Our reference: PRESIOUS

Predictive digitization, restoration and degradation assessment of cultural heritage objects

D5.1 – Evaluation Plan

Project Ref. No.	FP7-ICT-2011-9 – FP7-600533
Project Acronym	PRESIOUS
Project Start Date (duration)	1 Feb 2013 (36M)
Deliverable Due Date	31 Jan 2014 (M12)
Actual Delivery Date	
Deliverable Leader	Theoharis Theoharis (NTNU)
Document Status	Final
Dissemination Level	PU

Address
Sem Sælands vei 5, Gamle fysikk 3rd floor
Gløshaugen, NTNU, NO-7491 Trondheim,
Norway

http://presious.eu

Contact person
Catalina Hellesø
catalina.helleso@ime.ntnu.no
+ 47 73 59 1452

Predictive digitization, restoration and degradation assessment of cultural heritage objects

Deliverable Identification Sheet

Project Ref. No.	FP7-ICT-2011-9 – FP7-600533	
Project Acronym	PRESIOUS	
Document Name	PRESIOUS-D5.1	
Contractual Delivery Date	31 Jan 2014 (M12)	
Deliverable Number	D5.1	
Deliverable Name	Evaluation Plan	
Type	Document (.docx)	
Deliverable Version	2.8	
Status	Final	
Associated WP / Task	WP5 / T5.1	
Author(s)	T. Theoharis	NTNU
	G. Papaioannou	AUEB-RC
Other Contributors	R. Gregor, T. Schreck	UKON
	M. Savelonas, I. Pratikakis	ATHENA-RC
	P. Mavridis	AUEB-RC
	P. Perakis	NTNU
Project Officer	Philippe Gelin	
Abstract	The evaluation plan should include a description of the methodology, planned experiments and tests, validation software and tools that will be used and created for the task, the timing of activities and identified data sets that are required.	
Keywords	Evaluation, plan	
Sent to Internal Reviewer	8 Jan 2014	
Internal Review Completed	20 Jan 2014	
Circulated to Participants	24 Jan 2014	
Read by Participants	27 Jan 2014	
Approved by General Assembly	31 Jan 2014	

Document Revision History

Date	Version	Author/Editor/Contributor	Summary of Changes
18/10/2013	0.1	T. Theoharis	First draft of evaluation report. Discussed in GA. First planning of methodology for the evaluation.
20/11/2013	2.2	G. Papaioannou, P. Mavridis	Reformatted the draft, added structuring and criteria by component, split the scenaria from the evaluation experiments, added content for WP2 and WP4, added figures for scenaria.
1/12/2013	2.4	T. Theoharis	Unified the document, distributed to partners for contribution
22/12/2013	2.5	R. Gregor	WP4 and WP5 contribution. Restructuring of WP4 experiments section.
24/12/2013	2.41	M. Savelonas, I. Pratikakis	WP2 experiments added
27/12/2013	2.51	G. Papaioannou	Formatted and edited WP4 experiments chapter.
2/1/2014	2.6	G. Papaioannou	Added the evaluation section of WP5 tasks. Pending contribution from UKON, ATHENA.
6/1/2014	2.7	T. Theoharis	Section 4.4.2 completed. Two 2.6 versions consolidated. TT comments added.
27/1/2014	2.8	T. Theoharis	Changes from partners consolidated

Table of Contents

PRESIOUS - Deliverable 5.1: Evaluation Plan

1. EVALUATION METHODOLOGY & TERMINOLOGY

As detailed in the description of work, there will be two evaluation phases; the first will focus on the algorithms developed within WP2, WP3 and WP4 (deliverable D5.4), while the second will focus on the integrated systems (deliverable D5.8).

The method of evaluation is based upon **application scenaria** (see section 3), i.e. realistic scenaria that are defined to address each of the 3 main project objectives (corresponding to WP2, WP3 and WP4) as well as the integrated systems (WP5). An application scenario is parameterized for execution by defining the data that it will operate upon and the ground truth data that it will be evaluated against; it thus forms an **experiment** (see section 4). The relationship between application scenaria and experiments is 1-many. A scenario and consequently, an experiment, is evaluated through **evaluation criteria** (see below) which provide the evaluation feedback. The execution of an experiment may involve **evaluation data** as well as **relevant evaluation software** or other **tools**; these may be off-the-shelf or purpose-built.

A number of **evaluation criteria** will be established per Work Package (WP2, WP3, WP4, WP5). These criteria will serve as **indicators** that will show the degree of success of the project outcomes, in accordance to the agreed upon Description of Work. Specific target values cannot be set from the outset as the ranges of the associated variables are mostly unknown at the time this plan is drafted. Reaching the state-of-the-art will be ensured by peer-reviewed scientific publications and project reviews. The evaluation criteria along with the experiments define a 2D evaluation matrix that determines which criteria will be applied to which experiments.

The **timing** of the evaluation is as follows. After delivery of the first modules of WP2, WP3 and WP4 (M18), all suitable experiments will be conducted and initial values of the applicable evaluation criteria will be derived. At this point it may become necessary to make adjustments to the evaluation criteria. A pre-delivery version of the integrated systems (WP5) will also be evaluated via experiments; note that the first version of the integrated systems is due on M24. The above will be reported at the First Evaluation report (D5.4 - M24). The final versions of the modules of WP2, WP3 and WP4 are due on M30 and the final integrated systems (WP5) on M36. The Final Evaluation report (D5.8 - M36) will report on the final experiments and will contain the final values of the evaluation criteria.

2. SUMMARY OF OBJECTIVES

For convenience, we provide below a brief description of the project's objectives, as described in Part B of PRESIOUS Description of Work.

O1. On-the-fly auto-completion for 3D digitization. The shape of digitized CH objects can be potentially predicted during acquisition, based on the gradually available partial scans of an object. The stream of input point cloud data from the acquisition source will be used to interactively retrieve and fit the closest matching candidate shape from parts of digitized artifact repository models as well as template models (categorized primitive objects), onto the acquired geometry, thus predicting and automatically suggesting the geometry for the parts not yet scanned. As the acquisition data are the starting point for all subsequent processing steps, their quality and reliability are significant issues. Corrections and fine details can be locally applied, where necessary, using localized complementary scans, effectively minimizing the overall time and cost of the scanning procedure or eliminating the need to post-process the data or attempt scanning in hard to reach surfaces of the original CH artifact. Furthermore, the project team will investigate prediction confidence metrics for the automatic suggestion of additional digitization passes. CH objects are especially good candidates for such a system since they can often be categorized, possess regularity, symmetries or repeated patterns and salient features. Furthermore, typical acquisition cases involve immovable, large or heavy to lift parts and fragments, which can be digitized in place, since inaccessible parts could be predicted though auto-completion. Finally, it can greatly assist low-budget scanning equipment and techniques in delivering smoother and richer geometric information.

O2. Estimation and prediction of monument degradation. Based on present-time surface shape, material measurements and environmental data, the project will investigate highly efficient techniques for forward and inverse deterioration prediction. This will allow to essentially move the artifact's surface condition "back and forth in time" and visualize the dynamic state of the deteriorating object, in the context of geometric and textural alterations. In order to include geometric information in the simulation model, PRESIOUS will conduct a number of timed, high-accuracy differential surface scans on the degrading monuments. Also, using the digitized data of the monument in its current state, similar surface regions will be retrieved and fitted to the degrading surface. This will act as an additional constraint for the simulation by providing an indication of the intact state of the object.

O3. 3D CH fractured object restoration and completion (missing parts synthesis). By exploiting existing CH objects in an example-based object restoration process, automated procedures will be developed for fractured artifact reassembly in three dimensions. This resembles the solution to a three-dimensional puzzle where the pieces are either intact or broken artifacts retrieved from relevant CH object repositories and the target result is predicted from approximate model templates, which act as constraining guides. The developed techniques will rely on retrieval and feature extraction techniques, similar in nature to the predictive scanning problem. Subsequently, novel techniques for the recovery (prediction) and automatic geometry generation of missing elements will be developed; the missing elements prediction will be conducted at multiple levels of detail (general shape, detail sculpting) and will thereby aid the physical repair process of the actual objects.

3. APPLICATION SCENARIA

The following scenaria are typical use-cases that correspond to the 3 main objectives of PRESIOUS as well as to the integrated systems.

3.1. Application Scenaria for objective 1 (On-the-fly auto-completion for 3D digitization)

Predictive digitatization mainly involves a retrieval and a reshaping stage, with each stage comprising one or more components. The proposed scenario will encompass experiments to evaluate each component individually, as well as the overall pipeline. In the latter case, the dependency of the pipeline performance on the performance of individual components, as well as the interdependency of component performances will be investigated. Figure 1, presents a high-level user-space workflow of the predictive digitization pipeline. The proposed scenario maps to this pipeline, emphasizing on the novel aspects of the project. The methods covered are related to WP2, but also involve the interactive parts of the integrated platform (*predictive digitization platform* – WP5). Given a partial 3D object, the retrieval stage performs a query on the database of 3D objects and retrieves a ranked list of objects that are geometrically similar to the input. This geometric similarity is quantified by means of a distance metric, calculated over the output of appropriately defined 3D shape descriptors. The reshaping stage determines the transformation that is uniformly applied to each one of the retrieved objects in order to align it with the partial scan of the input. Given one or more pairs of aligned objects, fitting and detail imprinting are performed, resulting in local deformations to the template object that minimize the distance between the two objects of each pair. Finally, blending is performed on the deformed objects with the initial partial scanned data to produce the final output of our algorithm, which is the predicted scan.

The proposed scenario will involve a database of pottery objects, created from the Hampson Archeological Museum collection. This choice was based upon the challenges induced by the particular objects as well as the availability of a large dataset of such objects which can support the operational pipeline of WP2.

The Centre of Advanced Spatial Technologies - University of Arkansas worked on the digitisation of numerous artefacts from the Hampson museum collection using a Konica-Minolta Vivid 9i short-range 3D laser scanner. The digitisation was performed at a precision close to 0.2 mm. The 3D digital replicas are covered by the creative common 3.0 license and are offered for online browsing or downloading in both high (>1M polygons) and low (<= 25K polygons) resolutions.

As a testbed for content based retrieval and partial matching experiments, 384 models of low resolution were downloaded from the website of the museum along with associated metadata information. Initially the models were classified by the museum into six general classes (Bottle, Bowl, Jar, Effigy, Lithics and Others). As the current classification did not ensure similarities based on geometry within a given class, an extended geometry-oriented classification was performed. The models were divided into 16 distinct classes, namely appendages, ball-like-short-neck, bottles-wide-body-long-neck, bottles-wide-prism-body-long neck, bottles-wide-body-medium-neck, bottles-wide-body-short-neck, ducks, flat-long, open, open-curved, other, pipes, sphere-like-long-neck, tooth, tripod base and twins. The queries will be: i) real partial scans, acquired from pottery objects, which are of representative objects from the Hampson collection and have been specifically constructed for this purpose, ii) artificial partial queries, created by slicing and cap filling a number of complete 3D objects from the Hampson collection, originating from those classes that are densely populated. In the first case, scanning will be performed by means of the Breuckmann scanner, whereas in the latter case artificial partial queries will have a reduced surface compared to the original 3D object.

