4.0, web server IIS 7.5 and database Microsoft SQL Server 2008 R2
‹ **QM+MR**: (VM with) Java 6 or higher, Apache Tomcat, D2R Server

There are two possible setups, a training setup and a recommended setup:

‹ **RI training version**: One computer running the QMMR directly, and the OR+RI in one VM, and the Location webservice in another
‹ **RI complete version**: One computer for the QMMR, one for the OR+RI, and one further computer (or VM) for every location that is to be connected.

In particular, the following tasks were performed:
‹ Create the init scripts for the OR (Microsoft SQL Server):
- Creating the (internally used) MS SQL user, setting permissions for database access
- Generating all OR tables and initializing them with basic data
  Š Mimetypes, root user, root group and root usergroup
‹ Clone and clean the OR+RI VM: clear ingested data, uninstall all unimportant software
‹ Configure the IIS 7.5 Webserver of the deployable RI
‹ Configure the VM and the host system (IP addresses, Ports, Firewall, ...)
‹ Test the functionality and the usability of the APIs (C++ and Java) e.g. login, ingest, upload, …

‹ Test the solution to work with CORExplorer, MetadataGenerator, Ingestion Tool, Coyote, PowerShell

‹ Write an installation manual

6.2.3 Collection management using CorExplorer and MetadataGenerator

A modern PC can easily store hundreds of thousands of files. Users can still manage this quantity using the file system hierarchy. This proves that hierarchical grouping is an efficient way to ensure scalability.

The RI is for managing a very large number of digital artefacts. Therefore it also structures datasets in a group hierarchy, which is comparable to a directory hierarchy with hard links: the grouping is poly-hierarchical, so the same dataset (or group) can appear in any number of groups. This allows re-ordering and re-grouping without the overhead of actual file transfer. The RI group hierarchy can be built up automatically at ingest time from the directory hierarchy of the file system by using the MetadataGenerator. It also makes use of this hierarchy to speed up the generation of metadata, mapping the directory hierarchy to a hierarchy of a sequence of events.
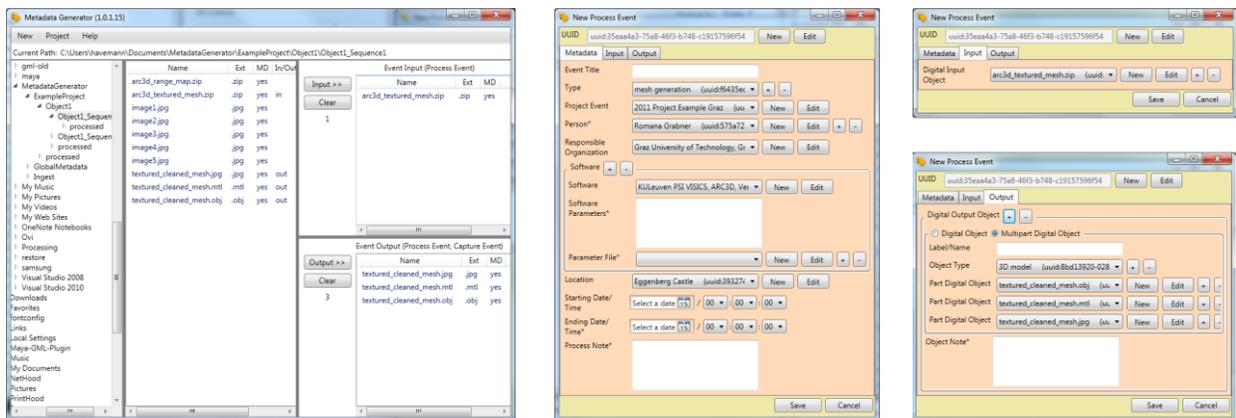
Figure 2 Metadata Generator generating metadata for a process event with a multipart digital object a output. All parts of a multipart digital object are downloaded together when viewing it in CorExplorer.

