Ref. Ares(2020)3382473 - 29/06/2020

SSHOC
social sciences & humanities open cloud

Research and Innovation Action

Social Sciences & Humanities Open Cloud

Project Number: 823782 Start Date of Project: 01/01/2019 Duration: 40 months

Deliverable 5.17 Implementation plan for the archeological case study

Dissemination Level	PU
Due Date of Deliverable	30/06/20 (M18)
Actual Submission Date	29/06/20
Work Package	WP5 Innovations in Data Access
Task	T5.7 Open Linked Data. Archaeology Case Study
Type	Report
Approval Status	Approved by EC - 03 November 2020
Version	V1.0
Number of Pages	p.1 – p.37

Abstract:

A virtual reconstruction of the Roman theatre in Catania will be created as an example of an actual transition of archaeological data to the cloud, i.e. from data silos on individual computers to webservices. The case study is based on a unified workflow that starts with the archaeological documentation and results in a virtual reconstruction. This deliverable describes the necessary work steps, the planned collaborations and the expected outcomes of this task.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability. This deliverable is licensed under a Creative Commons Attribution 4.0 International License.

SSHOC, "Social Sciences and Humanities Open Cloud", has received funding from the EU Horizon 2020 Research and Innovation Programme (2014-2020); H2020-INFRAEOSC-04-2018, under the agreement No. 823782

History

Version	Date	Reason	Revised by
0.1	29/04/2020	Initial version	All authors
0.2	05/05/2020	Peer review	Daan Broeder
0.3	22/06/2020	Address peer review comments	Wolfgang Schmidle
0.4	24/06/2020	Apply SSHOC guidelines	Wolfgang Schmidle
0.5	25/06/2020	Harmonise image captions	Wolfgang Schmidle
0.6	25/06/2020	Add list of acronyms and abbreviations	Wolfgang Schmidle
1.0	26/06/2020	Adressed Coordinator edit comments	Wolfgang Schmidle

Author List

Organisation	Name	Contact Information
DAI	Wolfgang Schmidle	wolfgang.schmidle@dainst.de
CNR (ITABC)	Sofia Pescarin	sofia.pescarin@itabc.cnr.it
CNR (ITABC)	Emanuel Demetrescu	emanuel.demetrescu@cnr.it
CNR (ITABC)	Bruno Fanini	bruno.fanini@ispc.cnr.it
CNR (IBAM)	Francesco Gabellone	francesco.gabellone@cnr.it
CNR (IBAM)	Alberto Bucciero	alberto.bucciero@cnr.it
CNR (IBAM)	Alessandra Chirivì	alessandra.chirivi@ispc.cnr.it
CNR (IBAM)	Ivan Ferrari	ivan.ferrari@cnr.it
CNR (IBAM)	Francesco Giuri	francesco.giuri@cnr.it

Executive Summary

In SSHOC Task 5.7 (Open Linked Data. Archaeology Case Study), a virtual reconstruction of the Roman theatre in Catania will be created as an example of an actual transition of archaeological data to the cloud, i.e. from data silos on individual computers to webservices. The case study is based on a unified workflow that starts with the archaeological documentation and results in a virtual reconstruction. With this workflow, data manually acquired during an excavation and traditionally stored on paper can now be stored in the cloud and used for 3D visualisations of the site.

The workflow uses tools that are being developed by the task partners, such as idai.field for exacavation documentation and the Extended Matrix for 3D reconstruction. The task partners will document the existing systems with their individual workflows and apply them to existing as well as newly created data on the Roman theatre. They will then work on combining these tools with the aim of creating an overarching cloud-based workflow.

The case study will rely on systems for normative data such as gazetteers for place and time information and based on the case study the task partners will evaluate additional data standardization strategies to enable data sharing and re-use of archaeological data. To this end, the underlying ontologies will be aligned with CIDOC CRM and provide the data as LOD wherever possible.

The present Deliverable 5.17 is the implementation plan for this case study. It details the necessary work steps, the planned collaborations and the expected outcomes of this task.

Abbreviations and Acronyms

API	Application programming interface, used by computer programs to communicate with each other
CIDOC CRM	CIDOC Conceptual Reference Model, a high-level ontology for the humanities, with domain-specific extensions such as CRMarcheo for archaeological excavations
CLARIN	European Research Infrastructure for Language Resources and Technology
EM	Extended Matrix, developed by ISPC Rome
GraphML	Graph Markup Language, an XML-based data format for storing semantic graphs
LOD	Linked Open Data (aka Linked Data, Open Linked Data)
norm data	short for normative data, e.g. controlled vocabularies
RM	representation model, contains 3D modelling details
SFM	Structure from Motion
STV	spacetime volume
UAV	Unmanned Aerial Vehicle (aka drone)
XML	Extensible Markup Language, used for saving structured data in plain text files
yED	an editor for drawing diagrams

Table of Contents

1. Introduction

Archaeological field work is a complex process in which many scientific resources and competences are allocated for long periods of time, sometimes even for several years. However, the data is often either not available in digital form or not public. For this reason, the sustainability of archaeological research would benefit immensely from optimized and effective systems of digital documentation, analysis and visualisation of archaeological sites.

The SSHOC Task 5.7 (Open Linked Data. Archaeology Case Study) presents an example of an actual transition of archaeological excavation data to the cloud. A workflow from archaeological documentation to virtual reconstruction is described, i.e. how data manually acquired during an excavation and traditionally stored on paper can be stored in the cloud by every archaeologist working on the same site and subsequently used for 3D visualisations of the site. The tools for this workflow, such as *idai.field* for exacavation documentation and the Extended Matrix for 3D reconstruction, are being developed by the Task 5.7 partners and are described in detail in section 2. The task's objective is to combine these tools and develop them further, make the tools easier to use and offer a cloud perspective for sharing and re-using data.

One focus will be on data transparency. The graphic representation of reconstructive hypotheses suffers from the so-called Black box effect, i.e. a user's cognitive difficulty to appreciate which parts of a reconstructive hypothesis are based on concrete data (archaeological remains) and which on less reliable sources, and which parts have been added with a merely evocative purpose. Solving this cognitive semantic ambiguity of construction models can drastically increase the communicative and educational strength of virtual reconstructions for Cultural Heritage.

The task partners are the German Archaeological Institute (DAI) in Berlin and two institutes of the Istituto di Scienze del Patrimonio Culturale (CNR-ISPC) in Rome and in Lecce. They will be referred in this document as DAI Berlin, ISPC Rome and ISPC Lecce.[1]

Task 5.7 brings together the workflow of ISPC Rome for generating 3D visualizations from excavation data, the excavation documentation system and the expertise in using normative data (*norm data* for short) of the DAI Berlin and the excavation data and archaeological expertise of the ISPC Lecce. This will be described in detail in section 2. A sketch of the work to be undertaken in this task is given in section 3. This includes as many connections to other SSHOC tasks as possible because, being a case study, task 5.7 intends to make use of the findings of other tasks wherever possible. The concrete work steps are detailed in section 4.

[1] ISPC Rome (CNR-ITABC) and ISPC Lecce (CNR-IBAM) were merged with two other CNR institutes to form the CNR-ISPC Institute of Heritage Science in October 2019 and are both listed simply as CNR in the SSHOC Grant Agreement.

The initial plan was to identify an archaeological community and to create a software module to help them take their existing data to the cloud, as evident in the original description of the present deliverable: “Plan for the software module for entering and storing the data, based on community input and findings and implementations from project partners, and new functionality”. The idea that the DAI Berlin group had in mind when formulating the task description was to improve and standardize the workflow for archaeometric lab data on bone analysis. However, in order to adequately use the expertise of the ISPC Rome and ISPC Lecce groups, rather than searching for a new archaeological community, the task partners have focused on what the task groups are good at and and will combine this expertise to achieve a new goal that the individual groups couldn’t have achieved on their own. Consequently, the focus of the deliverable as well as the whole task has shifted from creating a software module from scratch to connecting and improving selected existing systems. The partners still rely, of course, on input from the archaeological community to improve the workflows and user experience for documenting and working with excavation data.

The present document is the first of two deliverables (D5.17) in Task 5.7. It describes the implementation plan for the actual work in this task. The second deliverable will be the report on the final implementation (D5.18). Between the two deliverables there are two milestones. The first milestone (MS36) consists of two parts, namely (a) the identification of a suitable archaeological case study and (b) the implementation of prototypes. Part (a) is already done and described here. Part (b) as well as the second milestone (MS37: Publicly available Linked Open Data service) are steps in the implementation plan and will be reported on in D5.18.

2. Background

The work in this task requires a wide array of expertise, namely domain-specific knowledge about archaeology and archaeological excavations, IT knowledge for creating domain-specific software and computer-based workflows, experience with 3D modelling and reconstruction, and a thorough understanding of the use of norm data and long-term preservation of data. Thus, the main purpose of this section is to present the state of the art and the basis of the team's work and to introduce the main points and terminology before the descriptions of the concrete work presented in section 3. For IPCS Rome this is the ***Extended Matrix*** for generating 3D visualizations from data in an extended version of the archaeological Harris matrix and the ***EMBlender tools***, for DAI Berlin this is the ***idai.world*** and its excavation documentation system ***idai.field*** as well as its gazetteer for temporal terms ***ChronOntology***, and for ISPC Lecce this is the ***survey data*** of the ***Roman theatre in Catania***.

2.1 Extended Matrix

The Extended Matrix approach, defined by the ISPC Rome research team, is based on the *stratigraphic reading* approach[2] and aims to create a common framework connecting archaeological documentation and virtual reconstruction in the earliest stages of the excavation of a site or a 4D survey of a monument. The relationships between the stratigraphic units found during an excavation, e.g. on unit being on top of another unit, result in a so-called *Harris matrix*. However, stratigraphic units are not used only in archaeological excavations: they are also the base element in similar disciplines like building archaeology, restoration and landscape archaeology. Stratigraphic units indeed do not exclusively describe the information that derives from the excavation phase, which goes to document all those destructive activities that are typical of the phases of progressive removal of soil layers, but also map elements above ground (such as wall structures, plasters, decorations, frescoes). Through the population of the extended matrix, both objective information and interpretations derived from deductive, contextual or analogical information (i.e. derived by analogy) are documented (see Fig. 1).

