

TRILOGY: A Platform of Platforms for Advanced Digitization & Digital Curation



Stelios C. A. Thomopoulos*, Christina P. Thomopoulos, Giorgos Farazis, Panagiotis Tsimpiridis, Eleni-Ino Theodorou, Dimitris Zacharakis, Konstantinos Panou, Konstantinos I. Dimitros, Christos Maroglou, Efstathios Georgiou, Korina Kassianou, Sofia Mitsigkola, Michael Bampatsikos and Nick Argyreas

National Center of Scientific Research "Demokritos", Institute of Informatics & Telecommunications, Integrated Systems Laboratory

Submission: December 07, 2020; **Published:** December 21, 2020

***Corresponding author:** Stelios C. A. Thomopoulos, National Center of Scientific Research "Demokritos", Institute of Informatics & Telecommunications, Integrated Systems Laboratory, P.O. Box 60037, 15310, Agia Paraskevi, Greece

Abstracts

In this paper we present an overview of TRILOGY, a platform of platforms for Advanced Digitization & Digital Curation (ADDC), that results from the interlinking of the three digital platforms SYNTHESIS, NARRATION (AFIGISSI in Greek) and COSMOS. TRILOGY is developed at the Integrated Systems Laboratory¹ (ISL) of the Institute of Informatics & Telecommunications at the National Center of Scientific Research "Demokritos²" through a series of funded projects acknowledged at the end of the paper. TRILOGY aims at providing one coherent composite ADDC platform to enable Advanced Digitization (AD) and Digital Curation (as defined later on in the paper) and pave the way to the realization of a Digital Immersive Cultural Environment (DICE) [1]. The synthesis of the three platforms in TRILOGY provides an integrated platform for Advanced Digitization and Digital Curation (ADDC) of diverse tangible and intangible Cultural Heritage (CH) digital assets across different data formats, linguistic changes and, over time, semantic alterations. TRILOGY allows to create, represent and virtually interrelate CH digital content using graph models, where node interconnections lead to conceptually related narratives. Curating and interrelating semantically augmented digital assets is facilitated using AI (Artificial Intelligence) algorithms, in addition to human semantic assignment, enhanced by the implementation of a fuzzy semantic indexing system which leads to a common semantic framework capable of modelling digital assets that share affinities, including spatial and temporal information, even in cases of linguistic divergence. Moreover, a multi-modal and user-friendly visual representation of retrieved information serves as a comprehensive means to research and create affinities and associations between facts and data around this newly created, advanced, digital content using semantic representations, assisted by AI tools, across different data formats, linguistic changes and over time semantic alterations.

Keywords: Artificial Intelligence; Telecommunication; Cultural Heritage; Cultural Environment; Digitization; Archeology; Architecture

Introduction

At the moment, considerable Cultural Heritage (CH) material worldwide has already been digitized and exists in a variety of collections, databases in public and/or private archives. However, many digital CH assets and resources remain static, underused and decontextualized. Much digitized CH content is often disconnected from its origins, associated narratives and lacking in contextual framing that gives it its meaning and liveliness. In this way, many digital CH resources remain underutilized, underappreciated and unseen in public life. The creation of a Digital Immersive Cultural Environment (DICE) [1] alongside with tools for the creation of advanced digitizations and the curation of digital assets are paramount for a holistic, interactive, immersive and personalized

digital cultural experience, either as an entirely virtual reality experience, or an augmented reality experience harmoniously coordinated with an existing physical environment.

In the context of this paper we define Advanced Digitisation (AD) as "the creation and storage of multi modal digital assets alongside their deep contextual interrelations that interlink them with other tangible or intangible CH digitized artefacts, leading to the creation of dynamic cultural narratives and innovative pathways to the curation of new experiences." Based on the above definition, we provide an overview of how the three platforms, namely SYNTHESIS, NARRATION (AFIGISSI) and COSMOS, are integrated into TRILOGY, a platform of platforms, to provide a

¹<https://www.iit.demokritos.gr/labs/isl/>

²<https://www.iit.demokritos.gr/>

coherent composite environment for the creation of advanced digitizations, their curation alongside dynamic narratives, and subsequently their experiencing by end users, and the public at large, in an immersive and personalized way over a variety of different user interface modalities.