One problem we had to solve is that a 3D dataset can consist of multiple files: A Collada .dae file can load one or more .obj files, each of which can load one or more .mtl material libraries, each of which can load one or more .jpg textures. The same .obj can be used in many .dae files, etc. After discussions in the whole RI team, this was solved using a new type of annotation in the metadata scheme, introducing *multipart digital objects* into CRM-digital. This can now be generated by MetadataGenerator (see Figure 2) and CorExplorer (see Figure 3)

Figure 3: CorExplorer is the tool for hierarchical grouping and viewing of 3D assets in the RI. The context menu offers loading a 3D object into the second tab, which is the viewer tab using VSL (from T9.1)

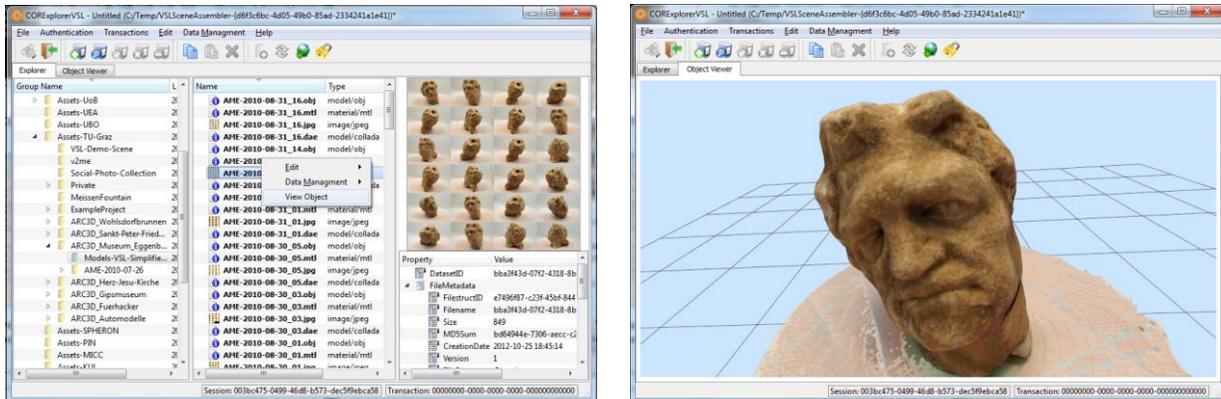CorExplorer is the tool for exploring and editing the group hierarchy, which is indispensable for collection management. In Year 4, it has been connected with the VSL in various ways (see WP9 deliverable D9.4). In particular, it is possible to directly load a 3D dataset into a viewer on the other tab in the same application (Figure 3). Both tools underwent various changes in Year 4 to increase usability and speed. They have become something like "work horses" for all kinds of RI-related tasks in the context of collection management and previewing; this is also because the CorExplorer provides a good basis for extensions, which only have to be added as new tabs in the tab row. One example is the object viewer in Figure 3, and other examples are presented in the WP9 report D9.4.

6.3 Deviation from work plan

A robust and simple solution has been realized: The indexer uses the RI API to download datasets to the indexing server (e.g. located in Leuven for shape indexing, located in Bonn for material indexing) in order to create the CRI database. The CRI database is then connected to the CRI webservice, which is registered with the QM of an RI instance. The query mechanism works as planned:

> Shape query software (client) ，RI API (client) ，QM webservice (Heraklion) ， CRI webservice (Leuven) ， CRI index query (Leuven), and then the result is sent back to the client

A perceived disadvantage of this solution is that binary datasets need to travel over the network for indexing, but the system as implemented demonstrates the functionality sought. The alternative would have been to seek an ambitious goal of running the CRI indexers directly on the Location computers. This setting would have the following advantages: The Location computers are physically located within the participating institutions, and this is where the valuable binary datasets reside, e.g. a 3D scan of Michelangelo's David statue. The shape index of the David statue is a feature vector (a sequence of numbers) describing the overall shape of the object. This shape index is then sent to a central database (e.g. the shape search CRI database at Leuven) so that it can be found by shape queries issued by clients.

Thus, the advantage is that (a) valuable binary datasets do not have to leave the institution for indexing; only the derived shape indices are sent to the CRI server; and (b) the indexing (which may be a sophisticated algorithm) works in a massively parallel fashion.

Two obstacles prevented the realization of this ambitious plan:

‹ First, in its current form, shape or material indexing does not work in an isolated fashion. Instead of deriving a feature vector from a given shape, the indexer actually compares a new dataset with the datasets that are already in the database. This form of indexing requires a bi-directional communication between indexer and database, not just unidirectional sending of data indexer , database.

‹ Second, the indexer requires a software infrastructure that is not available on the Location computers (Linux with a particular set of libraries and services, including a valid Matlab license). It was impractical to assume the availability of specific configurations of computing resources in Cultural Heritage institutions.