[2] OSIRIS website, details on statigraphic approach: http://osiris.itabc.cnr.it/extendedmatrix/index.php/why-a-stratigraphic-approach/; [24 June 2020]

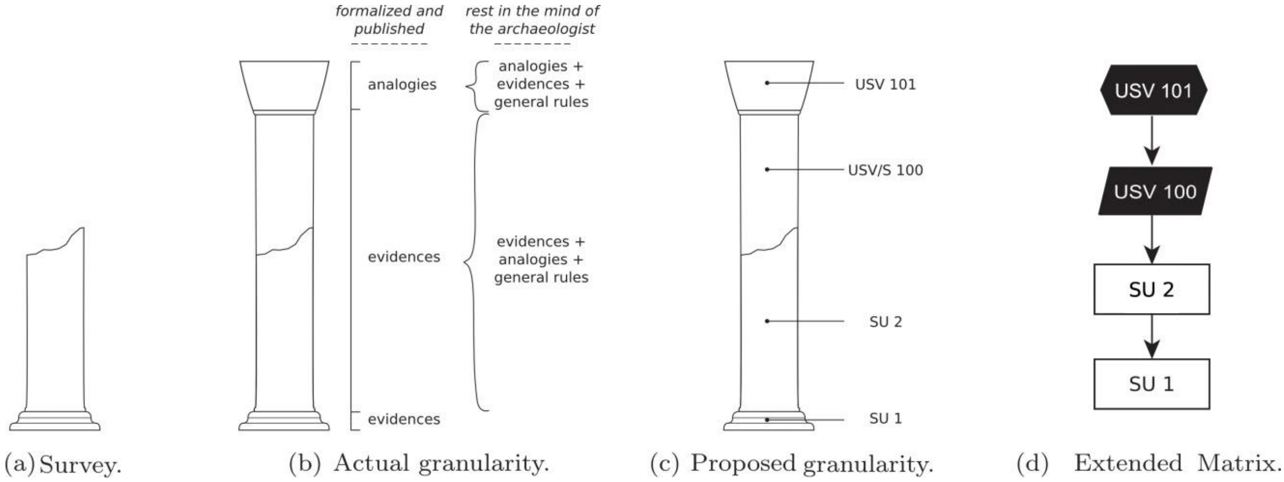

Fig. 1: 3D visualisation of archaeological stratigraphic units with the Extended Matrix (Source: Demetrescu 2015)

The Extended Matrix is populated through the free editor yED[3]. The yED editor is able to design Harris Matrices formally correctly by including some graphic elements that are often complicated to reproduce manually, such as bridges to avoid the crossing of connection lines between stratigraphic units. yED also has an effective automatic layout system useful to keep the dataset orderly and readable. The GraphML[4] XML file produced by yED is imported into Blender's[5] EMtools plugin[6], specifically designed to read the file and connect it with the geometries - simplified shapes called *proxies* - which represent the volumes associated with the various stratigraphic units, on which the virtual reconstruction will take shape.

Once the reconstruction phase is finished, the files generated by EMTools can be exported in a 3D format which in turn can be imported into the *EMviq* tool[7]: a complete, interactive 4D visualization and interrogation tool for Extended Matrices. By using this tool, the 3D reconstruction together with all related stratigraphic information can be interactively inspected, allowing users to virtually explore an entire archaeological context and spatially perform 3D queries. Once a proxy is hovered/selected through user input, a source-graph is presented to the user that validates that specific proxy.

[3] yED editor for drawing diagrams: https://www.yworks.com/products/yed; [24 June 2020]

[4] GraphML, an XML-based data format for storing mathematical graphs: http://graphml.graphdrawing.org/; [24 June 2020]

[5] Blender website: https://www.blender.org/; [24 June 2020]

[6] OSIRIS website, details on EMtools: http://osiris.itabc.cnr.it/extendedmatrix/index.php/extended-matrix-framework-emf/embt/; [24 June 2020]

[7] OSIRIS website, details on EMviq: http://osiris.itabc.cnr.it/scenebaker/index.php/projects/emviq/; [24 June 2020]

In order to deploy an interactive 4D inspection context that the user can interact with, three different extraction steps are defined from an Extended Matrix stored in a graphDB[8] (Fanini and Demetrescu 2018):

1. *Timeline extraction*: This step will extract from the GraphDB a finite number of time-periods, identified by ID, including beginning and end values.
2. *Proxy-graph extraction*: This will create a graph of proxies for real-time queries and interrogation, arranging semantic 3D shapes (proxies) into the 3D space including procedural rules (e.g. so-called seriation nodes).
3. *Source-graph extraction*: This will generate an internal runtime representation of EM sources relationships (paradata).

These computational steps (parsing the XML file) need to be performed only when the Extended Matrix (the GraphML file) is modified by the user. This is particularly important for cloud-based scenarios, where remote users are operating on the same Extended Matrix (GraphML file) through the yED editor, or other tools able to visually manipulate GraphML files.

EMblender tools: The Extend Matrix tool is an add-on written in Python for Blender 3D to facilitate the creation of 3D reconstructive hypotheses from an Extended Matrix (in GraphML format). The tool provides the concurrent preparation of a GraphML file in the yED editor according to the rules of the EM formal language. A technically skilled archaeologist edits the GraphML file in yED (OP1) while a modeling archaeologist (OP2) creates the proxy model inside Blender-EMtool. The tool allows a hot reload of the GraphML file inside Blender (which affects any stratigraphic data previously loaded). The entire workflow is carried out in four main work steps corresponding to four groups of actions. The work process is not destructive so that it is possible at any time to return to one or more previous steps:

(1) **Creation of the EM** concerning only the material remains of the study context (e.g. excavation stratigraphy or stratigraphic reading of elevations); import of the results of a 3D survey (if present, in the form of photogrammetric models or laser scanners) or of any plans and graphic documents related to the context; import of the GraphML file coming from yED; on the basis of this data it is now possible to model proxies or the simplified representation of the stratigraphy of the elements in situ; connection of the stratigraphic units represented in the extended matrix and the proxies modeled in 3D; Normally the workflow involves the creation of stratigraphic data in the Extended Matrix and the transfering the information to Blender, but in the case of very high resolution photogrammetric models it is possible to make the stratigraphic reading directly on the 3D model so it is OP2 to ask OP1 to add new stratigraphic units to the EM.

---

[8] GraphDB website: http://graphdb.ontotext.com/; [24 June 2020]

Fig. 2: EMtools workflow step 1- import photogrammetric model and archaeological plans

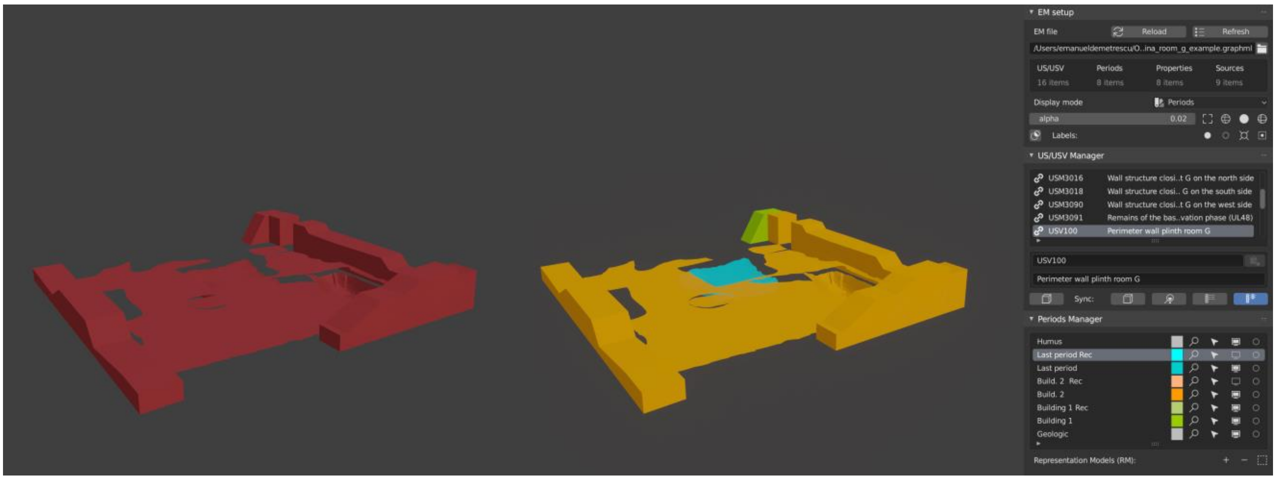

Fig. 3: EMtools workflow step 1- 3D modelling of the proxies (only real stratigraphic units) in EM color schema (left) and epoch color schema (right)

(2) **Creation/modification of the reconstructive hypothesis** in the EM by OP1 (if it has not been done yet); 3D modeling of proxies (OP2); connection between stratigraphic units and proxies; at the same time a diary of virtual activities is drawn up that represents a textual version of the reconstructive hypothesis, divided according to the supposed evolution/construction phases of the context; the 3D model is now ready for a first glimpse by people outside the team.

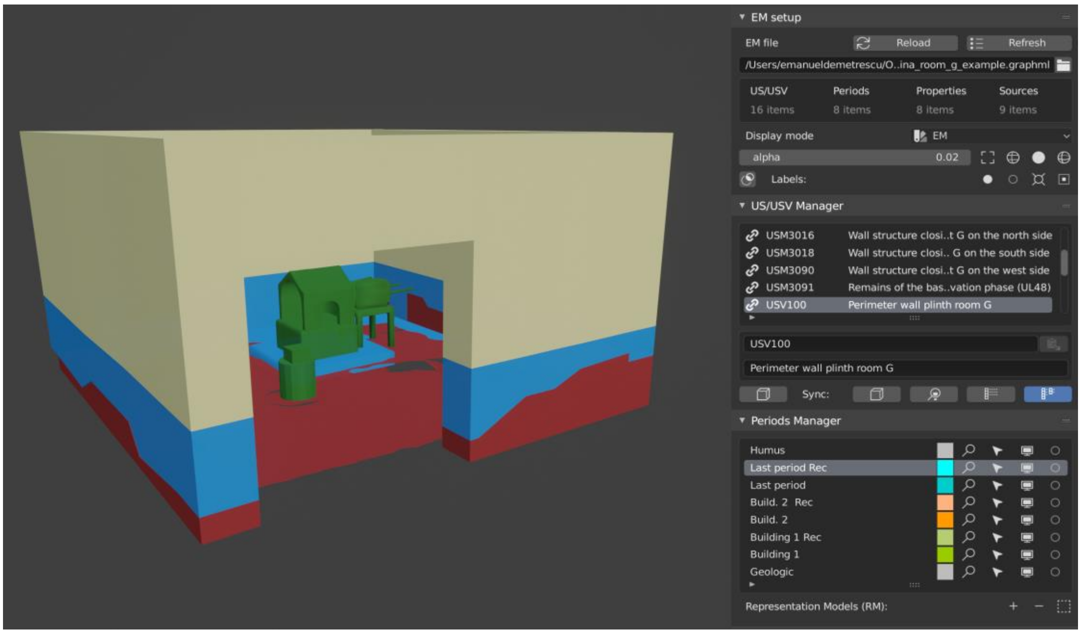

Fig. 4: EMtools workflow step 2- 3D modelling of the proxies (both real and reconstructive stratigraphic units) in EM color schema

(3) **3D modeling details**, including materials and textures, in a *representation model* (RM) for each phase over time of the context to be reconstructed; association of the 3D models with the chronological phases to which they belong.