Digital Platforms for Cultural Heritage

At the moment, there exists a wide variety of digital platforms and repositories for cultural heritage developed over the past years. Important examples of this include Europeana³. In this paper we describe an ADDC composite platform that, instead of acting as a digital repository of CH, goes beyond the state of the art by actually creating tools with which such repositories can be of renewed social, cultural and economic use. This is both due to the kind of tools in each of the three platforms in the trilogy, as well as to the results made possible due to the combination of these tools within TRILOGY. The TRILOGYADD platform demonstrates advanced curation capabilities within the scope of a specific CH thematic using (digital and digitized) CH artefacts to show the platform's capability of representing different aspects of these artefacts from a multiplicity of perspectives (historical, socio - economic, cultural) along with their potential relations and affinities. Moreover, TRILOGY extends static information storing procedures with dynamic conceptual graph modeling capabilities of tangible and intangible cultural assets and their potential correlations, semantically annotated and represented using state of the art visual, textual and audio semantic representation techniques, validated by related stakeholders both in scientific and research field, and creative industry. Furthermore, the ADDC platform includes developing a creative user-friendly interface which highlights the importance of visualizing contextual commonalities (and differences) in order to create new connections between disparate CH content. The way TRILOGY goes beyond the state of the art through the use of multiple levels of metadata and semantics makes possible to access simultaneously, making it easier for stakeholders and end-users to gain access to material and knowledge that remains off-limits to them due to barriers posed by linguistic and other localization parameters. The ambition behind TRILOGY is to facilitate processes for Creative Industries working with audiovisual materials to not only host materials but inspire them and incubate the very connections that lead to innovative storytelling to take place in the first place.

The TRILOGY ADDC platform of platforms provides an integrated platform for advanced digitization of diverse tangible and intangible Cultural Heritage (CH) digital assets across different data formats, linguistic changes and over time semantic alterations. The proposed ADDC Platform allows to create, represent and virtually interrelate CH digital content using graph models, where node interconnections lead to conceptually related narratives. Curating and interrelating semantically augmented

digital assets is facilitated using AI (artificial intelligence) algorithms, in addition to human semantic assignment, enhanced by the implementation of a fuzzy semantic indexing system which leads to a common semantic framework capable of modelling digital assets that share affinities, including spatial and temporal information, even in cases of linguistic divergence. Moreover, a multi-modal and user-friendly visual representation of retrieved information serves as a comprehensive means to research and create affinities and associations between facts and data around this newly created, advanced, digital content using semantic representations, assisted by AI tools, across different data formats, linguistic changes and over time semantic alterations.

Description of the TRILOGY ADDC platform of platforms

The TRILOGY ADDC platform of platforms features semantic annotation tools for enriching each digitized asset with a set of descriptive tags, corresponding to a summarization of the conceptual and contextual aspects of the asset. Semantics annotation tools are proposed that comprise not only manual annotation capabilities, but also automatic semantic representation of audio, visual and textual resources, by the application of state-of-the-art AI algorithms refined by domain expert knowledge. TRILOGY tools provide rich socio-historical context in order to add coherence to these cultural assets and their interconnected manifestations. TRILOGY is designed to process vast amounts of multimodal data from either pre-existing datasets or created from scratch. Platform users (professionals, operators, content providers, etc.) are able to enrich platform databases with diverse data in order to be further processed and annotated manually or in an automatic way. Furthermore, automated semantic representation algorithms are incorporated for producing semantic representations for each digitized artefact in a unified way for all resource disciplines (audio, image, text). These representations are supposed to reflect cultural (artistic and scientific) aspects of digital CH assets, evaluated and validated by expert stakeholders in the respective domain fields. Based on the extracted semantic representations, the proposed solution foresees a further construction of a conceptual knowledge graph representing the cultural, technical and conceptual aspects of digital artifacts, by taking into consideration their evolution in the progress of time, along with their potential interrelations. The main goal of the processing is the information extraction/retrieval and the creation of diverse semantic representations depending on the data formats (e.g. the process of semantic annotation differs between music audio, music scores and text corpora). Semantic representations formed by the NLP (Natural Language Processing) tools will yield new relationships in the Graph Database hence increasing the efficiency of querying for interrelated objects. TRILOGY uses graph databases for knowledge representation and indexing of perfect and non-perfect data, a