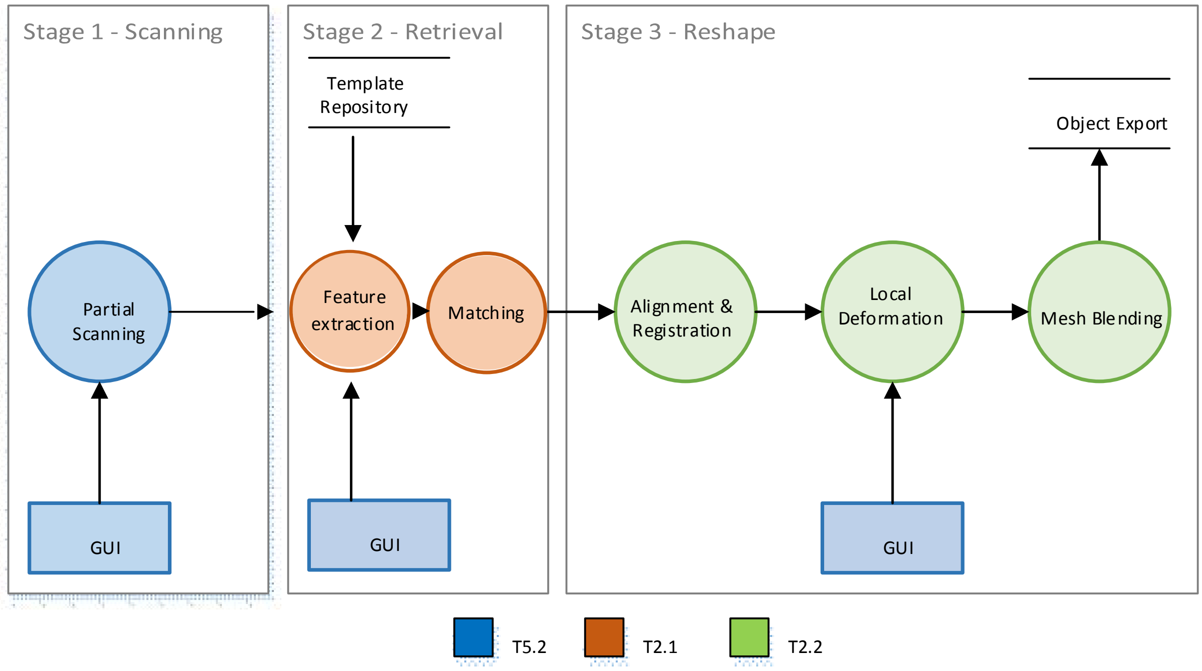

Figure 1. User workflow for objective 1 and mapping of its procedures to the tasks of WP2 and WP5.

3.2. Application Scenaria for objective 2 (Estimation and prediction of monument degradation)

WP3 consists of three components, which will be briefly described (see Figure 2). The main component is the *Erosion Simulator*, which simulates the erosion process on the surface or volume of an input stone. The *Stone Builder* is used to synthetically generate a specific type of surface or volume stone data and is used to either fill in plausible stone data that fits to given stone samples, or to synthesize the stone data from scratch for a specific stone type. The *Differential Geometry Measurer* is an auxiliary module that aligns subsequent scans of the same stone surface and measures the difference (erosion) between them; it is not subject to evaluation but it is an evaluation tool (s/w).

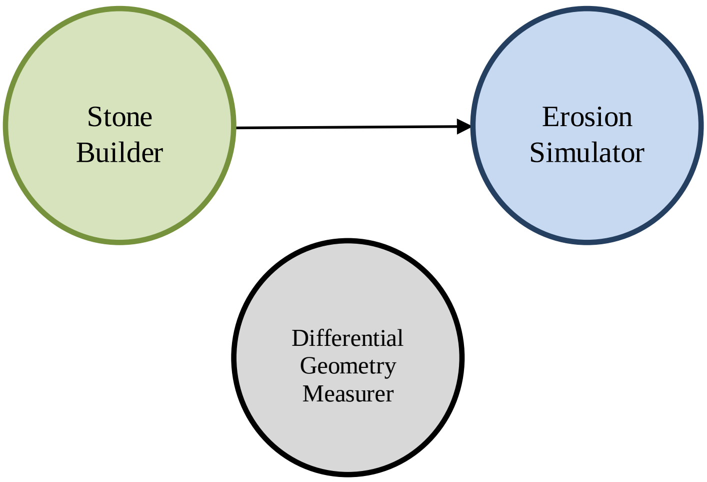

Figure 2. Erosion Simulator Components

The following scenaria concern the forward (in time) projection of the erosion prediction process. Reverse erosion prediction (going back in time) will be evaluated by swapping the initial and final datasets. However, since erosion is a non-deterministic physiochemical process with inherent loss of information, the reverse algorithmic procedure can only approximate the initial data and is unlikely to result in the initial data again. Furthermore, this process, i.e. going back in time – is complemented by the object repair process of WP4. We outline 3 scenaria below.

3.2.1. Object Erosion: Surface Geometry Only (E-GE).

In the Erosion-GEometry only (E-GE) scenario the input is geometric surface data of a known stone type. The erosion simulator produces the eroded surface geometry after certain period(s) of time, given a set of erosion parameters. The output surface can be displayed.

3.2.2. Object Erosion: Surface Geometry and surface physicochemical data (E-GESC).

In the Erosion-GEometry and Surface Chemistry (E-GESC) scenario, the input is physicochemical data samples over the surface geometry. The surface physicochemical data for larger areas may be synthetically generated from known sample points. The surface monitored will be limited to an area of a few millimeters if Scanning Electron Microscopy (SEM) is used for monitoring the changes in chemical composition and surface morphology. It may also be possible to use X-ray microtomography (micro-CT) for covering a larger area and assessing changes in the interior and exterior parts. The erosion simulator produces the eroded surface geometry and surface physicochemical data after certain period(s) of time, given a set of erosion parameters. The output surface can be displayed.

3.2.3. Object Erosion: Surface Geometry and volume physicochemical data (E-GEVC).

In the Erosion-GEometry and Volume Chemistry (E-GEVC) scenario, the input is physicochemical data samples over the volume of the object plus surface geometry. The volume physicochemical data may be synthetically generated from known surface sample points (obtained using SEM) or volume sample points (obtained using micro-CT). The erosion simulator produces the eroded surface geometry and volume physicochemical data after certain period(s) of time, given a set of erosion parameters. The output surface can be displayed. This scenario will be applied provided that the above volume physicochemical data are available.

3.3. Application Scenaria for objective 3 (3D CH fractured object restoration and completion)

In Figure 3, we present a high-level user-space workflow of the object restoration and automatic completion pipeline. The three proposed scenaria map to this pipeline, but each one emphasizes on a specific novel aspect of the project. The methods covered in the scenaria are related to WP4, but also cover the interactive parts of the final platform (*virtual repair and measurement platform* – WP5) which are involved in the presented use cases. **All components** of the pipeline are tested through at least one of the provided scenaria.

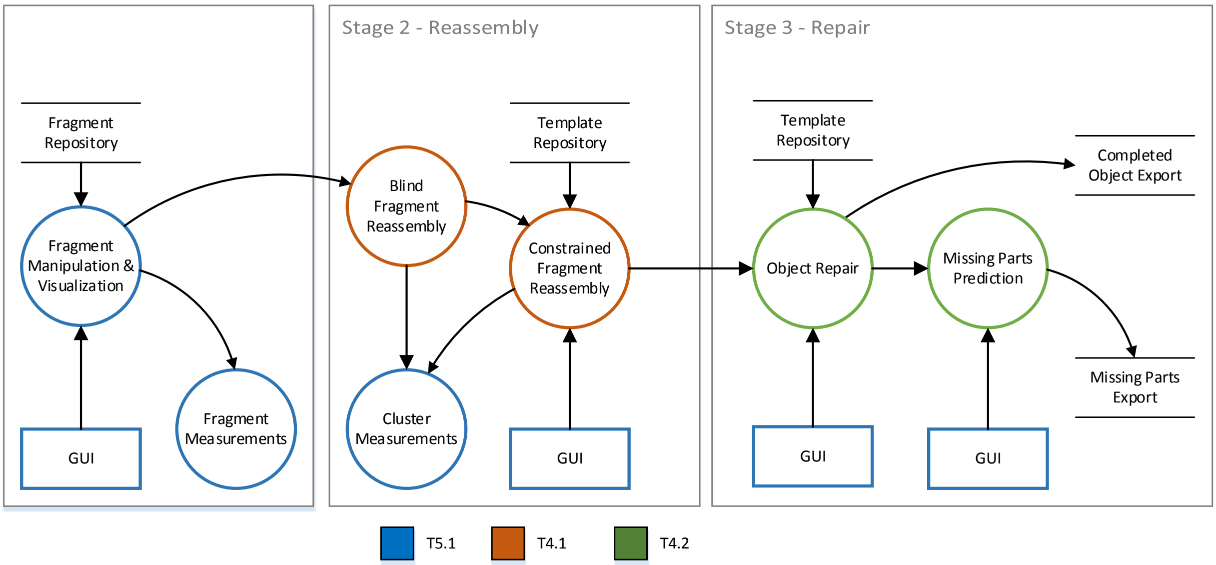

Figure 3. User workflow for objective 3 and mapping of its procedures to the tasks of WP4 and WP5.

3.3.1. Shape-guided reassembly

One of the scientific goals of PRESIOUS is to combine object matching techniques with retrieval and feature extraction methods in order to enable the reassembly of fragmented artefacts that have been severely damaged or whose parts are missing. This functionality is embodied within the current scenario, where apart from fragments with contact surfaces that can provide adequate support for geometric matching based on them alone, we will attempt to reassemble loosely coupled fragments. The test objects will be incomplete, consisting of fragments that do not precisely match, have large gaps between them or, in the worst case, are disjoint. To solve such a 3D puzzle, we will demonstrate how external criteria such as external surface features and predicted or user-provided overall shape of the reconstructed object may contribute to this effort. This scenario essentially maps to T4.1 and the first stages of an object repair workflow using the restoration environment of T5.1, including the GUI interaction. Figure 4 highlights the components that are involved in the scenario.

From the perspective of a user, the scenario involves the following tasks in the virtual repair and measurement platform:

- Objects (fragments) from one or more collections are loaded into the workspace. The objects may be annotated and searchable / categorized, although this is not a requirement for the subsequent stages (we may only have the raw geometric data of scanned objects). The user may additionally limit the “reassembly set” to specific objects only by selection, or can even manually provide some known associations between fragments.
- Objects can be manipulated using 3D gizmos and transformation tools, they can be measured, have their attributes, identified fractured faces etc displayed and can be also manually aligned (with computer-assisted snapping of fractured surfaces).
- At any time, given the current reassembly set, the user can initiate a “blind” reassembly, i.e. a geometric reassembly with no external constraints (see next).
- At any time, usually when the results from the blind reassembly are not satisfactory due to major fragment incompatibilities, the user may request a constrained reassembly using one or more of the specialized algorithms developed for T4.1 that take into account one or more of external features such as a “guide” object (user-defined or based on shape similarity), or surface features of the fragments.

- After a reassembly session or an assisted manual pairing, the user can assess the validity of a solution, by measuring object inter-penetration, gaps (matching error) and by visually inspecting the fragment positioning.

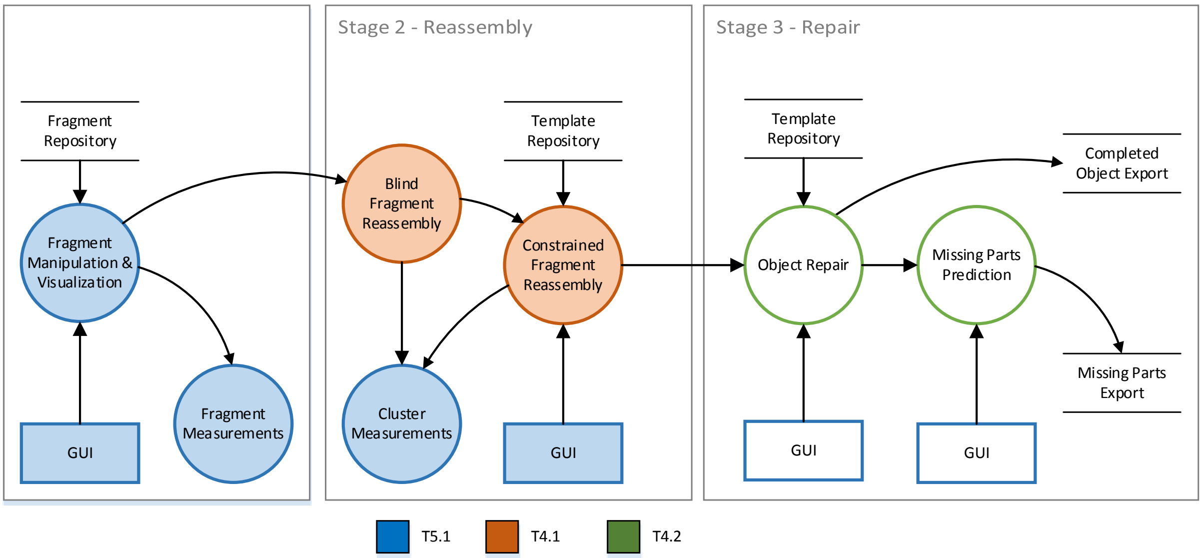

Figure 4. Components involved in scenario 3.3.1.

3.3.2. Object repair

After digitized fragments of objects have been reassembled[1] in stage 2, according to T4.2, the *Object Repair Scenario* aims to repair the remaining defects of the partial object (i.e. the result of Reassembly Scenario 3.3.1). As a result, a *single mesh* shall be computed that represents the completed, repaired object. The repair scenario is targeted at defects that have not been addressed by previous processing steps. Also all aligned fragment shapes have to be merged into a single shape. Figure 5 provides an overview of the involved tasks.