Therefore, the decision was taken to adopt the strategy above despite the Year 3 report having planned to adopt a strategy that avoids communicating the shape description to the indexer. However, this is considered to be a minor deviation only.

7 T3.5 Design and implementation of a query manager

7.1 Work planned

Apart from integration, testing, debugging and optimizations in the QMMR service it was planned to complete the communication with the CRI modules (see also the deliverable D7.3 where the CRI modules are described in detail). Thus, a new web service stub would be developed in order to perform the queries on the CRI service.

7.2 Work performed

The 3D-COFORM RI supports querying in parallel of all the available heterogeneous databases and returns an intersection of the separately returned results. The databases that can be queried are: OR (via D2R Server), MR, Shape Index and Material Index.

The QM manages a uniform way via SPARQL for querying heterogeneous data sources such as MR and OR, either individually or in conjunction with CRI and then handles the query results. Query preparation involves the appropriate parsing and transformation of SPARQL queries in order to redirect their parts to the proper RI modules. In particular, to query the OR module with SPARQL there is an intermediate module, called D2R server which stands as an rdf wrapper on the remote relational OR database.

The CRI comprises two separate services one for indexing shape features and the other for material features of 3D models. Indexing is an independent process performed on remote OR locations (for details, please refer to the D7.4 WP7 Final report deliverable). The result of the indexing can be queried by QM calls via REST and finally QM merges the CRI results with the rest of the results from other data sources. The parameters used for querying the CRI services are: the uuid of a specific object used as source of similarity on shape or material, as well as filtering parameters on similarity levels such as max-similarity, min-similarity, the top-n similar or dissimilar objects. When CRI is called in queries QM adds the shapeSimilarity and materialSimilarity attributes in the attribute set of the query result.

Apart from completing the QM and CRI integration, much work was done on querying performance. Semantic fundamental relations based on CIDOC-CRM used for providing accurate and complete knowledge about the resources of the semantic network had to be optimized for both higher querying speed and a more efficient way of combining simpler relations into composite ones. Optimization was achieved by defining new semantic relations on CIDOC-CRM used as shortcuts of longer and complicated graphs. These new relations were mirrored on new rules for the BigOWLIM reasoner engine of MR. The new rules were formulated and written according to the language specification of Ontotext's BigOWLIM. Every rule consists of one or more premises and one or more corollaries. Both premises and corollaries are a set of basic rdf triples pattern with a meaningful combination between them. Every time a rule premise pattern is found in the semantic data the triples of the appropriate corollaries are produced by

the reasoner. Thus SPARQL queries in many cases can now use the new shortcut relations instead of long graph chains. That makes it easier to manipulate the query statement and offers a significantly better performance if compared with the former situation.

7.3 Deviation from work plan

There was no deviation from the work plan. The fourth year work plan has been fulfilled according to the DoW.

8 T3.6 Design and implementation of an annotation and co-reference manager

8.1 Work planned

The ACoRM was planned to follow the integration, testing, debugging and optimization of the QMMR webservice.

8.2 Work performed

Annotation and Co-reference practices were designed and modelled with CIDOC-CRM. For Annotation and Co-reference it was decided to implement the source rdf files based on TriG format taking advantage of the notion of Named-Graph.

Queries for retrieving annotations, areas and their relative source files were finalized.

Constraints about area manipulation such as creation, update and deletion regarding propagation on multiple 3d models were defined in the QMMR manual and applied by the related tools. It was decided to include specific information about areas in the METS area definition files.

Co-reference was designed to satisfy two main procedures:

a. Merge of URIs into one URI, which solves the issue of referencing the same resource with different names. The merge solution foresees the global replace of selected resource URIs from all rdf source files with one called as primary and the creation of history files describing this conversion. This result of merging URIs leads to more efficient querying, but the change to the metadata repository cannot be undone.

b. Same-as correlation of resources, satisfies the need to annotate different resources as the same, in a more flexible way than permanent merging. In this case querying for same objects brings up more complexity, but can be modified with the annotation update mechanism of the RI and when it becomes stable then a merge procedure may take place.

Finally, constraints and rules on Merge and Same-as procedures were defined.

8.3 Deviation from work plan

There was no deviation from the work plan. The fourth year work plan has been fulfilled according to the DoW.