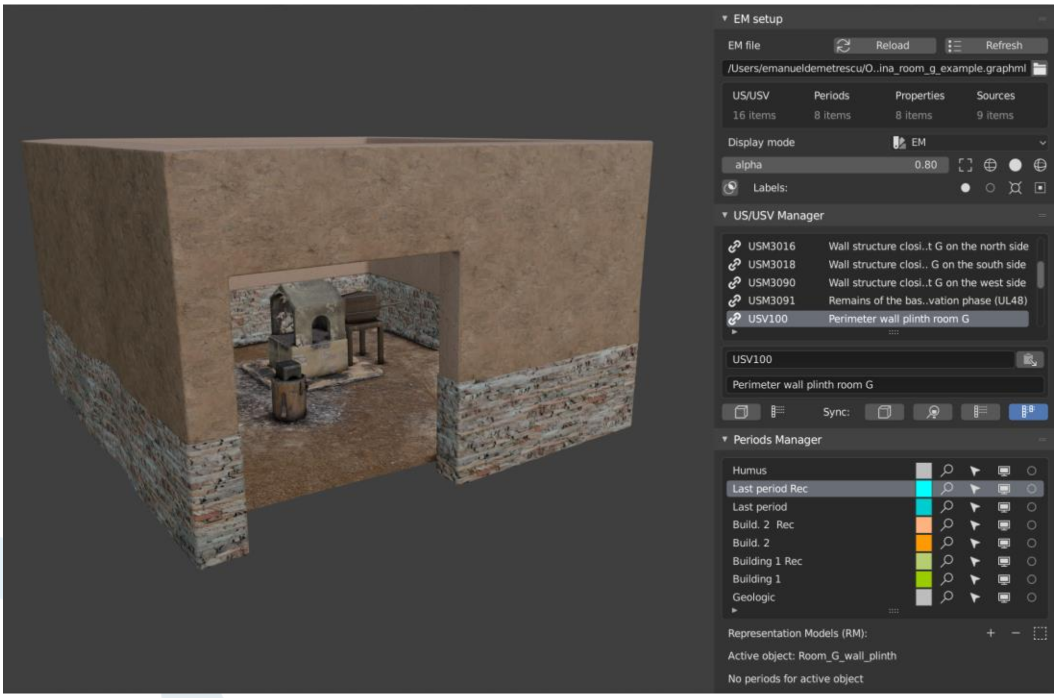

Fig. 5: EMtools workflow step 3- 3D modelling of the Representation Model (RM)

(4) After the modelling of an RM for a given epoch, it is time to **publish the reconstruction**. The Export manager will help the user to create tables for an article or a book chapter (CSV export) and an EMviq interactive web app (through ATON[9]) or a desktop installation (e.g. Unreal, Unity or Godot). The export of the reconstructive hypothesis within an "EMviq" coded folder directly into a Cloud space for archiving and sharing coincides with the online publication on the EMviq platform.

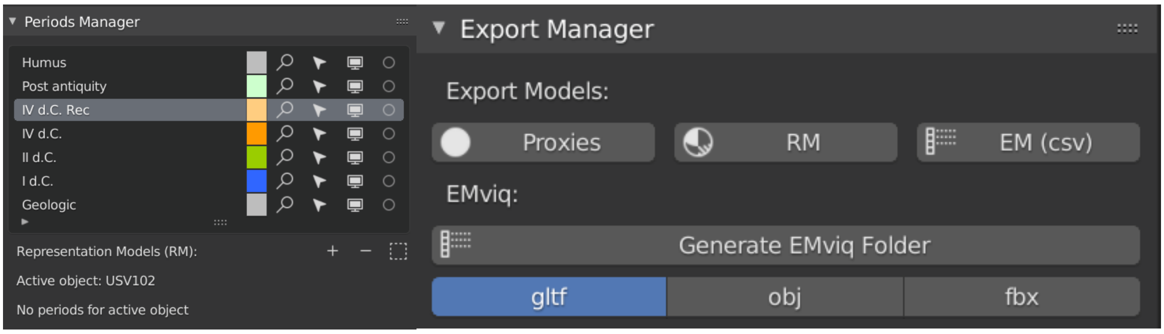

Fig. 6: EMtools panels - "Period Manager" and "Export manager"

2.2 idai.field and ChronOntology

The German Archaeological Institute (DAI) has developed the *iDAI.world*,[10] a system of open access webservices for archaeological data, based on and themselves open source software. Two categories of services are relevant here, namely (1) data services including *Arachne*, a large database of archaeological objects, and *idai.field*, a system for documenting excavation data, and (2) norm data services for data interoperability and standardization, notably *idai.gazetteer* for place-related data and idai.chronontology aka *ChronOntology* for temporal terms.

Arachne[11] is the overarching object database of the DAI. All other webservices are connected with Arachne. idai.field[12] can be used by archaeologists doing field work at an excavation with or without internet connection, in which case the data will be synced with a server later on. idai.field supports a concurrent workflow and is modular and scalable. The main difference between EM and idai.field is the focus in EM on 3D reconstruction and in idai.field on excavation documentation. Harris matrices and relations between stratigraphic units can be

[9] OSIRIS website, details on the ATON framework: http://osiris.itabc.cnr.it/scenebaker/index.php/projects/aton/; [24 June 2020]
[10] iDAI.world website, overview of all DAI webservices: https://idai.world/; [24 June 2020]
[11] iDAI.objects Arachne website: https://arachne.dainst.org/; [24 June 2020]
[12] iDAI.field download page: http://field.dainst.org/; [24 June 2020]

represented in idai.field but they are not the focus of the system, so the possibilities to visualize a Harris matrix in idai.field are limited.

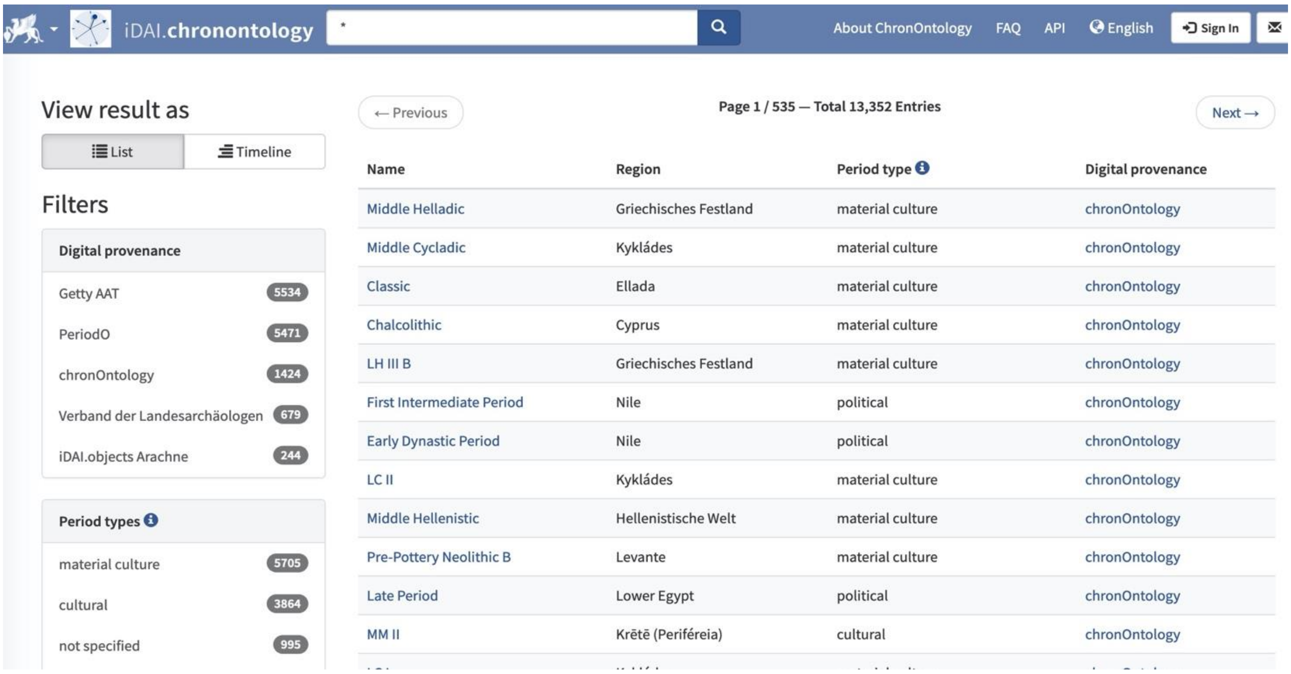

Fig. 7: Records of periods in ChronOntology

ChronOntology[13] is a gazetteer for temporal terms, i.e. a system for storing, managing, mapping and publishing temporal terms (i.e. periods such as Middle Helladic). ChronOntology allows a rich semantic modeling of various terminological systems for cultural periods. The data model can describe their semantic relations and is based on CIDOC CRM[14]. A central tenet is the separation of a temporal term's definition from any dating information about the term. For example, one can denote the reign of Augustus without stating or agreeing on strict start and end dates. Only the definition determines whether one is talking about the same thing, and temporal information is meaningful only in connection with the definition. However, since strict period definitions are often not the focus of the research literature, ChronOntology currently represents them by a system of period types such as "political" or "material culture" or subtypes such as "pottery". The view of periods as concepts with separate definitions and dating information complements other approaches to period gazetteers where spatiotemporal information is seen as part of the definition of the terms themselves.

Each defined temporal term is automatically associated with a phenomenological *spacetime volume* (STV), i.e. the area in space and time where it happened, regardless of what is known about its extent. STVs allow for

[13] iDAI.chronontology website: https://chronontology.dainst.org/; [24 June 2020]

[14] CIDOC CRM website: http://www.cidoc-crm.org/; [24 June 2020]

periods such as "Neolithic" to take place at different times in different regions. (Note, however, that the usage of terms is often not consistent. For example, "Neolithic in the Levant" might seem to denote the same as "Neolithic, limited in a strictly geographic sense to the Levant", but has in fact a different definition.) Any concretely given spatial and temporal information is an approximation of the phenomenological spacetime volume. ChronOntology uses widgets for space and time to visualise the concretely given information. ChronOntology itself concentrates on the temporal aspect, while the geographical aspect is handled by DAI's gazetteer for place names idai.gazetteer[15].

2.3 The Roman theatre in Catania

The ISPC Lecce group intends to apply the EM to their extensive data on the Roman theatre in Catania and connect it to the DAI norm data systems. The following subsections give a sketch of the history of the Roman theatre, explain why it will be a good illustration of the task's goal, and give an overview of the already existing data on the Roman theatre.

2.3.1 History of the Roman theatre

The Graeco-Roman theatre of Catania stands in the heart of the historic centre, on the south-eastern slopes of the hill of Montevergine, within an area that played a central role in the dynamics of the city's urban development. The original structure seems to date to the Greek period, as suggested by the remains of the walls in limestone blocks identified in several sectors of the monument.

The building visible today was built during the Julio-Claudian period as part of a programme that saw the rebuilding of the monument, which probably reused structures and materials from the earlier Greek theatre. Between the Antonine and Severan periods, the structure was extended and underwent transformations that gave the entire complex a monumental aspect. The stage building was substantially modified by the insertion of curving exedra into the rear wall of the side doors and the enrichment of the decorative scheme, while the cavea was enlarged by the addition of the third walkway and creation of a new seating tier. The theatre was decorated with marble columns, statues, and decorative reliefs celebrating public events associated with mythological themes. During the final period of use, which can be dated to around the 4th century AD, the building was probably adapted for aquatic shows, with the creation of channels and systems for filling it and draining water, work which also involved the repaving of the orchestra.