[1] more specifically: the reassembly scenario computes an aligned set of fragments

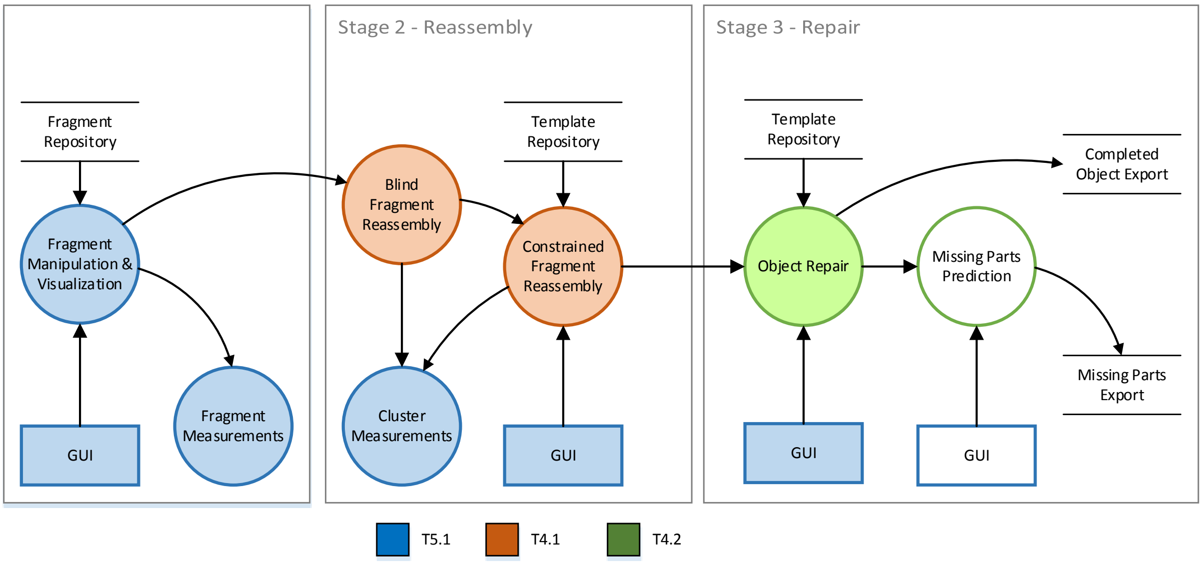

Figure 5. Components involved in scenario 3.3.2.

From the user's point of view, the following sub-tasks are accomplished in the Object Repair Scenario:

- Missing large parts: Not all fragments of a CH object might have been recovered and digitized, thus the repair scenario aims to fill in these missing parts based on the retrieval of similar shapes from a shape repository. Depending on the available shapes in the repository, this can be achieved in a partially automated workflow. It is however expected that the user is required to provide additional information that aims in finding, selecting or synthesizing surfaces that approximate the overall shape without such missing large parts.
- Fracture edges: Fragment shapes are not modified in Stage 2. However, between complementary aligned fragments, there exist a certain number of surface patches that correspond to the (exposed) fracture geometry itself and do not belong to the original, intact surface. These patches have to be removed.
- Merging of shapes: Subsequently, all remaining surface patches have to be merged into a single shape.
- Removal of small scale defects: After relevant input surface parts have been merged into a single shape, a certain number of them will contain surface patches that are affected by smaller defects of the fragment of the physical object. Yet still the corresponding surface patches are potentially located near the exterior surface of the completed object. Hence these patches have to be identified – either by user input or by automated heuristics - and removed in the Object Repair Stage.
- Inpainting: Following the removal of small-scale-defect areas, the merged surface patches are expected to contain holes and gaps. In addition, certain parts of the intact surface might be derived from a template shape of a rather primitive nature that does not provide realistic small-scale features while still encoding plausible information of global shape curvature (see next scenario). The inpainting task synthesizes small-scale local shape features on both, the aforementioned holes and gaps as well as surface areas that stem from rather primitive template shapes without a plausible amount of small scale local features. By default, this synthesis of small-scale features relies on local features that are present on adjacent surface areas. Alternatively, the user might choose a different strategy as e.g. a symmetry-based approach or manual selection of source and target areas.

3.3.3. Missing parts prediction

After the computation of a "completed" object in the object repair scenario, the Missing Parts Prediction Scenario is aimed to compute missing parts, based on the aligned fragments that are computed in Stage 2 (Object Reassembly) and the "completed" object from the preceding scenario. Essentially, the missing parts can be computed as the difference of the partial object (i.e. the aligned fragments) and the "completed" object. Subsequently, this difference then has to be cleaned from segments of too small diameter. Such segments could e.g. result from the small volume between two closely aligned fragments. Figure 6 illustrates the involved Stages for computing the missing parts.

These segments could then be used for the production of physical parts (e.g. by printing or milling) to produce physical artefacts that could in turn be combined with the existing physical fragments of the object to reassemble the originating CH object.

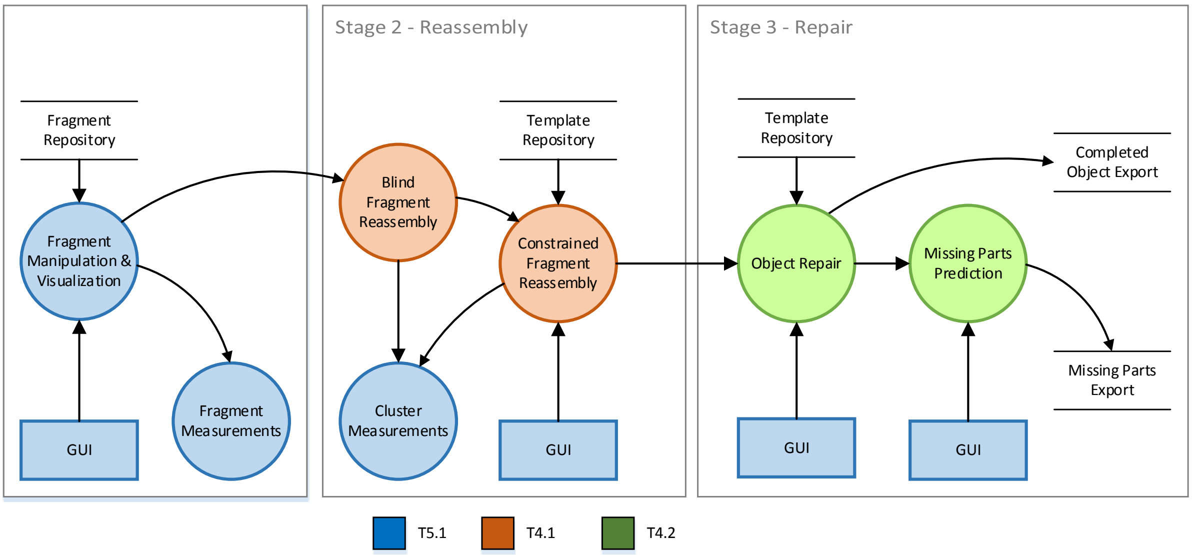

Figure 6. Components involved in scenario 3.3.3.

4. EVALUATION PROCEDURES AND EXPERIMENTS

In this section we define the evaluation procedures and experiments that will be performed per work package, giving the appropriate evaluation criteria. Each work package consists of a number of **components** to be tested. The evaluation criteria are applied per application scenario and per component.

4.1. Experiments for WP2 (Predictive 3D Digitization)

In order to ensure the robustness of our system, it is essential to first ensure that each component of the pipeline works as expected. To this end, in the remainder of this section we present a set of experiments that test the functionality of each component separately. One exception to that is the evaluation of the retrieval stage, which although comprises two distinct components, one for the extraction of 3D descriptors and one for the matching of the query objects with the repository objects, will be evaluated as a whole. Finally, we present a set of experiments that evaluate the functionality of the complete system.

4.1.1. Retrieval Component Evaluation

The task of this component is to retrieve from a database a ranked list of objects that have similar characteristics/features as the input object.

Input Data:

- Real partial queries, acquired by scanning pottery objects with the use of the Breuckmann scanner. The pottery objects used as queries are copies of representative objects of the Hampson collection and have been specifically constructed for this purpose.
- Artificial partial queries, created either with plain slicing or viewpoint-dependant slicing of template 3D objects from the Hampson collection. In the latter case, the partial queries will be similar to a range scan.

Further categorization of the input data result from the following considerations:

- The employed partial queries, either real or artificial, may or may not include distinctive object parts (e.g. handles of a bowl).
- The employed partial queries may or may not comprise disjoint parts.
- Queries could either be represented as a mesh or as a point cloud.

The artificial queries will have various degrees of partiality, indicatively ranging from 25% to 40%, with a step of 5%. In the case of real scanning, partiality will be determined by the scanner field-of-view.

Output Data: A ranked list of objects.

Evaluation Data (Ground truth): To validate the correctness of the retrieval component we will use as ground truth the classification of the pottery database, which is described in Section 3.1.

Validation Method: Validation will be performed by taking into account the type of the partial query:

- Given a partial query artificially created from an existing complete object in the database, the retrieval should return a ranked list, where the top position is held by the corresponding complete object, followed by objects of the same class.
- Given a partially scanned query of a pottery object, the retrieval component should return objects of the same class.

The retrieval performance will be assessed by means of precision-recall (P-R) plots and five quantitative measures: nearest neighbor (NN), first tier (FT), second tier (ST), E-measure (E) and discounted cumulative gain (DCG). For every query object that belongs to a class, recall denotes the percentage of models of the class that are retrieved and precision denotes the proportion of retrieved models that belong to the class over the total number of retrieved models. The maximum score is 100% for both. NN is the precision at the first object of the retrieved list, when C objects have been retrieved, where C is the number of relevant objects to the query. In a similar fashion, ST is the precision when 2C objects have been retrieved. E-measure combines precision and recall metrics into a single number and the DCG statistic weights correct results near the front of the list more than correct results later in the ranked list under the assumption that a user is less likely to consider elements near the end of the list. In addition to these retrieval performance measures, high precision timers will be used for assessing execution time of the retrieval component.

Relevant tools: The retrieved ranked lists will be visualized as thumbnails, along with the quantitative measures mentioned. Each thumbnail could be further inspected in its 3D form via a 3D mesh or point cloud viewer.

4.1.2. Registration Component Evaluation

The task of this component is to correctly align two input objects. The first mesh will be transformed in order to be aligned with the second one.

Input Data: Two meshes. For this synthetic test, the first mesh for this experiment will be a subset of the second mesh, randomly positioned in space. This methodology ensures that we have a well-defined and known ground truth solution.

Output Data: The transformation (rotation, translation, scale) that aligns the two objects.

Evaluation Data (Ground truth): The ground truth solution in this experiment is the original position of the first mesh.

Validation Method: To validate the correctness of this component we will compare the ground truth solution with the solution proposed by the algorithm. When the ground truth is known in advance, this can be performed by directly examining the two transforms. In this case, the components of the transforms should be nearly (within an expected error range) identical. In the general case, where the ground truth is unknown or difficult to determine, the solution can be verified by calculating the minimum, maximum, average and/or median distances between the transformed meshes. In the ideal case, these measures should be near zero. Finally, the results can also be inspected visually, in order to ensure the correctness of the results.

Relevant tools: For the visual inspection of the aligned meshes we will use a 3D mesh viewer. The exact error (distance) measurements will be provided by the software modules developed for T2.2.

4.1.3. Deformation and Detail Imprinting Component Evaluation

This component deforms an input mesh (source), in order to match a second mesh (target). The deformation should be performed in such a way that minimises the distance between the source and the target, but also it should not create excessive stretching/distortion on the source mesh. Since both the source and target meshes are expected to have outliers, holes and other imperfections from fuzzy measurements, the local deformation procedure will not enforce interpolation of the target mesh by the surface of the source mesh, since this can lead to highly distorted results.

Furthermore, this deformation stage will potentially involve the imprinting or in-painting of geometric details to the template object, in order to match the appearance of the input scanned data.

Input Data: Two meshes, source and target.

Output Data: The deformed mesh or a non-rigid transformation that leads to the result.

Validation Method: To validate the correctness of this component, we will directly measure the minimum, maximum, average and/or median distances between the two meshes. Ideally all these measures should be near zero, thus the source interpolates the target data. Furthermore we will measure the distortion of the source mesh after the deformation. The distortion should be uniform around the mesh, without many peaks. Finally, the results can also be inspected visually, in order to ensure the correctness and the quality of the solution.