9 T3.7 Implementation of a Long term Digital Preservation Management Tool

9.1 Work planned

Our principal plan for the next period was to continue to evaluate the preservation manager and ensure its continued integration with the repository infrastructure as it and its associated API develops.

9.2 Work performed

The long term preservation component was completed prior to the end of the previous review period but conflicting timescales meant that integration with the latest version of the Repository Infrastructure (RIv3) was not completed at that point. Therefore a small amount of additional investment has been made in this reporting period to complete integration with the latest version of the software.

This has been reasonably straightforward and comprised mainly a rewrite of RI connectivity code within the long term preservation tool. The corresponding class (CoformRIManager) was edited to use the coform.client package and the newest version (1.4.1) of the Ingestion Tool's SimpleIngestionHandler. Since both the previous and current versions of the long term preservation component utilised the Ingestion Tool (developed at FORTH), albeit different versions, this helped to ensure the integration was completed satisfactorily. Authentication tokens and URLs are similarly stored in a properties file that is based on the one used by the Ingestion Tool.

A limited amount of time was spent supporting and informing the process of long term preservation evaluation conducted as part of WP10. This amounted mostly to explaining the long term preservation component and offering further contextual information regarding the Repository Infrastructure and its role in 3D-COFORM information workflows.

9.3 Deviation from work plan

None.

10 T3.8 Implementation of Watermarking Tool

10.1 Work planned

This task was concluded in Year 3. In the following the final algorithm is presented.

10.2 Work performed

The watermarking algorithm has been adapted to the MeshLab user needs of using generic 3D mesh models (please refer also to Deliverable D5.4 – Final Report of WP5 – 3D Artefact Processing). A frequency based approach on the vertices coordinates has been chosen obtaining a format independent tool. The use of a Neyman-Person detector based on the vertices positions makes the system robust to slight decimation of the model and changing of topological connections.

Hereafter a description of the whole final algorithm (embedding and detection phases) is reported:

The embedding rule

The watermarking embedding algorithm works according to two parameters:
- ‹ a secret key K
- ‹ a coefficient γ determining the strength of the watermark.

An overall picture of the watermark insertion process is shown in the following Figure 4.
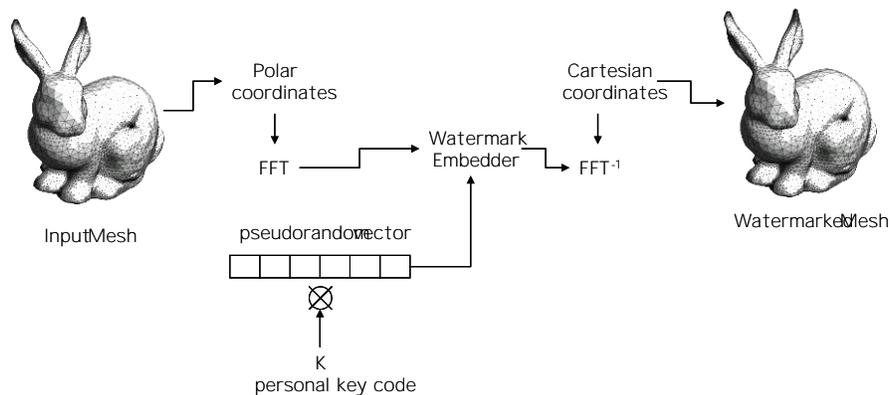
Figure 4: Watermark insertion process

In particular, the following steps are performed:
- ‹ A watermark-dependent structure called watermarking vector is generated according to K
- ‹ A polar coordinate system is considered for the input mesh
- ‹ The vertices of the mesh and the corresponding watermark element are used by the embedding algorithm to embed the watermark by altering the modulus of a subset of the FFT coefficients
- ‹ The watermarked FFT coefficients are used to reconstruct the watermarked model

25

Watermark detection

During the detection the user specifies the personal key K and the detector provides a positive or negative answer. To do so the watermarking vector is generated by starting from K, then the FFT transform of the modulus of each of the vertices are evaluated, and the correlation ρ between the watermarking signal and the FFT coefficients is calculated. If ρ is greater than a certain threshold Tρ the watermark is present, else the model is declared non-marked. The value of Tρ is obtained by means of statistical considerations. Specifically the Neyman-Pearson criterion is adopted, which consists in maximizing the missed detection probability for a given probability of falsely revealing the watermark in a non-marked host.