Between the 5th and 6th centuries AD the theatre lost its original function, as attested by the building of structures in the orchestra and gradual filling of the monument's lower part. From this moment onwards, the building slowly deteriorated, beginning with the robbing that took place throughout the medieval period and then with the gradual obliteration of the ancient structures caused by the unstoppable building activity, which

[15] iDAI.gazetteer website: https://gazetteer.dainst.org; [24 June 2020]

between the late medieval and modern periods determined the formation of a residential quarter known as "Grotte" in the theatre area (Malfitana et al. 2016).

2.3.2 Why the Roman theatre was selected

The Roman theatre in Catania was selected for this case study for several reasons. For one, at least at this early stage of our collaboration work a certain level of preservation is preferable for testing the workflow. The Roman theatre is certainly one of the best-preserved examples of this type of Roman architecture, and even here the characteristics represent peculiarities that have still to be fully documented. The problem faced by an interpretative and reconstructive study essentially lies in the absence of accurate and up to date graphic documentation. To remedy this lacuna, ISPC Lecce has undertaken a survey campaign using different techniques and tools (see below).

In addition, this archeological site presents evidence of different construction phases, which allows a focus on temporal data:

- Greek phase: wall structures built with large sandstone blocks with Greek letters in the V-VI century BC
- first Roman phase: This period sees a remaking of the previous wall structures, the *cavea* and the construction of the *scena* using square lava blocks.
- second Roman phase: Between the Antonine and Severan periods, the structure was extended and underwent transformations that gave the entire complex a monumental aspect. The stage building was substantially modified by the insertion of curving exedra into the rear wall of the side doors and the enrichment of the decorative scheme, while the *cavea* was enlarged by the addition of the third walkway and creation of a new seating tier. The theatre was decorated with marble columns, statues, and decorative reliefs celebrating public events associated with mythological themes.

The quantity and variety of available data, deriving from the use of different techniques, tools and methods, integrated to obtain an exhaustive representation of architectural geometry and its specific materiality, also contributes to making this case study particularly interesting. Thus, the Roman theatre gives the chance to fully test and validate the Extended Matrix scientific approach as well as the DAI norm data systems.

The selected case study also offers the opportunity to reflect on the crucial issue of the so-called scientific transparency of virtual archaeology projects. To achieve a high scientific rigour in a reconstructive archaeology project, it is essential to collect documentation and transparently present the entire work process. This allows the validation of the reconstruction and enables other researchers to review the results without necessarily starting a new study from scratch.

Finally, the case study of the Roman theatre in Catania will allow the ISPC Lecce team to compare the 3D reconstruction already carried out by the Lecce Laboratory with the reconstructive hypothesis that they will develop starting from the sources organized in the EM, offering a term for comparison and validation of the work done.

2.3.3 The survey campaign at the Roman theatre

To overcome the absence of accurate and updated graphic documentation of the Roman theatre that is indispensable for developing the reconstructive hypothesis of the site, a documentation and analysis campaign was conducted using these tools and methodologies:

(A) 3D model by laser scanning;
(B) digital photogrammetry images;
(C) photos;
(D) video shooting with a UAV (Unmanned Aerial Vehicle) and from the ground;
(E) scientific literature.

(A) **3D model by laser scanning**, with Leica ScanStation P20. The work in situ was set up with the aim of covering the entire complex, which due to its substantial size required the planning of a series of stations that would cover the entire external perimeter, the cavea and each individual walkway including the single flight of steps connecting them and those of the vomitoria. The coverage of all wall surfaces required the use of 39 stations, each set at a resolution of 6 mm on a dome of 10 m. Due to the quantity of data in the fusion process for each individual scan in a single points cloud, (about 3 billion points – 100 giga .pts) per extrapolation of the mesh, it was necessary to subdivide the files into four parts and work on each one separately (fig.8).

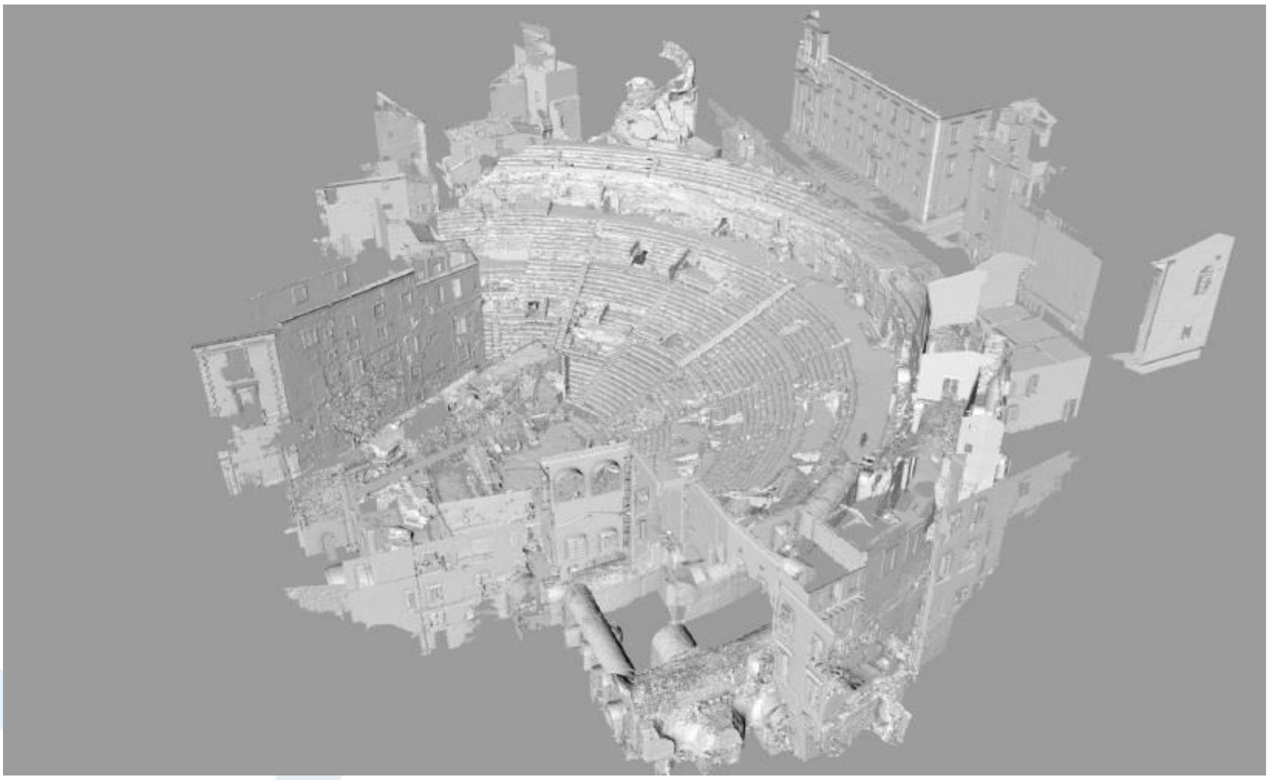

Fig. 8: 3D Model from laser scanner

(B) 3D graphic reproductions of of about 300 representative decorative architectural elements preserved within the structure, such as capitals, columns, cornices, pedestals, friezes, and statues, using **digital photogrammetry** (Camera canon 5D Mark II, software for data processing Agisoft Photoscan, fig. 9). Photogrammetry is a technique that uses multiple images of the same area to rebuild the 3D geometry of the objects in an entirely automatic manner with the help of complex algorithms from computer vision, called SFM (Structure from Motion)[16].

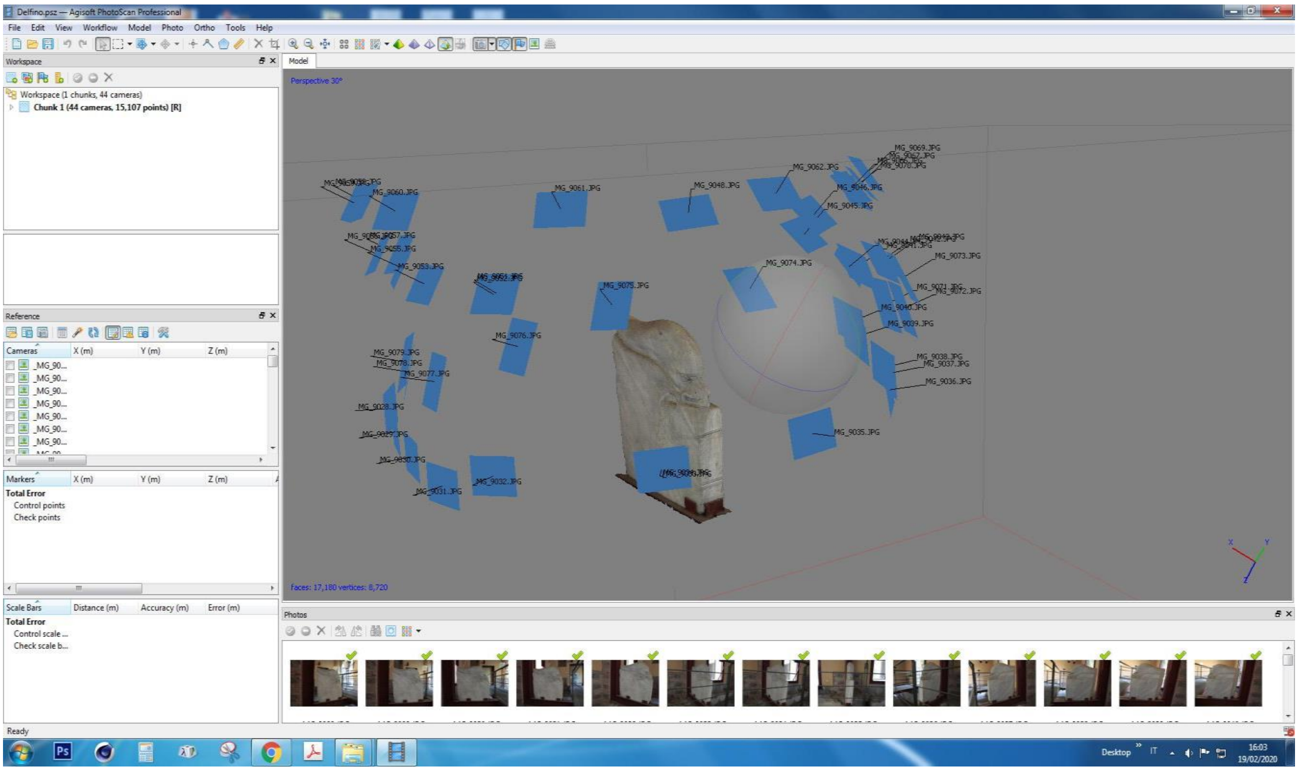

Fig. 9: 3D model from digital photogrammetry

(C) **Photographic survey** performed with a Canon Mark II 24 mpx (photos of architectural, decorative, construction techniques, about 700, fig. 10).