Relevant tools: For the visual inspection of the aligned meshes we will use a 3D mesh viewer. The exact distance and distortion measurements will be provided by the software modules developed for T2.2.

4.1.4. Blending Component Evaluation

This component combines two or more meshes to one continuous mesh without any discontinuities. In our pipeline, this component will take the predicted/deformed surface and it will merge it with the initial input scan data.

Input Data: Two or more meshes.

Output Data: The final combined mesh.

Validation Method: The output mesh should be smooth and continuous. Ideally, we should be able to mathematically prove that the blending method ensures the continuity of normals (C^1 continuity) and the continuity of curvature (C^2 continuity) in the generated mesh. If this is not possible, another way to validate the quality of the results is to detect and measure the discontinuities on the final mesh. Finally, the results can also be inspected visually, in order to ensure the correctness and the plausibility of the solution.

Relevant tools: For the visual inspection of the aligned meshes we will use a 3D mesh viewer. The exact continuity measurements will be provided by the software modules developed for T2.2.

4.1.5. Complete Pipeline Evaluation

This experiment evaluates the validity of the complete system. We will start scanning a physical 3D object, the partial scans at various stages of completion will be fed to our system, the output will be recorded and will be compared to finished scanned object.

Input data: Partial scans at 25%, 50%, 75% completion. The percentages mentioned here are indicative.

Output data: The predicted object for each stage of completion.

Evaluation Data (Ground truth): The complete (finished) scan.

Validation Method: The output of our algorithm (predicted scan), should match the ground truth (finished scan). We will examine whether this is true at various stages of scan completion. To measure if the two meshes match, we will measure the minimum, maximum, average and/or median distance between the two surfaces. Initially we expect that the distance (prediction error) will be large, but it should gradually converge to zero, as the scanning procedure advances and the algorithm is fed with more data. The rate of this convergence is also a measure for the performance of our method. Finally, the results can also be inspected visually, in order to ensure the correctness and the plausibility of the solution.

Relevant tools: For the visual inspection of the aligned meshes we will use a 3D mesh viewer. The exact distance and distortion measurements will be provided by the software modules developed for T2.2.

4.1.6. Evaluation Criteria

Correctness: The correctness of the results will be validated by a direct comparison against the ground truth. In the experiments described in 4.1.1, the retrieval performance will be quantified by means of precision-recall (P-R) plots and five quantitative measures: nearest neighbor (NN), first tier

(FT), second tier (ST), E-measure (E) and discounted cumulative gain (DCG). In the experiments described in 4.1.2-4.1.4, the distance between the ground truth and the given solution will be measured as the minimum, maximum, average and/or median distance between the two meshes.

Objective Quality: For the experiments where we don't have a ground truth solution, we have described exact measures of the quality of the results (distance, distortion, continuity). For more details the interested reader is referred to the description of each experiment.

Subjective Quality (Plausibility): The subjective and therefore qualitative acceptance of a result as realistic and seemingly correct under inspection by an expert.

Speed: Total execution time, measured for the complete system and for each component separately. High precision timers will be used for this task.

Scalability: The impact on the algorithmic complexity of a method of the linear increase of one of the problem's parameters, such as number of data tokens, solution granularity etc. Expressed in theoretical complexity notation and validated through experiments (actual measurements).

Robustness: The ability to tolerate errors introduced in the input of an algorithm. It also involves the reproducibility if the results under different starting conditions for the same input set. It affects all algorithms developed.

User Intervention: The amount of time, gestural syntax and level of skills required for a user to provide (manual) input to a process. The smaller the user interaction, the more automated is the process.

Table 4.1.6 summarizes the connection between the above criteria and the experiments described.

	Correctness	Objective Quality	Subjective Quality	Speed	Scalability	Robustness	User intervention
Retrieval Evaluation	●		●	●		●	
Alignment Evaluation	●		●	●	●	●	●
Local Deformation Evaluation		●	●	●	●	●	●
Mesh Blending Evaluation		●	●	●	●	●	
Complete Pipeline Evaluation	●		●	●	●	●	●

Table 4.1.6. Mapping of evaluation criteria to the predictive digitisation experiments.

4.2. Experiments for WP3 (Object degradation prediction)

The experiments planned for WP3 follow and are based on the application scenaria defined above. The geometrical and physicochemical measurements on actual erosion can only be taken once, while the erosion simulator can be run multiple times and might produce different results if it involves stochastic processes. Therefore the ground truth data from the differential scans and the accelerated aging chamber experiments may be compared against a statistical combination of several erosion simulator runs.

4.2.1. Experiments based on application scenario E-GE

E-GE-1: The first set of differential scan data from Nidaros Cathedral (obtained at the beginning of the project) are used as input to the erosion simulator which is parameterised and run for ~18 months (time of second differential scan) and ~36 months (time of third differential scan). The simulator output is geometrically compared against the second and third differential scans from Nidaros Cathedral.

E-GE-2: As **E-GE-1** but based on the differential scans of the Elefsis column.

E-GE-3: The surface geometry of the Nidaros slabs and Elefsis slabs is measured. The erosion simulator is parameterized and applied to the above data. The slabs are subjected to accelerated aging and their surface geometry is measured again. The measured values are compared against the erosion simulator output. This experiment is a safety net in case no measurable geometric erosion is obtained from the differential scans in **E-GE-1** and **E-GE-2**.

Validation procedure: The differential scans on Nidaros Cathedral and the Elefsis column as well as the geometrical scans before and after the accelerated aging on the probe slabs, are used as ground truths to validate the geometrical accuracy of the erosion simulator.

Evaluation data (ground truth): differential scans on monuments or probe slabs.

Relevant tools: Differential Geometric Measurer (DGM): this piece of software aligns 3D meshes and measures the difference between them. It will be written at NTNU.

4.2.2. Experiments based on application scenario E-GESC

E-GESC-1: Geometrical and physicochemical data is measured on the surface of the Nidaros slabs. The erosion simulator is parameterized and applied to the above data. The Nidaros slabs are subjected to accelerated aging and geometrical and physicochemical data is measured on the surface of the eroded Nidaros slabs. The measured values are compared against the erosion simulator output.

E-GESC-2: As **E-GESC-1** but based on the Elefsis slabs.

Validation procedure: accelerated aging on probe slabs (compare surface geometry and physicochemical data before / after, e.g. via surface chemical analysis).

Evaluation data (ground truth): differential scans and physicochemical analysis of probe slabs before/after accelerated aging.

Relevant tools: DGM (see above). Stone Builder (SB): this piece of software will synthetically generate surface/volume physicochemical data of given stone types and map them onto the surface/volume of a 3D object. SEM, micro-CT.

4.2.3. Experiments based on application scenario E-GEVC

E-GEVC-1: Geometrical and physicochemical data is measured on the surface/volume of the Nidaros slabs. The erosion simulator is parameterized and applied to the above data. The Nidaros slabs are subjected to accelerated aging and geometrical and physicochemical data is measured on the surface/volume of the eroded Nidaros slabs. The measured values are compared against the erosion simulator output. Note that physicochemical changes within the volume of the slabs can only be expected within a certain depth from the surface, which depends on porosity and chemical composition.

E-GEVC-2: As **E-GEVC-1** but based on the Elefsis slabs.

Validation procedure: accelerated aging on probe slabs (compare surface or volume geometry and physicochemical data before / after, e.g. via physicochemical characterization using SEM coupled with micro-CT).

Evaluation data (ground truth): differential scans and physicochemical analysis of probe slabs before/after accelerated aging. The physicochemical analysis within the volume cannot be as accurate as the physicochemical analysis on the surface.

Relevant tools: DGM (see above). SB (see above). SEM, micro-CT.

4.2.4. Evaluation Criteria

The evaluation criteria for the above experiments follow:

- **C-WP3-1. Erosion prediction accuracy against geometric ground-truths.** Accuracy is defined as a measure of the deviation of the predicted erosion against the ground truth geometric data in physical units (surface distance measures).
- **C-WP3-2. Erosion prediction accuracy against physicochemical ground-truths.** Accuracy is defined as a measure of the deviation of the predicted chemical composition of the eroded object against the ground truth chemical composition in terms of chemical concentration and distribution.
- **C-WP3-3. Erosion prediction geometric robustness** (in the presence of noise, holes and spikes, missing data). Robustness will be measured from the geometric accuracy deviation resulting from variations in the amounts of noise, holes and spikes and missing data. These variations will be synthetically created on the input data.
- **C-WP3-4. Erosion prediction physicochemical robustness** (in the presence of noise, missing data). Robustness will be measured from the physicochemical accuracy deviation resulting from variations in the amounts of noise and missing data in the chemical composition. These variations will be synthetically created on the input data.
- **C-WP3-5. Erosion prediction tolerance against object geometric resolution and scale variations** (macro vs microscopic data). Tolerance will be estimated from the different geometric resolutions and scales of the input data obtained from differential scans on monuments and probe slabs. Different geometric resolutions will also be produced by aggregating the original geometric data.
- **C-WP3-6. Erosion prediction tolerance against object physicochemical resolution and scale variations** (macro vs microscopic data). Tolerance will be estimated from the different physicochemical resolutions and scales that will be acquired from probe slabs using SEM or micro-CT. Different physicochemical resolutions can also be derived by aggregating the original physicochemical input data.
- **C-WP3-7. Erosion prediction procedure complexity and absolute processing time evaluation.** The erosion prediction algorithm complexity will be determined and the processing times will be experimentally measured.
- **C-WP3-8. Synthetic stone builder global and local consistency against real stone materials.** The stone builder module is used to synthetically create physicochemical stone data. Its physicochemical consistency will be measured against known stone materials, both globally (macroscopically) and locally (microscopically).

Table 4.2.4 shows the evaluation criteria applicable per experiment; the colour is a third dimension that shows the WP3 component to which they apply: stone builder=green, erosion simulator=blue.

	C-WP3-1	C-WP3-2	C-WP3-3	C-WP3-4	C-WP3-5	C-WP3-6	C-WP3-7	C-WP3-8
E-GE1	●		●		●		●	
E-GE2	●		●		●		●	
E-GE3	●		●		●		●	
E-GESC1	●	●	●	●	●	●	●	●
E-GESC2	●	●	●	●	●	●	●	●
E-GEVC1	●	●		●	●	●	●	●
E-GEVC2	●	●		●	●	●	●	●

Table 4.2.4. Mapping of evaluation criteria to the degradation experiments.

4.3. Experiments for WP4 (Object Repair)

The experiments that we have designed for WP 4 are component-driven, since their focus is on testing the variety of underlying algorithms and not examine the user-centric procedure itself (this is part of WP5 evaluation). Therefore, in the following text, we start with establishing a generic set of evaluation criteria first. Subsequently, for both tasks (T4.1) and (T4.2), a set of module descriptions is provided. For each of these modules, we specify a subset of the previously introduced criteria that will be measured for the module's evaluation. In addition, the module descriptions provide specific details on how certain criteria can be measured for the given module.

4.3.1. Evaluation Criteria

General Effectiveness

For each experiment in Stage 2 (Reassembly) and each processing step in Stage 3 (Object Repair and Missing Parts Prediction), general effectiveness establishes a basic, experimental measure of the fraction of input datasets, that actually lead to desired (i.e. valid) outputs for this experiment or processing step. Note that for individual experiments or processing steps, the evaluation of effectiveness can be further refined as described in subsequent criteria. Evaluation of general effectiveness implies that for each experiment / processing step a set of required input and output properties[2] along with a list of actually used input datasets will be specified.

General Efficiency

For each experiment or processing step related to the reassembly or repair scenario general efficiency relates to the resource consumption of an algorithm and will be determined both experimentally by measuring the runtime, memory consumption or other performance metric of a processing step across a range of input datasets, and theoretically, by evaluating the complexity of a given algorithm. The above measurements could then either be used directly as an absolute measurement (milliseconds / Megabytes) or relative scales when comparing alternate strategies (e.g. as done for the experiments for the reassembly stage). Resource consumption measurements are specific to the hardware (e.g. processor, amount of RAM installed, type of hard disk drive) and software platform (e.g. operating system, file system) that is used for testing. Hence experimental measurement implies a specification of those external test parameters. For a subset of experiments or processing steps, the evaluation of efficiency is further refined in subsequent criteria.