10.3 Deviation from work plan

No deviation from the plan. The task is concluded.

11 T3.9 Integration and testing

The objective of this task is to coordinate and enable the integration of the different components developed within the project such as integrating tools and tool chains, and to test the interfaces and communication between the tools for data and workflows provided by Cultural Heritage (CH) partners.

11.1 Work planned

This task followed the general technical objective for the fourth period of the project, in other words, consolidation of the development. Thus, the final integration and the corresponding testing of the tools would be accomplished in this period. The plan was to have the practitioners heavily involved in this process, in order to generate realistic scenarios / stories, serving for reliably finalizing the development process.

11.2 Work performed

The integration of the technical development was achieved at all levels: workflow, system and component. The first level of integration was achieved by means of defining stories and scenarios. The second level of integration was based on deployment experiments. The third and final level of integration was conducted in professional environments with practitioners and for typical CH activities: investigation, reconstruction, learning, replication, interpretation, and sharing.

Testing activities were conducted for integrated tools and within integrated workflows. These activities were supported with the corresponding documentation and the collection of feedback from the participants. Some of these activities were steered for the preparation of the content for the travelling exhibitions and for the 3D-COFORM Reshaping History Exhibition. The IVB was specifically tested in three events:

‹ Testing Workshop at the V&A in London, June 25 and 26, 2012

‹ Resources Room at the 3D-COFORM Reshaping History Exhibition in Brighton, July 27 to August 25, 2012

‹ German National Summer School in Schmitten, September 3 and 4, 2012

The results of these exercises led to the consolidation of the IVB, which is now much more matured and stable for professional use.

An Ingestion Tool training session was also scheduled during the Testing Workshop at the V&A in London, June 25 and 26, 2012. Important feedback on the use of the Ingestion Tool was provided by the users.

A significantachievement of this period is the export of 3D content to Europeana. The system integration and the deployment experiments allowed for an automatic mechanism of providing Europeana compatible metadata and 3D artefacts for technology demonstration.

11.2.1 Watermarking Tool testing

The process of testing the 3D watermarking tool developed within T3.8 continued during the fourth and last year of the project. Again, attention has been paid to the crucial issues of robustness and visibility. In particular, a superior robustness has been experimented against the attack of changing topological connections.

11.2.2 Testing at UBonn

During the course of Year 4 UBonn has been using the RI and accompanying tools for several different aspects. First, for implementing the material CRI, the latest version of the RI-C++ API was used to implement the material-indexer. The material-indexer is a service that periodically queries the RI for UUIDs of objects in the semantic network that are a ".btf"-file using the SPARQL interface, and retrieves the corresponding files for not yet indexed entries from the OR. Second, the implemented material CRI itself offers a webservice that is used by the RI to allow material search using query-by-example. For this, a standardized communication protocol, using GET-queries and XML-responses, was agreed upon between FORTH and UBonn, closely following the protocol that is used for the shape-CRI. Finally, several versions of the Java-based ingestion tool developed by FORTH have been used to ingest Multiview Dome data into the RI. During these ingestion procedures, developers at UBonn were in close collaboration with their colleagues at FORTH. Bugs and feature requests were reported immediately, which in turn directly led to improvements in the next versions of the ingestion tool. Both, the IVB and the CORExplorer were used to double check the results of the ingestion process.

11.2.3 Stand-alone deployable RI for testing

In order to provide a stand-alone deployable RI to serve testing and training purposes, following issues were addressed:
- ‹ Created MS SQL Initial Scripts for:
  - generating all RI tables
  - creating MS SQL user and set permissions for database access
  - initializing the table with RI basic data e.g. Mimetypes, Root Group and User
  - inserting training system related initial data e.g. Training User, Training Location, Permissions
- ‹ Copy the existing RI-Debug VM (Windows 2008 R2 Server)
- ‹ Find in OR-Server and OR-Location source dependents from OR-Central and OR-Debug and make it configurable
- ‹ Clean the VM (Clear all ingested data, uninstall all unimportant software)
- ‹ Initialize the MS SQL 2008 Database for training (using created MS SQL scripts)
- ‹ Configure the IIS 7.5 Webserver of the new RI as Training system

‹ Export to a VMware Player usable Version

‹ Configure the VM and the host system (IP addresses, Ports, Firewall)

‹ Install QMMR on host system

‹ Configure QMMR for communicating with OR Training VM (IP addresses, Ports)

‹ Initialize QMMR

‹ Test the functionality and the usability of the APIs (C++ and Java) e.g. login, ingest, upload, download

‹ Test the solution to work with CORExplorer, MetadataGenerator, Ingestion Tool, Coyote, PowerShell

‹ Clean the solution and make it deliverable

‹ Write an install manual

11.3 Deviation from work plan

No deviation from the work plan. The fourth year work plan concerning T3.9 has been fulfilled according to the DoW.