16 Structure from Motion: for introductions see e.g. https://alicevision.org/; [24 June 2020]; https://peterfalkingham.com/blog/; [24 June 2020]

Fig. 10: Photo of the Roman theatre in Catania (Source: CNR-IBAM, 2015 survey conducted by DiCeT project)

(D) **Video shooting** with a UAV (Unmanned Aerial Vehicle) and from the ground (together about 30 videos, fig. 11).

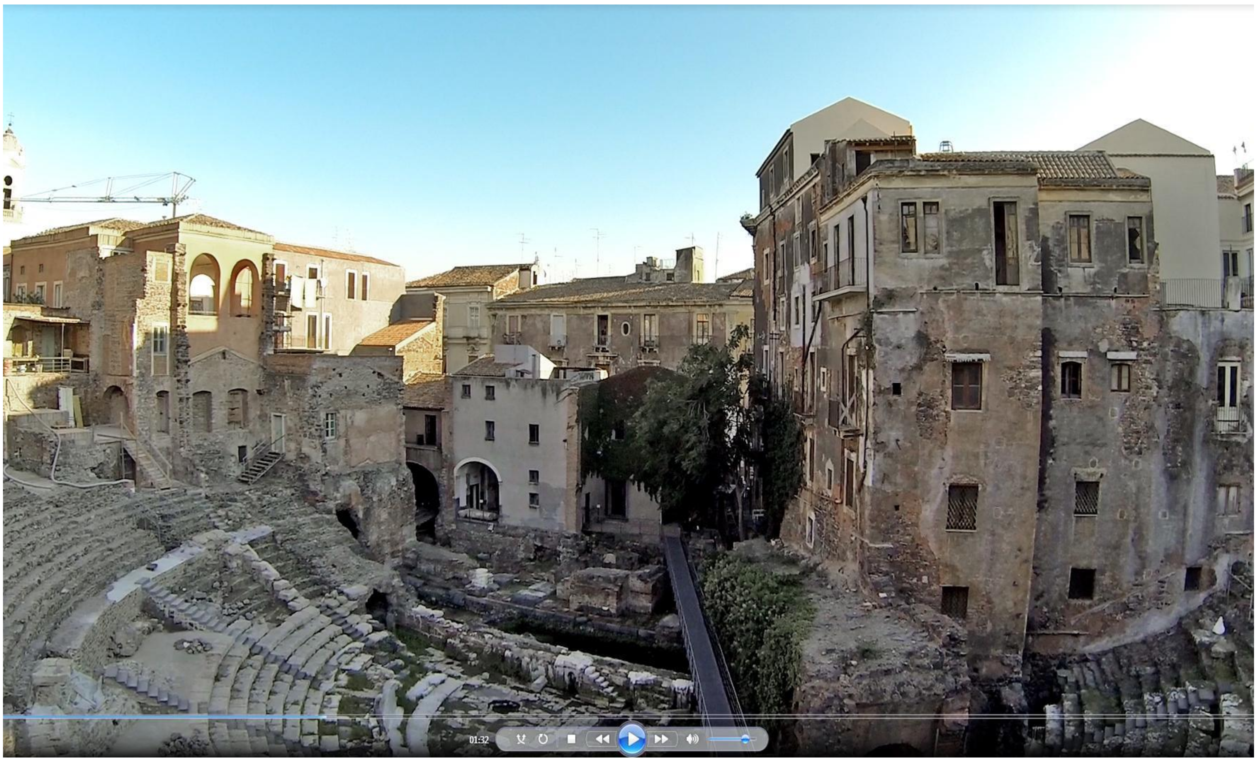

Fig. 11: Video of the theatre from UAV; screenshot (Source: CNR-IBAM, 2015 survey conducted by DiCeT project)

(E) In order to advance a reconstructive hypothesis of the theatre, all the documentation produced and analyzed was integrated with the data and information inferred from the **scientific literature** on the site (see Malfitana et al. 2016), and with typological comparisons with other contemporary or otherwise worthy monuments of interest.

3. Planned Work and Rationale

This section describes the planned work in this task and its rationale. After an overview, the role and planned work of each partner is detailed. This will be the basis for section 4, where the planned work is grouped into more succinct Work Steps sorted by topic rather than by task partner.

The general question was how the task partners can combine their strengths in the SSHOC project to better serve the archaeological community and at the same time make the case in the SSHOC project. More precisely, how archaeologists could benefit more easily from digital workflows and from using the task partners' systems, and how sharing and re-using archaeological data can be encouraged. The following points were identified:

***documentation*:** The EM developed at ISPC Rome and the ChronOntology temporal gazetteer developed at DAI Berlin are not easy to use for archaeologists who are not in direct contact with the people who developed these systems. DAI Berlin and ISPC Rome will document their systems better and test the documentation with the ISPC Lecce group and other volunteers.

***digital workflows*:** Based on the EM documentation, the ISPC Lecce group will apply the EM workflow to their own data and report back to the ISPC Rome group. A central goal is making the workflows simpler to use.

***data interoperability and standardization*:** DAI Berlin will make the case study an example for using norm data wherever possible, by using the DAI's idai.gazetteer and ChronOntology norm data systems as well as other systems. All partners will evaluate data standardization strategies to enable data sharing and re-use. DAI Berlin and ISPC Rome will also work on aligning the ontologies underlying their systems with CIDOC CRM.

***connecting systems*:** A general problem for interoperability is that the systems are simply not connected. DAI Berlin and ISPC Rome will evaluate the possibility to connect idai.field and EM via a GraphML file that is exported from idai.field and imported into EM. If there are no unforeseeable hurdles, the partners will create a workflow that combines the systems.

***the cloud*:** The task partners will work on moving the workflows away from data silos on individual computers to webservices. This includes the the possibility to work offline when there is no internet available, as is standard during excavations, and to synchronise the data later on via a webservice. To this end, the workflows need to be more robust by allowing concurrent work.

In the following sections, the intended work for each task partner is detailed. The planned collaborations with other SSHOC tasks are assigned to the task leader DAI Berlin. The work in this task will be highly collaborative.

3.1 ISPC Rome

At ISPC Rome, the focus lies on documentation, working on the ontology, a better workflow, the cloud, and combining systems.

***documentation*:** In order to make the systems more easily usable without direct help from its developers, ISPC Rome will write a White Paper describing the EM workflow in more detail than before.

ontology: The development of the EM has been going on for several years, and the SSHOC task 5.7 gives the opportunity to create the mapping to the CIDOC CRM ontology earlier than planned, i.e. rather than in version 2.0 of the Extended Matrix, it will be probably part of version 1.5 or 1.6. DAI Berlin will support the ontology work.

***workflow*:** With the tools developed by ISPC Rome for the enhancement of the Extended Matrix, the archaeological community has at its disposal new tools to connect the increasingly widespread photogrammetric 3D surveys of the archaeological excavation with the stratigraphic study of the site. On the one hand the tools, developed specifically for the Extended Matrix, will allow to collect archaeological data in a more efficient way, and on the other hand they will allow to start the activities of enhancement and virtual reconstruction of the site already during the excavation activity. In other words, the Extended Matrix framework helps to include in the field activity also the production of preliminary reconstructive hypotheses (that support the interpretation process) and on the other hand simplify the creation of high quality graphic material for the needs of dissemination of the excavation activity to a wider audience of non-specialists (according to the guidelines of social archaeology). ISPC Rome will offer webservices for the visualization of reconstructive hypotheses and the inspection of extended matrices (EMviq). The services will be available online both for desktop, tablet, smartphones and head-mounted display devices.

The already existing tools of the Extended Matrix as well as those that will be developed in this task allow to integrate a reconstructive method based on stratigraphic evidence based on already existing standards (Harris Matrix) with cutting-edge technologies such as the online publication of 3D models, the use of graph databases as well as communicative metaphors specific for virtual reality.

the cloud: One of the main objectives of this task is also to create a completely *web-based* tool called EMviq, developed on top of the open-source ATON framework. This task (already in progress) will offer all the functionalities of the original desktop tool, without any installation required by final users. Furthermore, it will allow to investigate advanced 4D search interfaces (for instance, a specific proxy or subset of proxies) matching specific search criteria at runtime (see fig. 12).

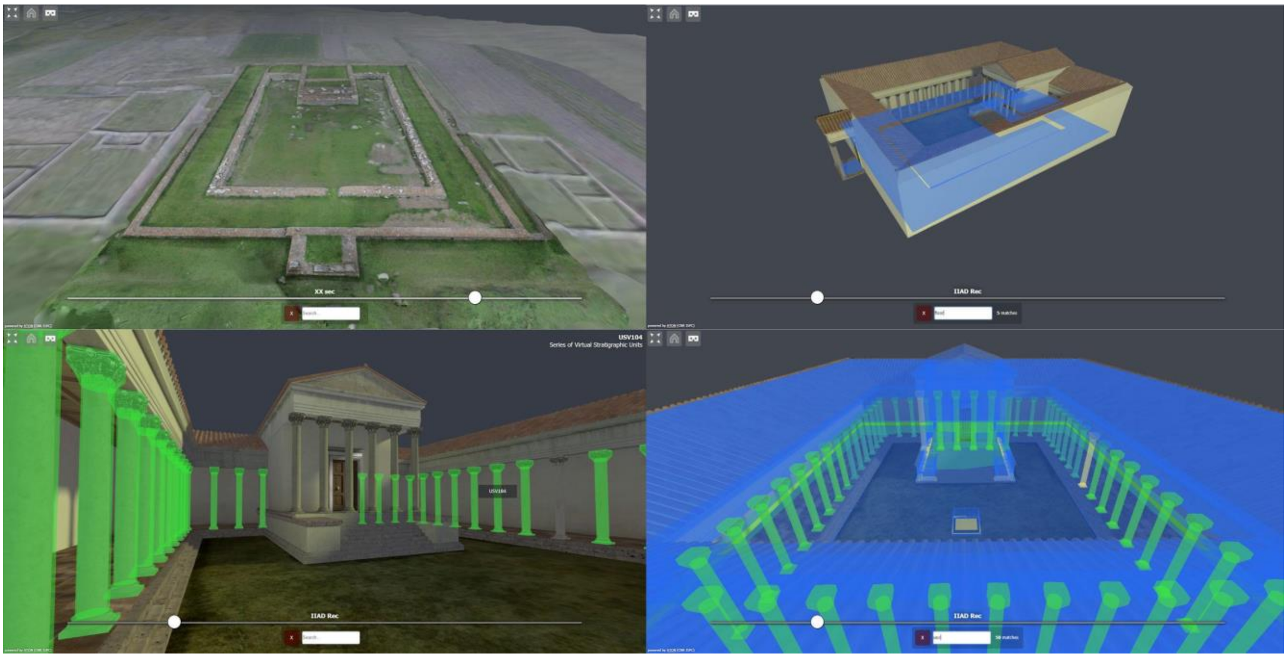

Fig. 12: A few snapshots of EMviq web prototype with timeline slider and integrated smart search to filter and highlight proxies at runtime

The tool will allow users to inspect 4D virtual environments interactively through a common web page, leveraging on modern standards like WebGL, WebXR API[17] and advanced HTML5 capabilities. Such integration will result in a "*universal*" tool, accessible from any device - including mobile devices (smartphones, tablets), desktop computers and immersive VR devices (consumer-level head-mounted display, HMD) - without user installation. The main user interface (UI) will provide functionalities to interactively explore time and space, inspecting and performing queries directly inside the multi-temporal virtual environment. The web-based tool will also offer simple but powerful visual validation features to assess and verify the *correctness* of Extended Matrices produced by the users. Concretely, the web-based service will support the following goals:

1. Boost the whole EM workflow among users collaborating on the same project, with the cloud system automatically resolving (potential) conflicting edits server-side

2. Adoption of existing open-source solutions (e.g. ownCloud[18]) for storing and sharing data will allow a fine-grained control on datasets (Fanini et al. 2019), enabling different policies that suit different *Extended Matrices* scenarios over the network. (Details about an authentication and authorization infrastructure are not decided yet.)