General Robustness

General robustness is related to the effectiveness of an algorithm and specifically refers to the ability to tolerate errors introduced in the input of an algorithm. It also involves the reproducibility if the results under different starting conditions for the same input set. It affects all algorithms developed in the reassembly stage. For processing steps in the object repair stage, a set of refined criteria is provided below.

Robustness against Missing Parts. Many processing steps in the object repair stage operate on a set of aligned input fragments. While some larger parts might be already missing in the available test datasets, robustness of these processing steps can be evaluated more comprehensively by omitting additional fragments from the input. This extension of the input data set will be used to obtain additional effectiveness ratings for robustness against missing parts depending on the number and size of omitted fragments.

2 more details concerning the design contracts of individual processing steps will be specified in D4.2 and D4.3 more details of the external design contract of the reassembly and repair stages will be specified in the system architecture specification D5.2

Robustness against Defects of Mesh Structure. Many components in the WP4 pipeline directly operate on mesh data. However as stated in Part A of D4.1, the mesh structure itself could contain a broad range of defects, where each defect class could potentially impact further processing in different ways. Coarsely, defects of mesh structure can be divided into violation of manifoldness and the introduction of noise. The former includes defects such as singularities and, in a stricter sense, self-intersections and violations of watertightness such as gaps, holes and boundary edges. The latter includes aliasing, sharp feature-chamfering, topological noise (i.e. changes in shape genus) and "spatial" noise that could e.g. be introduced by numerical instabilities. Depending on the input requirements[3] of individual processing steps, a selection of these defects will be artificially inserted into datasets to extend the available test data sets. These extended input data sets will then be used to obtain additional effectiveness ratings against specific mesh defect types.

Scale & Resolution Invariance. Scale and Resolution Invariance is a refinement of the robustness criteria. Many sources should in principle be exploitable for populating the various shape repositories used in the object repair stage. However different sources will provide digitized objects in non-uniform scale and resolution. Hence it is important that certain processing steps can simultaneously work with objects of different scale and resolution. To extend the available test data sets, scale and resolution of the available input data will be varied. The extended input data can in turn be used to determine additional effectiveness ratings. In addition to the general efficiency measures, it will be possible to better estimate the scalability of a processing step by using higher/lower resolution models to gain additional runtime and memory consumption measurements.

Introduction of Mesh Defects. Introduction of Mesh Defects is a refinement of the effectiveness criteria but is not related to the previously mentioned robustness criteria. Instead of relating to defects in the input, this criteria is related to occurrence of defects in the algorithms output. Many processing steps in the repair stage directly modify the input meshes. All of those steps could potentially introduce new defects of the mesh structure. When evaluating effectiveness and depending on the output requirements of individual processing steps, this will be reflected in the effectiveness rating of a processing step as well as separate measurement concerning the fraction of inputs that result in outputs with defects in mesh structure. This implies that additional tests of the output will be conducted to detect these defects.

Scalability

Scalability refers to the algorithmic complexity of a method. That is, the relative increase of required computational time or overall runtime for a certain increase of the input problem, such as number of data tokens, solution granularity etc. Scalability will be assessed, where possible, by theoretical analysis (complexity notation) and validated through experiments (actual measurements).

Discrimination.

The ability to disambiguate between valid and invalid results by a safe margin in a systematic manner. Used in the matching procedures within the reassembly stage to verify a valid solution and determine the fault-tolerance of a method. A good algorithm should present a large metric deviation when comparing a correct solution to an invalid one, in comparison to the difference of measurements between two valid solutions.

Retrieval Precision and Recall

Several processing steps in the repair stage retrieve objects from a shape repository. Precision and Recall are well-known measurements of the fraction of false positives and negatives for a content-

3 i.e. the design contract of the processing steps which will be reported in D4.2 and D4.3

based similarity search and will be experimentally determined for a specified list of queries and shapes in the repository.

User Input Requirements

Relying on user input will often increase the effectiveness of a particular task and is a flexible way to indirectly support the exploitation of very specific domain knowledge for object repair and assembly. Yet, for certain complex and ill-posed problems, relying on user-input might in many cases be the only way to actually address the problem. However, this might negatively impact the efficiency as well as the applicability of an approach. Gathering user input is likely to slow down the workflow, let alone the differences in individual, user-specific inclinations and preferences that lead to problems when multiple users are involved. When evaluating the user input requirements of a processing step, the number of required and optional user interactions per execution will be stated.

4.3.2. Fragment Reassembly Experiments

(Stage 2, Task 4.1)

Experiments in this section are designed in a progressive manner, from smaller problem scale to larger and from simpler to complex and more general cases. Four exemplary cases are defined, decomposing the reassembly into a focused pairwise matching experiment (0), two multipart experiments with a single target object and multiple target objects respectively (4.3.2.2, 4.3.2.3) and one with heavily degraded objects, requiring additional guidance apart from the contact surfaces themselves (4.3.2.4). The following table summarizes the connection between the above criteria and the experiments described in the subsections of 4.3.2. The generic criteria are specialized according to the experiment requirements and goals, where necessary.

	General Effectiveness	General Efficiency	Scalability	Discrimination	Robustness	Plausibility	User Input Requirements
0 Table 4.3.1. Mapping of general evaluation criteria to the reassembly experiments. Pairwise Matching	●	●	●	●	●		
4.3.2.2 Reassembly of a Single Object	●	●	●	●	●	●	●
4.3.2.3 Reassembly of Mixed Objects	●	●	●	●	●	●	●
4.3.2.4 Reassembly of Heavily Damaged Fragments	●		●	●	●	●	●

Table 4.3.1. Mapping of general evaluation criteria to the reassembly experiments.

4.3.2.1. Pairwise Matching

In this test, we focus on the basic algorithmic performance of our core techniques for the discovery of plausible combinations of fragment poses for the alignment of two fragments, regardless of them belonging to the same object or not. This is a generic test that evaluates the convergence speed, the plausibility of the set of provided solutions and the discrimination ability of correct and invalid poses.

This experiment maps to all three evaluation scenaria of objective O3.

Evaluation Procedure. The fragments are pre-processed to segment and classify their surfaces into *intact* and potentially *fractured regions* or *facets*. As intact regions are considered those surface patches that belong to the non-fractured part of a fragment, i.e. they (should) belong on the exterior of the final, reassembled object. The fractured regions consist of those polygons that have been marked as candidates for matching with the corresponding fractured regions of other fragments. A single fragment may have more than one fractured surface (see Figure 7). The fragment pre-processor is responsible for segmenting the fragments and classifying the segments accordingly, as an offline process. Bear in mind that the pre-processing stage, although important, is non critical in the sense that it allows for false positive detection of fragmented surfaces and only rejects parts of the fragment that are trivially smooth and/or flat. This surface classification is only used to drive the pairwise matching in a more focused set of potential solutions.

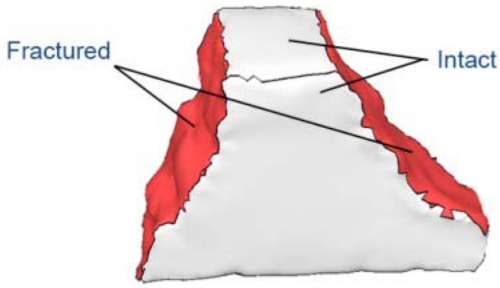

Figure 7. A pre-processed fragment. The surface is split into segments (facets), each classified as fractured or intact. The combinations of the fractured facets of two fragments comprise the input to the pair-wise matching procedure.

The matching procedure itself examines all combinations of facets of the two fragments at hand and outputs a set of values for the *matching distance*, inter-fragment *penetration volume* and *relative pose transformations* per facet pair. The tighter a match is, without having any fragment penetration, the better. However, it is also important to identify how well were the two fragments aligned under a resulting pose transformation. Important additional criteria are the discrimination power of an invalid facet match against a correct one, since these scores will drive the multi-part matching of the reassembly, and also the convergence speed and robustness of the pose estimation. The method should also scale well with respect to the geometric complexity of the given 3D models.

For the fragment matching, general effectiveness is derived via the amount of penetration between two fragments, their relative alignment and residual gap between them. Penetration assesses how much a fragment penetrates another during their mutual alignment in the reassembly stage. It is measured in volume units (mm^3). Smaller penetration means better alignment. This metric is used in both the pairwise matching experiment and the entire set of multi-part (reassembly) tests. In the latter case, apart from measuring the aggregate pairwise error, it also indicates how the alignment error is propagated along the fragment graph of a reassembly solution.

The fragment alignment is a qualitative measure indicating the correctness of fragment pose (rotational part) in the fragment matching and reassembly stage. It can be validated by visual inspection of a pair of fragments against a “ground-truth” manual alignment or visual consistency with the reassembled shape. For shape-guided reassembly, it is also possible to derive a set of quantitative metrics with respect to the overall shape, but this is specific to the implementation of the algorithm(s) and cannot be determined at this time.

The measurement of residual gap between two fragments is a quantitative aspect of the correctness criterion that can be applied to both the reassembly and the object repair stages. It measures the point to point aggregate distance (integral, i.e. residual volume) among the contact surfaces of two or more fragments.

Input. The segmented and classified facets of two 3D objects (fragments). These are exported by the fragment pre-processor in a ready-to-use custom format for the reassembly purposes.

Expected Result. A ranked list of matching scores, one per facet combination.

Relevant Tools. The quality of the results can be a) measured (penetration, fragment distance, i.e. contact gap) and b) visually inspected using one of the following: the stand-alone matching utility GUI (custom software), the resulting transformed objects when loaded into MeshLab or any other mesh viewer, the virtual workspace of T5.1 (only available at the final stages of the integration). Measurements are at least provided as output of the stand-alone matching utility. The prevailing score between two fragments can be also read through the GUI of the virtual repair and measurement platform of WP5.

4.3.2.2. Reassembly of a Single Object

The next assembly test after the pairwise matching in terms of increasing complexity is the multi-part joining of fragments belonging to a single entity. The final object needs not be complete, but both cases should be evaluated.

This experiment maps to all three evaluation scenaria of objective O3.

Evaluation Procedure. The user initializes a reassembly set with fragments that are known to belong to a single final (reassembled) object and the method tries to optimize the fragment combinations and their relative pose so that the final object is correctly represented by the pairwise rigid transformations, with as minimum inter-fragment penetration as possible and without skipping a part. The final solution should be plausible, which is easy to validate through comparison with a manually provided alignment of the pieces. Important factors are the robustness of the procedure in terms of alignment, noise tolerance, fragment degradation and penetration, as well as the scalability of the method itself with respect to the number of fragments presented.

Essentially, for each reassembly solution, a graph is built, potentially containing many islands. In this particular case however, where only one object is present, a single graph should be the result of the process. The robustness of the method is evaluated here as its capability to properly group the fragments correctly and without leaving isolated islands in the solution.

Input. The set of fragments to be used for the reassembly. Pre-evaluated combinations of fractured facets are used whenever available, or in the opposite case, they are calculated on the fly. The latter is hardly necessary however, since all the parts in a collection are known and can be processed offline.

Expected Result. A ranked list of proposed reassembly solutions, each one given as a set of object transformations (per fragment) and a set of “links” between fragments. The ranked list may be empty or contain a single graph, depending on the strictness of the quality settings.

Relevant Tools. The quality of the results can be a) measured (total penetration, total fragment distance, graph integrity) and b) visually inspected using one of the following: the resulting transformed objects when loaded into MeshLab or any other mesh viewer, the virtual workspace of T5.1 (only available at the final stages of the integration).

4.3.2.3. Reassembly of Mixed Objects

The generalization of the multi-part matching is the reassembly of more than one objects in a single reassembly set. For this mixed set of objects, there should be no guarantee about the number of final objects or the fact that all fragments actually contribute to a valid cluster. This test realizes a typical scenario of a large database of objects that have not been categorized by experts. This is the third reassembly test with respect to difficulty ranking.

This experiment maps to the first two evaluation scenaria of objective O3 (3.3.1, 3.3.2).

Evaluation Procedure. The user initializes a reassembly set with fragments that are known to belong to more than one reassembled object. Similar to the single-object reassembly, the final solution should be plausible, which is easy to validate through comparison with a manually provided alignment of the pieces. Additionally to the factors of the single-object reassembly case, the mixed object reassembly is tested for its discrimination capability. In contrast to the former, this test should result to clusters of fragments and potentially isolated fragments as well.

Input. The set of fragments to be used for the reassembly.

Expected Result. A ranked list of proposed reassembly solutions (could also be empty or with a single solution, depending on the quality requirements), each one given as a set of object transformations (per fragment) and a set of "links" between fragments.