12 Advance in State of the Art

The implicit assumption in existing working repositories is that the object to be stored has undergone all possible processing steps and reached its final stage. Repositories that store only the final product are well supported by models such as OAIS and ORE [DTT*10] [PBH*10].

In contrast to that, the innovation of the RI is providing an infrastructure for the primary storage space in a working environment for digital object production, and a primary place of tools integration. Here things are created and not "imported" as in an OAIS repository, and Submission Information Packages (SIP) for export can effectively be defined.

From this fundamental idea all requirements, their successful implementation and validation in this project emerged. No competitive product of this kind exists.

[DTT*10] Doerr, M., Tzompanaki, K., Theodoridou, M., Georgis, Ch., Axaridou, A., & Havemann, S. (2010). A Repository for 3D Model Production and Interpretation in Culture and Beyond. *Proceedings of VAST 2010: The 11th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, 21-24 September, 2010, Palais du Louvre, Paris, France. (pp. 97-104). Aire-La-Ville, Geneva, Switzerland: Eurographics

[PBH*10] Pan, X., Beckmann, Ph., Havemann, S., Tzompanaki, K., Doerr, M., & Fellner, D.W. (2010). A Distributed Object Repository for Cultural Heritage. *Proceedings of VAST 2010: The 11th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, 21-24 September, 2010, Palais du Louvre, Paris, France. (pp. 105-114). Aire-La-Ville, Geneva, Switzerland: Eurographics

13 Publications

Awards:

‹ The paper [PSS*12] has won the best paper award on EuroMed 2012.

[TD12a] TZOMPANAKI, K., & DOERR, M.: Fundamental Categories and Relationships for Intuitive querying CIDOC-CRM based repositories. TECHNICAL REPORT: ICS-FORTH/TR-429, 2012

[TD12b] TZOMPANAKI, K., & DOERR, M.: A New Framework For Querying Semantic Networks. Museums amd the Web 2012: the international conference for culture and heritage on-line. April 11-14, 2012, San Diego, CA, USA

[SSD*11] Peña Serna, S., Scopigno, R., Doerr, M., Theodoridou, M., Georgis, Ch., Ponchio, F., & Stork , A. : 3D-centered media linking and semantic enrichment through integrated searching, browsing, viewing and annotating. Proceedings of VAST11: The 12th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage. October 18-21, 2011, Prato, Italy

[Tzo12] TZOMPANAKI, A.: Design and Implementation of a tool for formulating recall-oriented structured queries on semantic networks. Master of Science Thesis, Department of Computer Science, University of Crete, September 2012.

[PSH*12] PAN X., SCHIFFER T., HECHER M., HAVEMANN S., BERNDT R., SCHRÖTTNER M., FELLNER D. W.: A scalable repository infrastructure for ch digital object management. In 18th International Conference on Virtual Systems and Multimedia (VSMM) (Milan, Italy, Sept. 2012), IEEEXplore digital library.

[SHT*12] SCHRÖTTNER M., HAVEMANN S., THEODORIDOU M., DOERR M., FELLNER D. W.: A generic approach for generating cultural heritage metadata. In 4th International Euro-Mediterranean Conference on Digital Heritage (EuroMed) (Limassol, Cyprus, Oct. 2012), Springer LNCS (in press).

[PSS*12] PAN X., SCHIFFER T., SCHRÖTTNER M., HECHER M., HAVEMANN S., BERNDT R., FELLNER D. W.: An enhanced distributed repository for working with 3d assets in cultural heritage. In 4th International Euro-Mediterranean Conference on Digital Heritage (EuroMed) (Limassol, Cyprus, Oct. 2012), Springer LNCS (in press).

14 References

In this section, we list references to works conducted outside 3D-COFORM.