[17] W3C website, documentation of the WebXR Device API: https://www.w3.org/TR/webxr/; [24 June 2020]

[18] OwnCloud, a self-hosted cloud service: https://owncloud.org/; [24 June 2020]

3. Boost validation process of Extended Matrices (correctness of produced GraphML data using EM formal language)

combining systems: ISPC Rome will work together with DAI Berlin to evaluate the possibility of combining the EM framework with the DAI's idai.field. The EMF is intended to provide tools to ingest graph databases (Extended Matrices) that describe the relations between the stratigraphic units. Although idai.field is already able to manage a concurrent workflow and can handle Harris matrices, it is not focused on drawing stratigraphic diagrams (i.e. the graphical rendering of stratigraphic relations). The EMF provides a modular and scalable approach to extend the archaeological record stored in the idai.field database. The stratigraphic unit number is a robust primary key to connect these two databases (EM and idai.field) covering the same archaeological context. Stratigraphic unit numbers are therefore candidates for persistent identifiers and and LOD (see below).

The partners will seek collaborations inside (see below) as well as outside the SSHOC project. For example, the development modalities of the Extended Matrix will change in the coming months with the planned introduction of an *Open Extended Matrix Initiative* that comprises the creation of a scientific board and the collaboration of a group of laboratories focused on the concrete improvement of the semantic tool and software tools connected to them.

3.2 DAI Berlin

At DAI Berlin, the focus lies on documentation, data interoperability and standardization and on connecting systems. As the task leader, DAI Berlin will also liaise with other SSHOC tasks.

3.2.1 Documentation and interoperability

ChronOntology is a powerful system for time-related norm data but it is not yet intuitively understandable. DAI Berlin will document ChronOntology and its time ontology for users as well as contributors.

ChronOntology's time ontology is based on CIDOC CRM. DAI Berlin will publish an explicit mapping to CIDOC CRM, help create the CIDOC CRM mapping of the EM and coordinate the two mappings. This will be done in collaboration with SSHOC Task 4.6 (Semantic annotation of Heritage Science Data). DAI Berlin will also support the project partners in using and contributing to norm data systems from the DAI and others.

Interoperability shines where systems are connected, so the project partners will connect idai.field and EM if possible. (The DAI's systems are already ready for the cloud.) From idai.field's perspective, EM could serve as a tool that handles the visualization of Harris matrices. As a first step, DAI Berlin will likely add a GraphML export to idai.field that can be read by EM.

Another important aspect of the interoperability of the connected systems will be to provide the data as Linked Open Data (LOD). LOD is even part of the task title (as Open Linked Data). The project partners will provide LOD as far as it fits the workflows. The data exchange between idai.field and EM is likely a good place to start.

3.2.2 Collaboration with other SSHOC tasks

As task 5.7 is a case study for the goals that SSHOC stands for, communication with potential users and others is especially important. The task partners will collaborate with the tasks and infrastructures of the SSHOC project wherever possible, to the benefit of both sides. This will be mostly done by the DAI Berlin group as part of the project management. Sorted by services, data, interoperability and communities, the following collaborations are planned:

Services: DAI and ISPC are both prepared to host the data, but the project partners will evaluate whether they can use the SSHOC Dataverse to be created in Task 5.2 *Hosting and sharing data repositories*. The overall idea in Task 5.2, i.e. creating the SSHOC Dataverse as a workflow repository and data repository where the workflow descriptions can be stored alongside the datasets, seems very fitting for task 5.7.

The project partners will also publish their services and workflows to the SSHOC marketplace (WP7) if possible and will evaluate if there is a chance to include them in the SSHOC Switchboard (Task 3.6: *Making Data Re-usable*). Although this is unlikely at the moment since the Switchboard works predominantly with text-based services and needs the services to function without much user intervention, the project partners will monitor the progress of the Switchboard and see if there is a way to collaborate. For example, the Switchboard intends to integrate 3D Visualisation services as well, for example the Ariadne visual media service (AVMS)[19].

Data: The project partners will provide their data as an example to task 4.7 *Modeling the SSHOC data life cycle* and work closely with task 5.6 *Issues in providing Open Data in Heritage Science and Archaeology*. They also expect the exchange of ideas with task 4.6 *Semantic annotation of Heritage Science Data*.

Interoperability**:** The partners will contribute to task 3.2 *Selected Ontologies and Vocabularies* and task 3.5 *Data and Metadata Interoperability Hub*. Especially for the issue of scientific transparency they will collaborate again with task 5.2 *Hosting and sharing data repositories* if possible. In addition, they will collaborate with the new CLARIN initiative in SSHOC that aims to compile the controlled vocabularies used in SSHOC.

Communities**:** The partners will work with volunteers in SSHOC who are interested in archaeology. For this, the project partners will work closely with all tasks in WP6 (Fostering Communities, Empowering Users & Building Expertise) as well as task 9.1 *Identifying shared and unique challenges for SSH data communities; Evaluation and Usability Report*. The project partners may also organize a SSHOC workshop towards the end of the project, i.e. late 2021 or early 2022. The details need to be discussed.

[19] Website of the EU project ARIADNE, details on the Ariadne's Visual Media Service: http://visual.ariadne-infrastructure.eu/; [24 June 2020]

3.2.3 Outline of the collaboration with SSHOC T5.2 and T4.7

Collaboration with task 5.2 *Hosting and sharing data repositories* has the potential to bring services, data and interoperability together. The planned collaboration format is based on the case study of the Roman theatre in Catania, where ISPC Rome and DAI Berlin will evaluate whether they can make the workflows for individual 3D reconstruction projects re-usable. Task 5.2 aims at using *Dataverse* for storing and executing workflows for scientific experiments along with the data itself. This enables other researchers to reproduce the experiment or to replicate it with different datasets, both of which is vital for the focus on data transparency in Task 5.7.

Dataverse is a repository service that can be used to store datasets, programs, workflows, etc. One of the goals of Task 5.2 is to connect it with *Taverna*, a software system for creating and executing workflows. Dataverse and Taverna work with automated workflows, and there is no direct counterpart in our workflows. One could argue that once a user has prepared all the data to fit in the EM workflow, the system can create the 3D reconstruction with the single press of a button. However, this is not an experiment in the sense of the natural sciences. The real work has taken place before that, which is the part task 5.7 will focus on. In other words, the focus is not on automated workflows but on proposing workflows that can be used in archaeological situations. The idea is to track and annotate the archaeological workflow so that it is shareable and reusable by the community.

The term "workflow" is defined differently in the present task and in Task 5.2, and the tasks will first need to agree on a shared meaning of the term. Workflows in the sense of task 5.2 were initially used in the business environment as a way to describe the flow of activities through an organization and were later adopted also for scientific applications. A scientific workflow is a composition of interconnected and possibly heterogeneous scripts that are used in an experiment. Scientific workflow languages provide statements to define the logic that relates calls of scripts; for certain processes, such as statistical analysis, a linear flow might be sufficient, but more complex flows may allow for parallel execution, event handling, compensation handling and error handling. A scientific workflow may be considered as a way to record the origins of a result, how it is obtained, experimental methods used, machine calibrations and parameters, etc. Some examples of scientific workflows are (1) data chaining pipelines that gather and merge data from multiple sources, (2) sequences of steps automating repetitive tasks (e.g. data access, data transformation), and (3) complex iterative chains of MapReduce jobs etc.

The question of whether the approaches can be joined stays an open question the team will be working on, i.e. whether parts of the archaeological workflow have a sufficiently strong resemblance to lab experiments to make it possible to adapt the methods and to standardise the workflow description in a way similar to scientific experiments. The Task 5.7 project partners see a reasonable chance to connect their systems. The Dataverse tools may help to formulate how to store and share our workflows. Since the workflows in Task 5.7 are closely connected to individual 3D reconstructions, a focus might be on reproducibility and on providing examples of workflows, rather than a library of existing workflows for simple re-use.

It will be necessary that the Task's partners determine whether applying Dataverse and Taverna to the workflows in Task 5.7 can be successful. The task partners will need to analyse how Taverna and Dataverse work and play with experimental data in the tools. The taks partners will go through all steps of the archaeological workflow, i.e. collecting survey data in the field, the reconstruction process, the final reconstruction and dissemination, and see if they can apply the experiment idea as it is understood e.g. in molecular biology to archaeology and share the archaeological process along with the dataset. From a qualitative point of view, 3D reconstruction is different to e.g. data collection in the field, so some parts maybe be more easily adaptable to Dataverse than others.

There is an obvious connection with Task 4.7 *Modeling the SSHOC data life cycle* and the deliverable D4.18 *SSHOCro beta version here*. In addition, D4.18 highlights the ontological aspects of Dataverse. In a similar vein, the task partners see a strong connection of this work with their efforts to align our systems with CIDOC CRM. Since mapping data to CIDOC CRM is an exercise in clarifying the semantics of the data and the Dataverse ontology is also based on CIDOC CRM, the mappings to CIDOC CRM may help align the workflows with the Dataverse approach.

3.3 ISPC Lecce

At ISPC Lecce, the focus lies on digital workflows and the case study of the Roman theatre in Catania. The research group in Lecce will (1) share with the task 5.7 partners the different data used and produced by the ITLab (The Information Technologies Lab) to present its existing reconstructive hypothesis of the Roman Theatre in Catania; (2) enrich the dataset with the acquisition of new data in situ and in collaboration with ISPC Rome; (3) use the Extended Matrix approach to formulate a revised virtual reconstruction proposal for the archaeological site; (4) document the different sources used in order to ensure scientific accuracy, which will also allow testing and validating EM solutions; and (5) make a comparison between the existing hypothesis of 3D reconstruction of the monument and the proposal derived from Extended Matrix.