Relevant Tools. The quality of the results can be a) measured (total penetration, total fragment distance, graph integrity) and b) visually inspected using one of the following: the resulting transformed objects when loaded into MeshLab or any other mesh viewer, the virtual workspace of T5.1 (only available at the final stages of the integration).

4.3.2.4. Reassembly of Heavily Damaged Fragments

In the presence of surface degradation and damage, the fragments may lose many of their distinctive features on the fractured surfaces (smoothed) or their contact surface may become significantly reduced. This test (fourth in terms of difficulty) is designed to evaluate a) the limits of the fracture-based reassembly algorithms and b) alternative methods proposed as part of T4.1 for addressing these hard cases, which are frequent in real-world situations, nevertheless. The present test as well as the next (disjoint fragments) are the ones that actually demonstrate the novel aspects of the new approaches taken in PRESIOUS for the reassembly topic. In the presence of severe deterioration, the matching algorithms developed so far (background work) fail. The limits of our new, improved fracture-based matching methods are examined and we employ shape-guided matching techniques to detect matches where the contact surface between two fragments becomes unreliable.

This experiment maps to all evaluation scenaria of objective O3.

Evaluation Procedure. The user initializes a reassembly set with fragments that have been previously inspected and are found not to match perfectly, due to their eroded and/or damaged (broken) state. Small pieces of material between them may be entirely missing, forming gaps and reducing their contact surface. The criteria and factors are similar to the other multi-part experiments.

Input. The set of fragments to be used for the reassembly.

Expected Result. A ranked list of proposed reassembly solutions, each one given as a set of object transformations (per fragment) and a set of "links" between fragments. The ranked list may be empty or contain a single graph, depending on the strictness of the quality settings.

Relevant Tools. The results can be practically evaluated only visually, since the gaps formed between the pieces do not allow for objective measurements. Later on, we may propose particular metrics that could be used, according to the specific algorithms designed for this task. Visual inspection will be enabled via a third-party software such as MeshLab or via the user interface of the virtual workspace of T5.1 (only available at the final stages of the integration).

4.3.3. Object Completion Stage

(Stage 3, Task 4.2)

The Object Completion Stage (Stage 3) directly follows the Reassembly Stage (Stage 2) and corresponds to T4.2. Given a partially reassembled object, access to external shape repositories and user input, a completed object and a set of missing parts is computed. Figure 8 provides an overview of the current, preliminary design of the object repair pipeline that is structured into eight processing steps. Table 1 summarizes the application of evaluation criteria to the individual processing steps. For each step a more detailed description along with is provided in the following subsections.

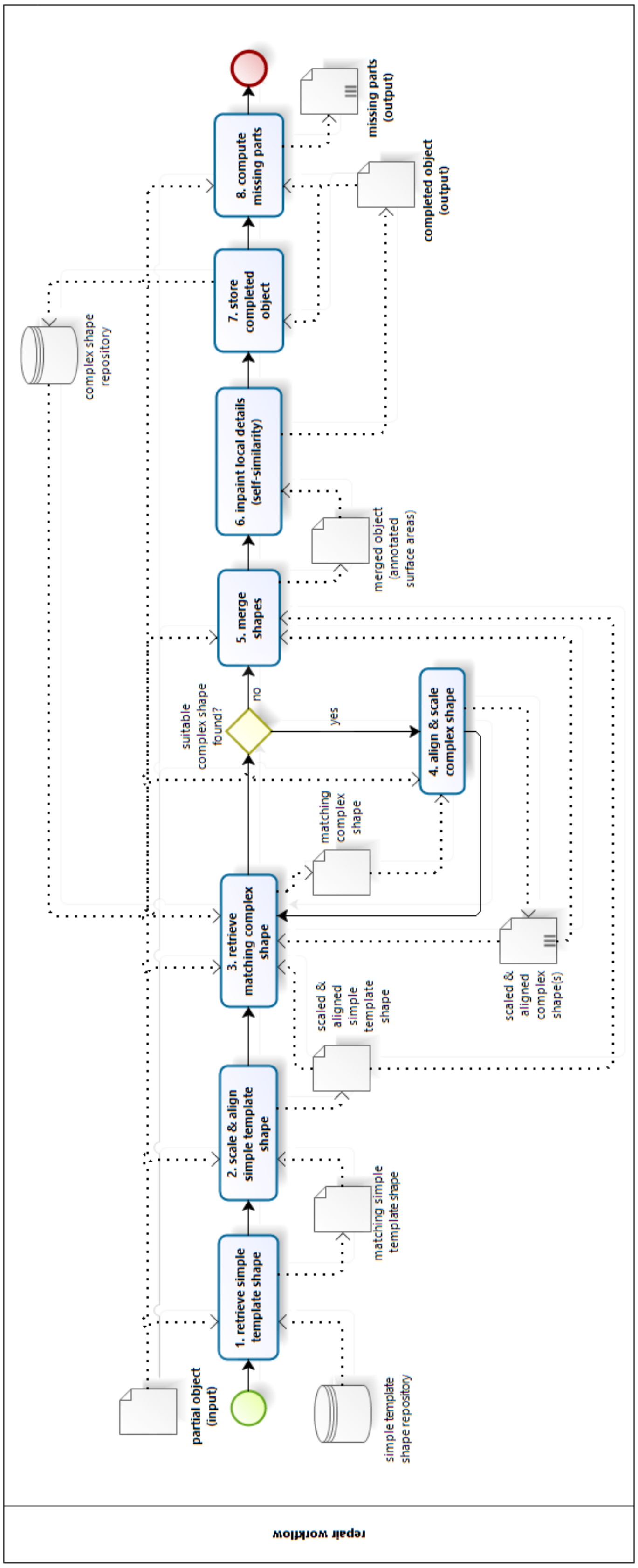

Figure 8. Preliminary design of the Stage 3 pipeline at M12.

	Simple Template Retrieval	Scale & Align Simple Template	Complex Shape Retrieval	Scale & Align Complex Shape	Merge Shapes	Inpaint Local Details	Store Completed Shape	Missing Parts Export
General Effectiveness	X	X	X	X	X	X	X	X
General Efficiency	X	X	X	X	X	X	X	X
Robustness / Missing Parts	X	X	X	X		X		
Robustness / Weathering	X	X	X	X		X		
Robustness / Mesh Defects	noise	noise	noise	noise				
Mesh Defect Introduction					noise + non-manifold	noise + non-manifold		noise + non-manifold
Scale / Resolution Invariance	X	X	X	X		X		
Precision / Recall	X		X					
User Input Requirements	X	X	X	X		X		X
Generality / Applicability	X	X	X	X		X		X

Table 4.3.2. Application of evaluation criteria (rows) to object repair processing steps (columns)

4.3.3.1. Primitive Shape Retrieval

Given the partial object, that is provided by the preceding stage, the Simple Shape Retrieval step queries a repository of primitive shapes with the partial object in order to retrieve a simple shape (Figure 9) that is on a global level similar to the overall shape of the completed object. This step assumes that completed objects are, in their overall shape, similar to at least one of the primitive shapes contained in the repository. Furthermore, it relies on a heuristic classification of the partial objects fragments surface into exterior and contact surface that is established in the preceding stage (Stage 2). It has not yet been decided if this classification should at some point be customizable by the user. The simple shape retrieval might internally modify the partial object input in arbitrary ways to perform multiple queries and subsequently compare their result rankings to settle on a template shape. It is also possible that a list of template candidates is presented to the user which then provides the final choice.

For evaluation, precision and recall will be determined for a specified set of partial objects and a simple template shape repository. Scale and Resolution independence is crucial for this step since partial objects will have a vastly higher resolution than template shapes. Yet the retrieval has to establish similarity measures that are largely invariant to difference in resolution to compare such objects. In addition this step is evaluated for scale / resolution invariance as well as robustness against missing parts, degradation and mesh defects in the form of noise.

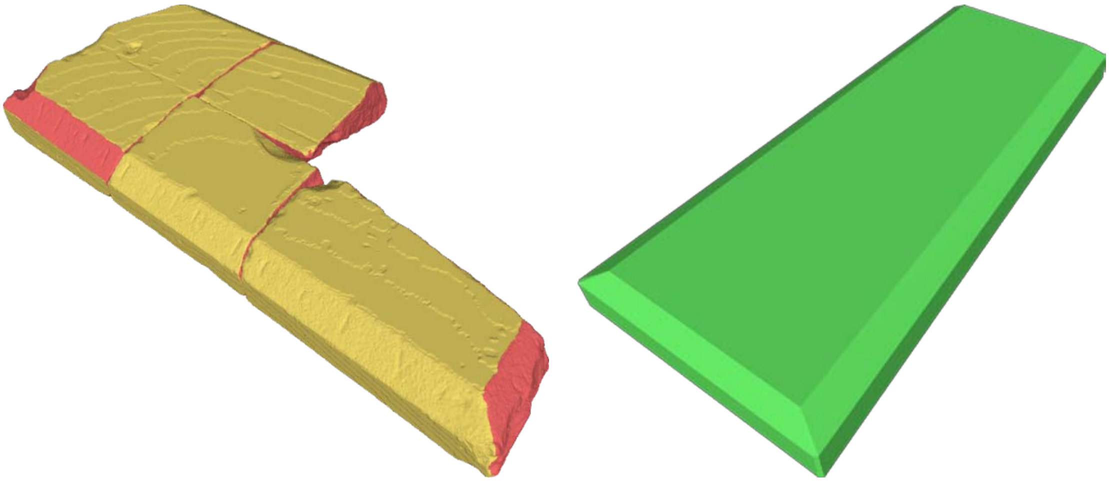

Figure 9. Left: Mock-up of a Partial object classified into exterior and fractured areas highlighted in yellow and red. Right: A matching simple template shape for the partially reassembled tombstone.

4.3.3.2. Primitive Shape Alignment and Scaling

After a simple template shape has been retrieved, its scale and rotation in coordinate space is very unlikely to match those of the partial object. Hence the Simple Shape Alignment and Scaling step is required to compute a transformation for the template shape to adjust its size and orientation. Figure 10 illustrates an alignment for the partial object and template shape shown in the previous subsection.

For evaluation, the actually computed transformation of the algorithm will be compared to an externally (manually) determined optimal scaling and alignment. Furthermore, given an algorithm based on an iterative approach, its convergence behaviour will be examined as well. Similar to the evaluation of the previous step, additional fragments will be omitted from the partial object and small scale defects could be artificially added by using the decay prediction components from WP3. Further evaluation criteria will be scale / resolution invariance, user input requirements and robustness to noise.

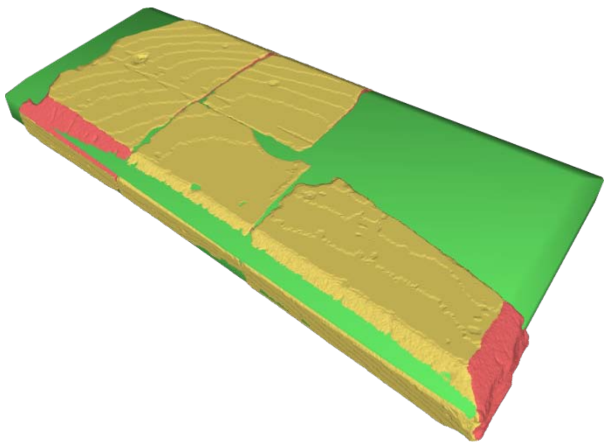

Figure 10. A mock-up of the aligned partial object (yellow / red) and the previously retrieved simple template shape (green).

4.3.3.3. Complex Shape Retrieval

In the Complex Shape Retrieval step, the aligned template shape and the partial object are used to retrieve more complex shapes such as digitized fragments of CH objects (Figure 11). The retrieval will respect local characteristics of the exterior surface of the partial object as well as the fractions of the aligned template volume that do not intersect with the partial object. As with the simple template shape retrieval, it is expected that the various input shapes are internally modified during the processing step and that multiple queries will be performed and used to establish an overall ranking of the results. Again user input might be gathered to select a suitable shape out of a candidate list.

As shown in the pipeline diagram (Figure 8), the Complex Shape Retrieval is part of a control flow loop that also includes the Scaling and Alignment of Complex Shapes.