3.3.1 Roman theatre: applying data to EM

ISPC Lecce will document the virtual reconstruction process of the Roman theatre in Catania, using the Extended Matrix approach, following a specific workflow that will include the following two operative steps:

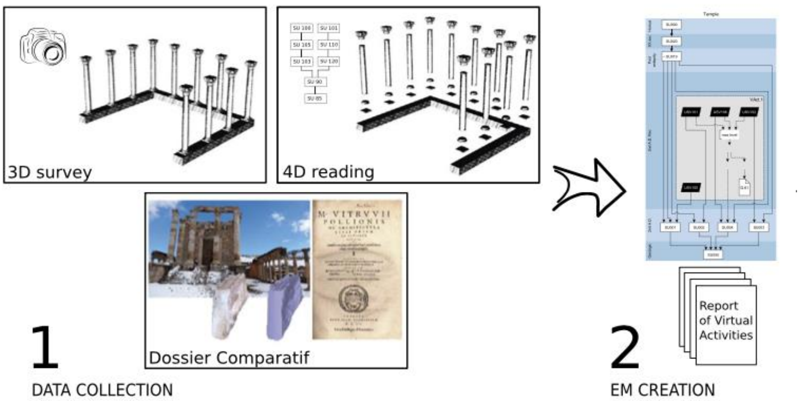

Fig. 13: Production work-flow, from data collection to EM creation *(Source: Demetrescu and Fanini 2017)*

1. **Data collection**: 3D survey of the remains and of the special findings (i.e. non in situ blocks of stone from the architectural apparatus), stratigraphic reading (4D analysis of the *palimpsest*), creation of a *Dossier Comparatif* (comparative study);

2. **EM model creation to organize data within the Extended Matrix**: writing down of the Extended Matrix and the Report of Virtual Activities;

These steps won't be separate phases of the reconstruction project: during data collection; ISPC Lecce will start populating the EM with interpretative elements as well as sketch out some visual representations of a given epoch. This synchronous workflow enables a cross-fertilization of the relationships between the archaeological elements, the external sources, and the hypotheses.

Data collection can include diversified methodologies and technical solutions. Different contexts require different approaches. Some scenarios cannot involve a 3D survey due to disparate limitations (physical, legal, etc.) but can make use of legacy data (drawings or blueprints). Furthermore, when the reconstruction project regards a completely lost context, all the data collection is focused on the sources organized within the Dossier Comparatif (DosCo).

3D Survey of the context: The Roman theatre in Catania is a palimpsest of different epochs in which each of these epochs must be purged of non-coeval elements in order to highlight the preserved ones and the overall shape of each chronological phase. Furthermore, a clear distinction between the different stratigraphies (grouped by epoch of belonging) will be a crucial phase to propose a valid reconstruction hypothesis. Alongside the *in situ* elements, the 3D acquisition will include the survey of all the non *in situ* objects. The 3D measurements result in accurate 3D models or digital replicas with colour information.

4D reading: the stratigraphic sequence and the Harris Matrix: During (or after) the 3D survey, a stratigraphic reading must be performed in order to highlight the temporal sequence (4th dimension) of the actions. The Harris Matrix visually summarizes the stratigraphic elements involved and the epochs which they belong to.

Dossier Comparatif: All the sources and the comparisons with other contexts will be stored in the Dossier Comparativ (Gros 1995): The Dossier Comparatif (DosCo) is a collection of documents which follows a specific nomenclature, a composite known as "D."(Document), plus a sequential numbering (i.e. D.01). The number is set according to the chronological sequence of data ingestion. All the documents are linked inside the EM and used to validate the reconstruction hypothesis. These are represented inside the EM through the *source node* (Demetrescu 2015).

Creation of the Extended Matrix and the Report of Virtual Activities: In archaeological practice, a virtual reconstruction is usually created at the end of an excavation project or 3D survey and is not part of the research itself. The 3D model, in this sense, is intended as an ex-post work which synthesizes different hypotheses made during the investigations (Medri 2003). Generally, these are not formally annotated. In some cases, experts write down intermediate reports to fix certain general ideas, but the collection of the archaeological record lacks a precise way to store the rich connections that are recognized between pieces of evidence during the fieldwork.

ISPC Lecce will use the EM approach to fix this issue. In this case study, virtual reconstruction will be part of the archaeological investigation starting from the bibliographic research about this archeological site and the data from the documentation and analysis of the campaign carried out. The reconstruction hypothesis will formally be "documented" along with the archaeological record (even if taken from already published archeological survey) which it is derived from. "Early collection" simplifies the management of the reconstructive record and "drives" the subsequent source-based modelling. This approach will make it possible to publish a thorough account of the scientific process, highlighting the connection between the archaeological and the reconstructive records and enabling other researchers to actively check, modify, and consciously reuse the virtual reconstruction.

Report of Virtual Activities: For real stratigraphic units, Virtual Reconstruction Units (USV) are combined in Virtual Activities (VActs, see Fig. 14) using rectangular shapes in the EM canvas. VActs are described in the Report of Virtual Activities (Demetrescu 2015), a textual report that acts as an intermediate output and is useful for quickly sharing a reconstruction hypothesis. A Report of Activities will be written down for each reconstructed epoch.

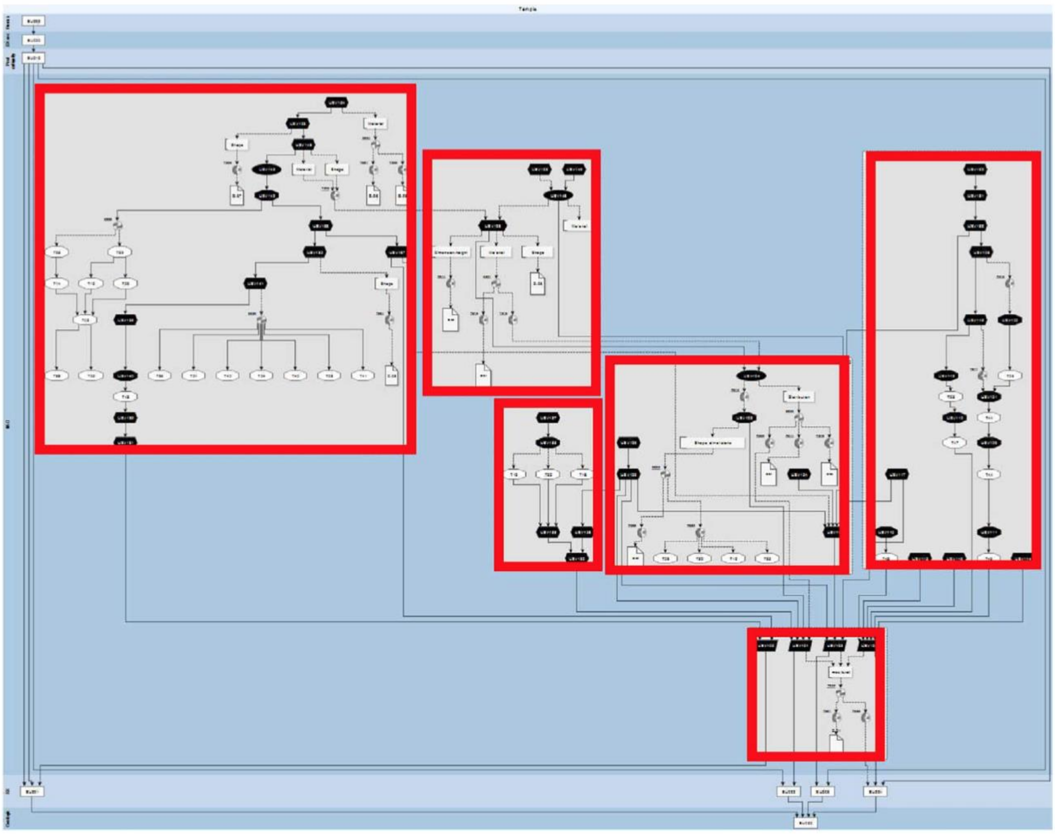

Fig. 14: Extended Matrix represents the Great Temple in all its epochs and hypotheses, making explicit all the elements involved in the reconstruction and their relationships. Marked in red are the Virtual Activities. (Source: Fanini and Demetrescu 2017)

3.3.2 Virtual reconstruction

Creation of 3D models (source-based modelling): Starting from the sources organized in the EM (see Fig. 15), ISPC Lecce will perform the 3D modelling, with different levels of representation from small details to a broad perspective:

- Digital restoration of digital replicas;
- Digital anastylosis of the elements in order to restore their original position and spatial relationships;
  Virtual Reconstruction of a given epoch starting from all the sources available and the anastyloses performed. It is important to take into account that from a reality-based model it is possible to make several reconstruction hypotheses: ideally at least one for each period formalized in the Harris Matrix.

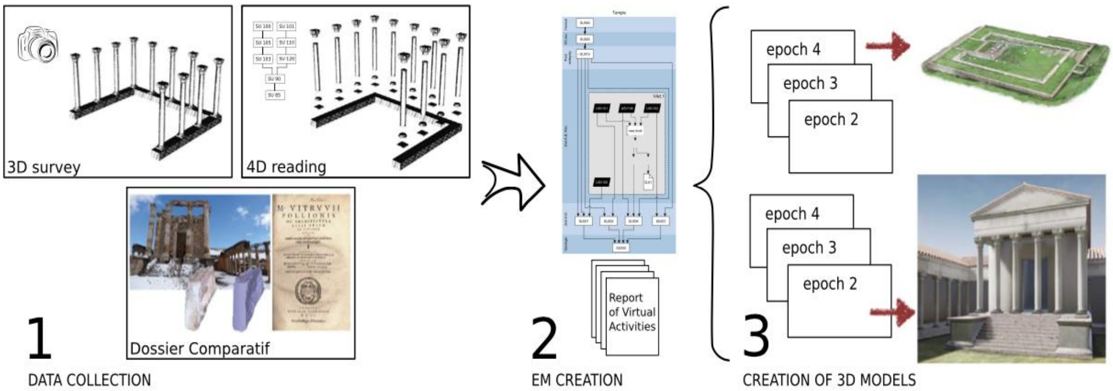

Fig. 15: Complete production work-flow, from data collection to EM creation and creation of 3D models (Source: Demetrescu and Fanini 2017)

The proxy-representation model approach: The creation of 3D content will follow two levels of abstraction: proxy models which are simplified representations of the reconstruction through basic geometrical shapes (cylinders, boxes, spheres, etc.) and representation models which are focused on fine geometries, colour, and material simulations resulting in the final, aesthetic depiction of the reconstruction hypothesis. The Proxy level will make it possible to highlight portions of the representation model and to interact with it. Furthermore, it also represents the first draft model of the reconstruction and is useful, along with the Report of Virtual Activities and the EM, for the sharing of intermediate results with colleagues and other experts and for providing feedback.

Graphical documentation for the reconstruction hypothesis: As mentioned before, the reconstruction hypothesis of a context has two depictions: a proxy model and a representation model. These models make it possible to explore and experience the research results in a virtual space. However, the EM approach is not focused only on digital media, but it pertains, first of all, to the archaeological documentation for a site's interpretation and scientific publication. Each stratigraphic unit - virtual or real - has a proxy representation. Making sections and plans out of proxy models helps provide technical documentation along with the description of virtual activities. Extended Matrix, Report of Virtual Activities, and graphical documentation published together allow for a coherent formal representation of the reconstruction hypothesis.