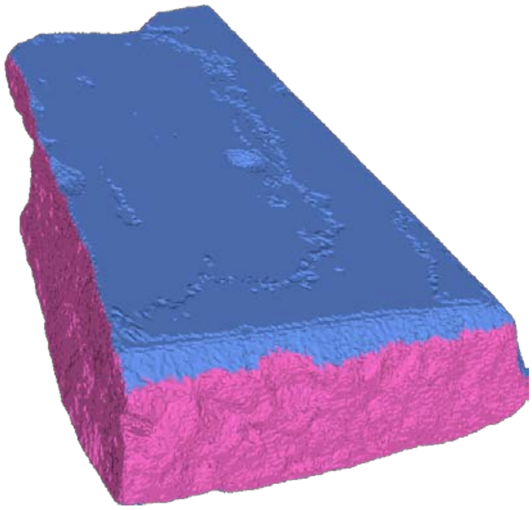

Figure 11. A mock-up of a retrieved complex shape, note that the surface is heuristically classified into possible exterior (blue) and contact surface (purple)

The complex shape retrieval step is executed several times until no further matching complex objects can be retrieved. However it has to be determined how the case should be handled if the remaining template volume is separated into several parts by the subtraction of previously aligned shapes. This will probably lead to user input requirements, where a user has to select or further edit one of the remaining sub-volumes.

For evaluation, precision and recall will be determined using a specified list of input shapes consisting of aligned template, partial object pairs that are combined with one or more complex shape (i.e. fragments). Invariance to noise and differences to resolution and scale are important as well. Similar to the Simple Template Shape Retrieval, robustness to missing large parts and degradation will be evaluated.

4.3.3.4. Scaling and Alignment of Complex Shapes

After a matching complex shape has been retrieved in the previous step, it has to be fitted to a part of the volume of the template shape that does not intersect with the partial object or previously aligned complex shapes (Figure 12).

As with the complex shape retrieval, it has not been determined yet how this step should handle cases in which the template volume is separated into several volumes by extruding previously aligned shapes. Analogical to the scaling, it is expected, that some form of user input will be required to address this problem.

The alignment approach is basically similar to the alignment of partial object and simple template shape. This assumes that the surface of the complex shape can be heuristically classified into external and contact surface (i.e. the breaking edges).

For evaluation, the same criteria as with the template shape scaling and alignment will be applied. The quality of computed solutions will be determined against manually created solutions.

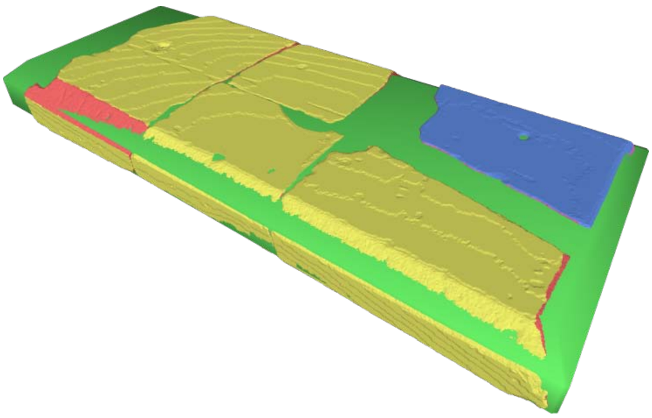

Figure 12. A mock-up of the aligned complex shape (blue, purple), partial object (yellow, red) and simple template (green)

4.3.3.5. Shape Merging

In this step, all previously aligned shapes are merged into a single shape.

Evaluation will focus on determining whether the result is a valid, piecewise linear 2-manifold[4]. In addition the result will be evaluated for introduction of other non-critical defects of the mesh structure such as changes in genus, noise as well as deviation of the total volume of the resulting shape against the union of the volumes of all aligned input shapes by comparing this output to an externally computed reference shape.

4.3.3.6. Local Detail Inpainting

After aligned shapes have been merged and annotated, surface areas stemming from the simple template shape as well as contact surfaces (i.e. fracture edges) will be inpainted using surrounding areas that belong to the external surface of the partial object as well as the external surface of other aligned complex shapes. Figure 13 showcases a completed object after inpainting. If, however, the areas that are automatically selected for inpainting are too large or if the source area features highly convoluted surface patches, a fully automated approach might not be able to achieve adequate results. Hence user input that manually determines source and target areas might be required. The actual inpainting approach will likely be based on one of the iterative approaches referenced in part A of D4.1.

Evaluation will focus on the introduction of mesh defects, especially self-intersection, singularities, and water-tightness.

4 This includes tests for singularities, self-intersections, boundaries and water-tightness

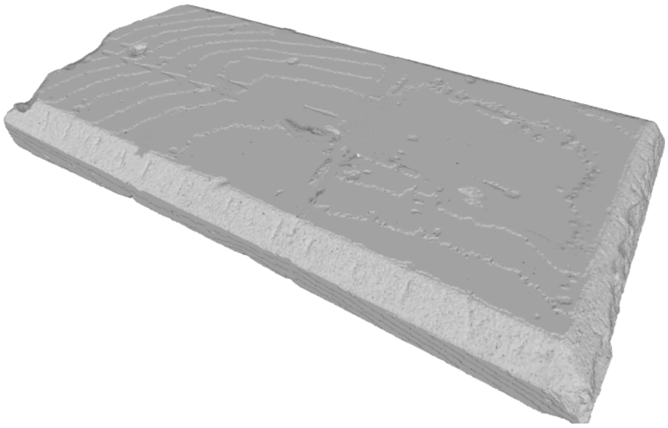

Figure 13. A mockup of the tombstone after merging and inpainting.

4.3.3.7. Completed Shape Storage

The output of the inpainting step is considered to be a completed object. This completed object will be stored in the complex shape repository. It has to be determined whether completed objects will be artificially decomposed into fragments again to provide a better basis for reuse as complex shapes in the completion of other partial objects. However it is expected that if such a decomposition will be used, it will be conducted externally and not within the object completion pipeline.

Technically this step does not change the mesh data, hence its evaluation will be limited to simple effectiveness and efficiency measures for storing a completed object.

4.3.3.8. Missing Parts Computation and Export

As stated above, the completed object has already been computed in previous steps. The missing parts are essentially considered to resemble the difference of the completed object and the partial object that was provided as an input at the beginning of the object repair pipeline. Similar to the shape merging step, computation will most likely be based on a volumetric representation of the completed and partial object shapes. In most cases, the extrusion will lead to several disconnected volumes. For shape export, each of the resulting volumes will be converted to a mesh. The export step will not consider specific hardware limitations of 3D printers or milling machines such as limited spatial resolution and limits on the minimum shape diameter that must be adhered to in order to obtain stable parts.

Evaluation will focus on the introduction of mesh-defects, with emphasis on singularities and water tightness. In addition the extrusion of the completed object by the partial object input will be computed externally to measure the output for introduction of noise and deviations in the volume of the parts.

4.4. Experiments for WP5 (Integrated Systems)

This section covers the evaluation of the Subsystems delivered as D5.3 and D5.7. The evaluation of the integrated systems has a different focus compared to that of the work packages targeting the scientific objectives of the project. It must:

a) Ensure the operational state of the delivered systems

b) Validate that the envisioned scenaria using the two platforms can be implemented

c) Demonstrate that the final tools and underlying workflows involved make a significant contribution to the objectives of the project, indicating a clear benefit from using them.

With the above in mind, the following general criteria must be met for the two platforms, which are further specialized in the following sections. The evaluation experiments are directly mapped to the envisioned scenaria described in detail in Section 3.

4.4.1. General Evaluation Criteria

Effectiveness

In simple terms, this boils down to answering the questions: "why should somebody use the solution provided by the platform" to perform the designated tasks and, in consequence, "how the currently established workflow is improved or otherwise affected by the use of the platform". The effectiveness must be clearly indicated by measuring performance, measuring potential improvements wrt to previous results by similar procedures, indicating new results that could not be previously achieved etc.

Usability

The usability of a system refers to the ease of use, completeness of functionality, interface and available options, the clarity of mapping the most common workflows and tasks to comprehensive functions etc. In some cases, where performance measurements also include and imply the user interaction with the application, the usability also affects the effectiveness of the system.

Responsiveness

For interactive tasks, this is a sub-criterion of the usability, requiring the timely or progressive feedback of an operation to avoid stalling the system. Lengthy processes should not withhold the entire system and have to progress in the background, so that other tasks can be completed simultaneously.

Automation level

For some tasks involved in the workflows of the two systems, it is important to maintain a high level of automated operation, without requiring constant input from the user. Failure to meet this, may also result in a poor effectiveness. Some tasks however, may require additional input and this is not objectionable as long as it corresponds to choices requiring semantic knowledge that cannot be deduced from the data being processed.

Completeness

The proposed systems must at least map to the scenaria proposed in this report, delivering all the functionality required to fulfill the described tasks, also described in the Description of Work.

Generality

This is the ability of the platforms to be configured properly to work for diverse scenaria and/or other application domains. This is important in order to prove that the designed methods, workflows and systems are worth exploiting further, and is also a link to the exploitation plan.

Robustness

Since deliverables of WP5 that are being evaluated are software platforms, these have to be stable and provide repeatable results; where random processes are involved, ‘repeatability’ will refer to similarity according to some statistical measure. This criterion should not be confused with commercial software stability, since it refers to prototypes, which are not production-grade pieces of software.

Plausibility

All tasks involved in the Virtual Repair and Measurement Workspace have no easily quantifiable indicators with respect to the quality and usefulness of the end result. Validation by an expert may be required in all cases, via visual inspection.

4.4.2. Virtual Repair and Measurement Workspace Experiments

The experiments for testing the Virtual Repair and Measurement Workspace (VRMW) are driven by the scenaria 3.2.1, 3.2.2, 3.2.3, 3.3.1, 3.3.2 and 3.3.3. Additionally, since there is a common core of collection management tools that transcends the VRMW scenaria, for 3D object manipulation, management and visualization, the evaluation guidelines and experiments for these are presented separately, first.

	Plausibility	Robustness	Generality	Completeness	Automation level	Responsiveness	Usability	Effectiveness
4.4.2.1 Collection Management Tools		●		●		●	●	
4.4.2.2 Object Erosion: Surface Geometry Only	●	●	●	●	●		●	●
4.4.2.3 Object Erosion: Surface Geometry and surface physicochemical data	●	●	●	●	●		●	●
4.4.2.4 Object Erosion: Surface Geometry and volume physicochemical data	●	●	●	●	●		●	●
4.4.2.5 Shape-guided reassembly	●	●	●	●	●	●	●	●
4.4.2.6 Object Repair	●	●	●	●	●	●	●	●
4.4.2.7 Missing parts prediction		●	●	●	●	●	●	●

Table 4.4.1. Mapping of general evaluation criteria to the VRMW experiments.

4.4.2.1. Collection Management Tools

The fragments of objects to be reassembled or entire objects for which erosion is to be simulated may be organized in collections or they may be loaded as an unsorted and unified collection into the virtual workspace. Objects are in general loaded as collections, appended to those already opened in the VRMW, or imported individually. It is imperative that the user interface provides all tools necessary for searching, filtering, organizing (e.g. group-selecting by collection), exploring, and manipulating the individual pieces similar to the way they would be laid out on a physical workspace. The same holds for the results of the procedures hosted in the VRMW, i.e the eroded versions of given objects, part assemblies and repaired models.

The layout of the application must allow for both free-style experimentation / inspection and tabular/hierarchical entity management (two views). Any information conveyed by the objects must be also accessible (annotations). If the objects are pre-classified by some criterion, there should be a functionality to select similar objects or search by keyword the entire set.

The system has to be responsive enough to allow for interactive control of the environment and the displayed models. Additionally, according to the DoW, the objects must be presented to the user in the best representation possible, according to the recorded data (e.g. textures, detail, realistic materials etc), while achieving interactive rendering times where the object size allows this. This means that the system must provide all necessary mechanisms for graceful quality degradation, if the graphics subsystem cannot handle the amount of data simultaneously shown. The (3D) virtual workspace should provide all the necessary cues for intuitive spatial positioning of the assemblies and the collections in general, including measurement grids, shadows (for quick visual spacing distance-to-reference plane measurements) as well as adequate lighting for visual feature discrimination and general inspection.

Standard rigid transformation functionality (e.g. gizmos, gestures and/or shortcuts) must be available for the easy manipulation of the objects, in order to be able to "virtually test" them and compare them against others. Users could also benefit from alignment aids and snapping tools.

For the inspection of the collection, there must be a set of controls available to reposition the view (focus) on a selected item or group of objects and be able to rotate the camera around the interest point (also modifiable).

According to the DoW and in order to facilitate the comparison of the scale of various object features, measurement tools should be available in the workspace, either as user-manipulated tools or as automatic dimensioning aids.

4.4.2.2. Object Erosion: Surface Geometry Only

The erosion scenaria are described in section 3.2 and their internal evaluation in section 4.2. Here, we describe the evaluation of the erosion module as part of the integrated VRMW. According to the Description of Work, the first version of the integrated system (M24) should concentrate on providing the GUI of the VRMW to support the erosion procedures, a working set of import/export tools, connectivity with the data sources, collection management tools, logging, reporting and basic object manipulation in the three-dimensional environment. Primary criteria for this first version of the reassembly mechanisms should be the usability and robustness.

For the final platform version, all functionality detailed in the DoW should be present and the erosion module will be evaluated using its full functionality. With respect to the criteria mentioned in section 4.4.1, the following details apply here:

Effectiveness: Short of waiting for the duration of a physical erosion process (typically in the tens or hundreds of years), the physical alternative to the erosion simulator is the accelerated aging chamber. The effectiveness of the erosion simulator will thus be measured by comparing the time, workload and cost of using the erosion simulator (including 3D scanning time) against using physical or accelerated aging. Note that physical accelerated aging cannot be applied to actual Cultural Heritage fragments and this should be taken into account in assessing the effectiveness of the erosion simulation.

Usability: The erosion simulator should not require heavy interaction with the user. After selecting the object to be eroded, its physicochemical parameters and the environment parameters (e.g. average annual rainfall), the offline task of the erosion simulator should commence. Upon notification of completion the user should be able to visualize the eroded object, to perform measurements on it and to store it appropriately. The ease and intuitiveness involved in using the erosion simulator will be rated.

Responsiveness: This will not be evaluated as the erosion simulator is not expected to offer real-time response. A progress indicator will be provided.

Automation level: Once the object to be eroded has been selected and the erosion parameters have been input (object physicochemical and environment parameters), the erosion simulator should be totally automatic. This is particularly important since the erosion simulation may require considerable time, and user interaction cannot be expected.

Completeness: The data input functionality to allow object selection and erosion parameter specification should be complete to fulfill the scenario. The data output functionality to allow visualization, measurement and storage of the eroded object should also be present.

Generality: The erosion simulator should be applicable to any properly digitized object (of specific stone types) as well as to a range of physicochemical and environment parameters. The simulator will at least be applicable to two specific types of stone, those mainly found at the Nidaros and Elefsis sites.

Robustness: Given the same initial object and erosion parameters, the erosion simulator should produce a statistically similar eroded result. Since random processes will likely be involved, the results are expected to vary within a statistical range.

Plausibility: The system will be evaluated by both an expert user (from the development team) and a field specialist. The former will provide an assessment of the technical side of the result (e.g. quality of resulting mesh, quality of visualization) while the latter will be an archaeologist who will provide an additional assessment of the results of the erosion simulator.

4.4.2.3. Object Erosion: Surface Geometry and surface physicochemical data

Same as 4.4.2.2.

4.4.2.4. Object Erosion: Surface Geometry and volume physicochemical data

Same as 4.4.2.2.

4.4.2.5. Shape-guided reassembly

The experiment procedure description for the reassembly is detailed in the respective scenario (section 3.3.1). According to the Description of Work, the first version of the integrated system (M24) should concentrate on providing the GUI of the VRMW to support the reassembly activities, a working set of import/export tools, connectivity with the data sources, collection management tools, logging, reporting and basic object manipulation in the three-dimensional environment. Primary criteria for this first version of the reassembly mechanisms should be the usability, responsiveness and robustness.

For the final platform version, all functionality detailed in the DoW should be present and the reassembly pipeline will be evaluated using its full functionality. With respect to the criteria mentioned in section 4.4.1, the following details apply here:

Effectiveness: Can the automatic reassembly improve the workflow of an archaeologist, conservator, anthropologist or any other expert related to the fragmented parts? In particular, the system should improve the ability to locate matches among fragments, accelerate the procedure, help identify clusters of fragments for large pieces that would be physically difficult to compare, and perform accurate positioning. For loosely coupled fragments, it is important to also evaluate the plausibility of the provided solution(s), with respect to the specialist's interpretation of the assembly. Since the shape-guided aspects of the reassembly are novel and we also take a new approach to the entire shape matching paradigm, a concrete measure of the improvements realized is the publication of papers regarding the reassembly system as a whole.

Usability: The reassembly procedure should not require heavy interaction with the user, since that would impair the system's contribution (see also automation level). Therefore, the pre-assembly operations for hard constraint placement, collection organization and reassembly initiation should be straightforward. If multiple options are presented to the user at some stage of the reassembly, requiring feedback, this should be a very intuitive procedure, with an easy-to-use optional rotation mechanism and immediate visual feedback. Visual aids should be also provided in the case where external features and additional shapes are used for guiding the reassembly. Special highlighting of the guide object and/or the tracked features should be provided to help the user assert the validity of the currently selected option.

Responsiveness: The user interacts with the 3D objects in the VRMW to set up a reassembly session via a mouse/gesture driven interface. This means that all actions performed by the user at this stage should have an immediate feedback. For the computational procedures of the reassembly, these should be initialized in the background to allow the user to explore the workspace or operate on other parts simultaneously.

Automation level: The reassembly should be as automated as possible. Typically, and for small collections, there should be no interaction at all. Additional input in the form of a choice indication or a hard constraining of two parts is not objectionable since it typically corresponds to semantic knowledge of the specialist about the fragments.

Completeness: Obviously, the VRMW should provide all widgets and options to perform all tasks described in the scenario.

Generality: It is paramount to show that the solution is not content-specific. To this end we will test the system with additional third-party collections and datasets obtained from public repositories.

Robustness: Following the same constraints and procedures in the reassembly initialization and the subsequent choices during the reassembly session, the system should provide a repeatable result.

Plausibility: The system will be evaluated by both an expert user (from the development team) and a field specialist. The latter will be an archaeologist / architect (our test repository mainly consists of architectural parts, for now), who will provide an additional assessment of the reassembly workflow, ease of use and plausibility of the results.

4.4.2.6. Object repair

The object repair scenario is described in section 3.3.2 and their internal evaluation is part of section 4.3.1 and 4.3.3. Here, we describe the evaluation of the object repair module as part of the integrated VRMW. According to the Description of Work, the first version of the integrated system (M24) should concentrate on providing the GUI of the VRMW to support the repair procedures, a working set of import/export tools, connectivity with the data sources, collection management tools, logging, reporting and basic object manipulation in the three-dimensional environment. Primary criteria for this first version of the reassembly mechanisms should be the usability and robustness.

For the final platform version, all functionality detailed in the DoW should be present and the repair module will be evaluated using its full functionality. With respect to the criteria mentioned in section 4.4.1, the following details apply here:

Effectiveness: Can the repair module improve the workflow of an archaeologist, conservator, anthropologist or any other expert related to a partially reassembled model? In particular the subsystem should improve the ability to combine a digitized and aligned set of fragments into a single mesh representation of the object. For the repair scenario this also includes the repair of physical defects[5] that are located on exterior surface patches of the fragments. According to the perception of the user the resulting model should consist of a single piece and should feature less defects than the

[5] see D4.1 for detailed defect specification

input fragments. Note that for assessing effectiveness of the object repair it has to be taken into account, that it might not be meaningful or even possible to alter the physical object at all as this could e.g. be counter-productive for archival or further research on the physical CH object. In such cases the digitally repaired model could also provide a basis for more efficient production of physical object that resemble the restored object.

Usability: Generally, the repair scenario should aim for a high level of automation to effectively increase the systems contribution and value to end users. However it might not in all cases be possible to achieve a completely automated workflow. E.g. the repair subsystem will not be able to assess semantic details that are known to the user, hence the system should to a certain extent feature facilities for user based selection of template shape candidates or selection of defective surface areas. Also due to the computational complexity and stability of certain alignments in the object repair stage, usability of the system might be improved by allowing the user to manually provide additional constraints for alignment computation.

Responsiveness: The user interacts with the 3D objects in the VRMW to set up a repair session via a mouse/gesture driven interface. This means that all actions performed by the user at this stage should have an immediate feedback. For the computational procedures of the reassembly, these should be initialized in the background to allow further interact with UI during computationally expensive operations.

Automation level: As already indicated for the Usability criteria, the repair should be as automated as possible. Typically, for a partial object with only few large parts missing and a sufficient amount of exterior surface that is not affected too heavily by small-scale surface degradation, the amount of required user input should be limited to just a few user interactions with very limited complexity.

Completeness: As for the other subsystems, the VRMW should provide all necessary widgets and options to perform the tasks described in the repair scenario.

Generality: It is important to determine and clearly state to which types of objects the repair subsystem can be applied to. On the one hand this can be assessed through a theoretical analysis of the entire repair pipeline. On the other hand we will try obtain suitable third party data for additional practical testing of the repair subsystem.

Robustness: Given the same data and user input as well as an identical shape repository, the system should provide the same results across several executions. In addition, the subsystem should be as invariant as possible against shifts in scale, small-scale defects, digitization noise and even missing large parts in the input.

Plausibility: Besides requiring formally specified input and output criteria, the effectiveness criteria and measurement does not capture the subjective quality of the individual repair results to full extent. As concluded in D4.1 repairing digitized CH Objects is an ill-posed problem due to incomplete information on the missing parts. In turn, establishing a generic and formal definition of the plausibility of an object reassembly or completion can be regarded as an ill-posed problem as well. Thus, we will rely on human provided ground truth for assessing plausibility of repair results.

In [[6]] user studies were conducted, where a group of voluntary users directly rated the plausibility of generated models versus initial models in a double-blind test. In context of the object repair subsystem, a similar study could be conducted in which a set of non-fragmented models and a set of repair results are presented to CH domain experts. For each model in the sets the CH domain expert has to provide a rating whether the objected presented on screen is a result of the repair subsystem or just a digitized model of a physical object without significant defects. By this approach, the evaluation could incorporate implicit domain knowledge and intuitive classifications made by domain experts during the visual inspection. In an ideal case, the ratio of true and false classified models should be the

6 Chaudhuri et al. *Probabilistic reasoning for assembly-based 3D modeling.* Proceedings SIGGRAPH '11 ACM SIGGRAPH 2011 papers Article No. 35

same for both sets. However to yield accurate results a large number of fragmented models as well as a set of similar models without highly visible degradation and missing parts is required.

4.4.2.7. Missing parts prediction

According to the scenario description provided in 3.3.3, the missing parts prediction heavily relies on the preceding object repair and reassembly scenarios. The additional components that will be implemented for the missing parts prediction will essentially compute the difference of the repaired model (i.e. the result provided by the object repair subsystem) and the partial object (i.e. the result provided by the reassembly subsystem). From the perspective of a user of the system, evaluation criteria and methodology will generally be almost equivalent to the evaluation of the repair subsystem. However evaluation of Plausibility will be omitted for the missing parts prediction. Although it can be regarded as tightly related to the plausibility of the object repair results, there is no way to assess plausibility in an experimental way, that is either trivial or would mainly cover aspects that are not within the scope of PRESIOUS.

4.4.3. Predictive Digitization

The experiments for testing the Predictive Scanning platform are driven by the scenario described in Section 3.1.

Effectiveness: The predictive scanning platform should speed-up the scanning process, limiting as much as possible the time required to obtain an acceptable, plausible result. In the case of the pottery database, the latter will be quantitatively evaluated by comparing the result with the ground truth described in Section 3.1. An indicator of the novelty of the predictive scanning as a whole will be the publication of papers.

Usability: The predictive scanning platform should feature a GUI facilitating user interaction during the scanning process, as well as during retrieval and reshaping. In the latter case, user interactions will be limited to simple, corrective tasks. A domain expert will assess the overall ease of use of the platform.

Responsiveness: Intermediate predicted scans should be available in a timely, incremental fashion, as the scanning process progresses. The user should be able to evaluate the plausibility of the result obtained by each stage of predictive scanning and decide whether additional scans are required.

Automation level: Predictive scanning involves human intervention in the scanning part. On the other hand, the retrieval and reshaping parts should be as automated as possible.

Generality: Apart from the pottery dataset, publicly available benchmark datasets will be used, in order to demonstrate that the predictive digitization platform is not content-specific.

Robustness: The performance of predictive scanning, as indicated by the respective quantitative measures, should not suffer from statistically significant decrease in cases of alterations in the scanning sequence or in the exact position of the scanning device.