Validation of the reconstructed model: The construction of a 3D representation model of the archaeological site of the Roman theatre in Catania (or part of it) will be the final part of the practical application of the virtual reconstruction approach based on Extended Matrix. Among other reasons, the Roman theatre was chosen because, in addition to having already an important series of raw data (bibliographic, laser scans, photogrammetric and photographic surveys), it already has a well-documented virtual reconstruction that can certainly represent the first and most direct yardstick to evaluate the effectiveness of the EM approach.

In fact, although the new and the old representation are both 3D representations of the same archaeological site, they are the result of work processes that have different approaches, objectives and practices: the former more streamlined, expeditious, oriented to a broad audience, is particularly focussed to the high quality photo-realism of the final product; the second, the EM in fact, aims to link the sources to the corresponding reconstructive hypotheses and therefore also more onerous in carrying out the entire process previously analyzed also because of a much finer granularity of the elements.

Therefore, the primary validation method will be to evaluate how closely the new reconstruction adheres to the previously elaborated one, and critically analyze any discrepancies in order to understand if the use of EM has actually brought an advantage not only in the documentation phase of the reconstructive process, but also in its final scientific technical (photorealism) quality with a clear relationship between source and reconstructive hypothesis. This kind of primary validation will be performed jointly by the ISPC Lecce and Rome teams, and therefore this could be defined **internal validation**.

Beyond the case study in Task 5.7, an **external validation** process could be even more significant and ambitious. It could be possible to provide other interested working groups with access to the project data, together with the first stages of the virtual representation. This would leave open the finalization part of the virtual representation model, which would be performed independently by each working group, resulting in different virtual representations that can then be compared.

4. Work Steps by Topic

Based on the detailed description in section 3, in this section the work plan is grouped by topic and formulated as concrete Work Steps. The work may be shortened sensibly in the course of the project if the work shows to be more demanding than planned or some parts unnecessary. The teams involved the most in each Work Step are given in parentheses.

4.1 Documentation and support

In order to make the systems more easily usable without direct help from its developers, ISPC Rome will write a White Paper describing the EM workflow in more detail than before. DAI Berlin will write documentation on the ChronOntology temporal gazetteer (idai.gazetteer is already well documented).

To test the documentation, ISPC Lecce will apply the workflow to a case study (see the WSs about the Roman theatre). In addition, volunteers within SSHOC with an interest in digital archaeology can read a draft of the documentation and apply the systems to their own data (to a smaller amount than the reconstruction of the Roman theatre in Catania). The results will be used to create updated versions of the documentation and will be a sanity check whether the intended work plans will actually help people that are not part of the groups involved in Task 5.7, or where changes to the work plans need to be made.

In addition to the documentation, ISPC Rome will support ISPC Lecce (large case study), DAI Berlin (small case study) and other volunteers in SSHOC in applying the EM, i.e. it will provide tools and knowledge for using yED to populate the matrix. DAI Berlin will support ISPC Lecce and other volunteers in using ChronOntology and norm data systems in general.

(ISPC Rome, DAI Berlin)

4.2 Roman theatre: preparing the data

The case study for applying the EM workflow and using spatio-temporal norm data has been identified as the Roman theatre in Catania, where ISPC Lecce has already done an extensive survey. ISPC Lecce will make sure that all existing data is readily available and documented. The numerous and different kinds of data already retrieved and processed used to develop the reconstructive hypothesis of the Roman Theatre in Catania will be shared for a joint analysis in view of their uploading to the EM. This activity will also offer the opportunity to identify missing data, or other useful data that can be integrated with those already available.

In order to enrich the scientific dataset behind the virtual reconstruction hypothesis of the archaeological site through EM, and to make it a good example for using spatio-temporal norm data, ISPC Lecce and ISPC Rome (in a supporting role) will carry out a new graphic documentation campaign of the Roman theatre. Particular attention will be paid to the restoration works that have transformed the monument over time, to graphically represent these stratifications in EM.

A trip to the Roman theatre in Catania is planned towards the end of 2020. DAI Berlin may join ISPC Lecce and ISPC Rome in Catania to get a first-hand impression of the Roman theatre. Although travel planning is difficult at the moment due to COVID-19, this would be a good point in time to make this trip since it gives ISPC Lecce time to familiarize themselves with the EM system and to prepare and analyse the already existing data.

To enrich sources with the new data to be acquired in situ (Catania) it will be necessary to ask for authorizations both to access the monument and to carry out the new documentation campaign.

(ISPC Lecce)

4.3 Roman theatre: Virtual reconstruction

ISPC Lecce will apply the data to the EM, including the report of virtual activities. Based on the data in the EM, ISPC Lecce will create a virtual reconstruction of the Roman theatre. ISPC Lecce will formulate a virtual reconstruction proposal for the archaeological site and track the different sources used, in order to ensure scientific accuracy. This work will also allow them to test and validate EM solutions. A comparison will be made between the hypothesis of 3D reconstruction of the monument already made by the Lecce research group, and the proposal derived from Extended Matrix. All steps are described in more detail in section 3.

(ISPC Lecce)

4.4 Norm data

DAI Berlin will support ISPC Rome and Lecce to use norm data wherever possible. A focus will lie on DAI's systems, but other norm data sources will be used as well.

ChronOntology, and in extension the idai.gazetteer, will be used to record temporal terms and denote them by standardized IDs. As usual for projects using DAI's norm data systems, new norm data will be added if necessary, in this case periods and events that are not yet in ChronOntology. The task partners will also formulate the relationships between the different time layers in the Roman theatre, either as relationships between the periods in ChronOntology, or as additional relationships in a separate system.

DAI Berlin will use the case study to see if they can improve ChronOntology's search capabilities both for humans and machines (i.e. via the search API). A goal is to help find the right temporal norm data based on some given space/time data.

DAI Berlin will also evaluate the idea of using EM's visualisation capabilities to visualize ChronOntology periods and their relationships. ChronOntology has a limited way of visualizing spatio-temporal data. The EM may not only be a good visualization tool to help people visualize their archaeological data in place and time but also for temporal norm data with its associated spacetime volumes (STVs, see above) with their causal relationships (e.g. the Middle Augustan STV is by definition contained in the Augustan STV) and non-causal relationships (e.g. when two STVs just happen to overlap).

(DAI Berlin, ISPC Lecce)

4.5 LOD, interoperability and standardization

In addition to using norm data, the task partners want to increase the interoperability of their workflows and data. In particular, they want to foster the re-usability of the data. One aspect will be to provide the data as Linked Open Data (LOD) as far as possible. A promising starting point are the stratigraphic units of the Harris matrix, but if possible, LOD should be extended to the data from the whole workflow. Questions to be answered are: Can the tools import LOD? Is it a good archiving format? Who else would be interested in using this LOD? In addition, the task partners will collaborate with the new CLARIN initiative in SSHOC for compiling controlled vocabularies.

(DAI Berlin)

4.6 Connecting ontologies via CIDOC CRM

The Extended Matrix will see in future versions (1.5 or 1.6) a formal mapping to CIDOC CRM and its extensions, in particular with regard to the extractor node elements (as a scientific observation, see CRMsci), source node (as a cultural element), stratigraphic unit node (see CRMarchaeo) and epoch field (through ChronOntology). The task partners will also evaluate whether additional CIDOC CRM extensions can be applied, such as CRMba (documentation of archaeological buildings) and CRMgeo. This work will be done in collaboration with SSHOC Task 4.6 *Semantic annotation of Heritage Science Data*.

(ISPC Rome, DAI Berlin)

4.7 Connecting idai.field and EM

The task partners will evaluate the feasibility of connecting idai.field and EM with their respective foci on excavation documentation and 3D visualization. They will start by selecting some sample DAI excavation data as represented in idai.field that would benefit from a 3D visualization. They will then use the existing workflow for connecting it to EM and evaluate the effort as well as the synergies for combining the systems. One scenario would be to add a GraphML export to idai.field that can then be read by EM. This way, EM could handle the visualization of Harris matrices created in idai.field.

(DAI Berlin, ISPC Rome)

4.8 Cloud-based multi-user EM system

Within the *EMviq* tool being developed by ISPC Rome (see section 2), a cloud-based approach will allow archaeologists and other users to design and modify their EM through yED, while remote users concurrently inspect and validate the process inside the virtual 4D space. This step will be realized through the integration of a subset of EMviq functionalities with a cloud system, allowing users to operate remotely on Extended

Matrices. Concretely, ISPC Rome envisions running EMviq as a web-based multi-user application with a mechanism for automated data conflict resolution. It still needs to be decided which activities and data should be centralised on a server and which activities and data should remain local. The project partners will also clarify the roles of different actors in the workflow.

(ISPC Rome)

4.9 Implementation report

All three groups will write up the results of this task in Deliverable D5.18 *Implementation report (Archaeology case study)*, which is due in M38 of the project (February 2022). They may also organize a SSHOC workshop towards the end of the project, i.e. late 2021 or early 2022. The details need to be discussed.

(DAI Berlin, ISPC Rome, ISPC Lecce)

5. References

Demetrescu, E. (2015). Archaeological Stratigraphy as a formal language for virtual reconstruction. Theory and practice. J. Archaeol. Sci. 57, 42–55.

Demetrescu, E. & B. Fanini (2017), A White-Box Framework to Oversee Archaeological Virtual Reconstructions in Space and Time: Methods and Tools. Journal of Archaeological Science: Reports 14: 500–514, 2017.

Fanini, B. & Demetrescu, E. (2018). Carving Time and Space: A Mutual Stimulation of IT and Archaeology to Craft Multidimensional VR Data-Inspection. In International and Interdisciplinary Conference on Digital Environments for Education, Arts and Heritage (pp. 553-565). Springer, Cham.

Fanini, B., Pescarin, S., & Palombini, A. (2019). A cloud-based architecture for processing and dissemination of 3D landscapes online. Digital Applications in Archaeology and Cultural Heritage, 14.

Gros, P.P. (1995). Hercule à Glanum: Sanctuaires de transhumance et développement urbain. Galliapp. 311–331.

Malfitana, D., Gabellone, F., Cacciaguerra, G., Ferrari, I., Giuri, F., Pantellaro, C. (2016), Critical reading of surviving structures starting from old studies for new reconstructive proposal of the Roman theatre of Catania, in Proceedings of the 8th International Congress on Archaeology, Computer Graphics, Cultural Heritage and Innovation 'ARQUEOLÓGICA 2.0' in Valencia (Spain), Sept. 5 – 7, 2016. Available online: http://itlab.ibam.cnr.it/new_itlab/wp-content/uploads/2017/12/Arqueologica-Critical-reading-of-surviving-structure-Theatre-of-Catania.pdf; [24 June 2020]

Medri, M.M. (2003). Manuale di rilievo archeologico. Roma-Bari.

List of Figures