³<https://www.europeana.eu/>

method capable of handling vast amounts of graph connections, as well as a fuzzy indexing method for fast and efficient updating of the graphs over time in polynomial time, making the ADDC platform capable of handling and reconciling inconsistencies across knowledge graphs, while accommodating data formats for both tangible and intangible CH content.

As stated already, TRILOGY is a platform of platforms that results from the interlinking of the three platforms SYNTHESIS [2], NARRATION (AFIGISSI) [3] and COSMOS [4], alongside a set of tools for semantic annotation and analysis of digitized artefacts (including AI processing tools) and the interplay with the Waygoo geolocation platform [5], functional component in SYNTHESIS. TRILOGY features a back-end system enabling content editing, managing, and sharing; and a front-end platform which will serve inter-associated pieces of information in an appealing and comprehensive visual manner as shown in the preceding graphics. The system exploits semantically tagged resources and incorporate deep learning algorithms with the aim of extracting, from textual, audio, and visual resources, crucial characteristics that distinguish each resource from the others, while also contributing effectively to the creation of a conceptual graph that holds the linking information among each resource item.

Maintaining and verifying the integrity of CH assets and content processed in the platform in the context of creating new AD digital assets, IPRs (Intellectual Property Rights) need to be taken into consideration and preserved in the AD creation process. The aim of TRILOGY is to preserve IPRs associated with processing CH assets through AD using a blockchain tool that will act as a record for holding metadata about digital assets, in a language independent specification (e.g. json) format. This metadata is stored on the blockchain as transactions. The blockchain tool also contains information regarding derivative IPRs that are produced using TRILOGY IPR management tools. Next, we describe briefly each one of the three platforms in that make up TRILOGY platform of platforms. We first give an example

of digital curation to motivate the creation process for AD and the need for TRILOGY as an enabler of ADDC.

The ADDC process description

The ADDC TRILOGY platform makes extensive use of contemporary technologies in order to allow efficient storage, processing and interlinking of multi-modal big data. Prior to describing the three platforms in TRILOGY, we motivate the need for an ADDC platform through a concrete example. We consider an existing paintings museum with a traditional set-up of diverse modern paintings hanging from the walls of the museum in a conventional way. The challenge at hand is to create a digital replica of the museum by organizing the CH content of the museum in a accordance to dynamic narratives that will lead to a multiplicity of means of experiencing the curated collection of digitized paintings both as a personalized virtual museum as well as an AR (Augmented Reality) experience in the physical museum alongside with the paintings themselves. Figure 1 provides an infographic of the example that was described above in order to demonstrate how the proposed ADDC platform can be used to achieve the augmented digitization associated with the museum paintings and the associated digital curation required for achieving immersive and personalizable virtual (in vitro) and augmented (in vitro) reality experiences of the museum content. Figure 1. (a) shows the actual paintings as they are arranged in the actual physical museum today. Figure 1. (b) represents the augmented digitizations of the paintings that consist of high-quality digital replicas of the original paintings alongside with other digital, possibly multimedia, content associated with the paintings. The augmented digitization content depicted in Figure 1. (b) is stored in a graph database and can dynamically be curated in accordance to different narratives, (Figure 1. (c) & 1. (d)) demonstrates the use of the outcome of the ADDC process for an in vivo personalized experience of the museum content using augmented reality and a non-linear visualization.

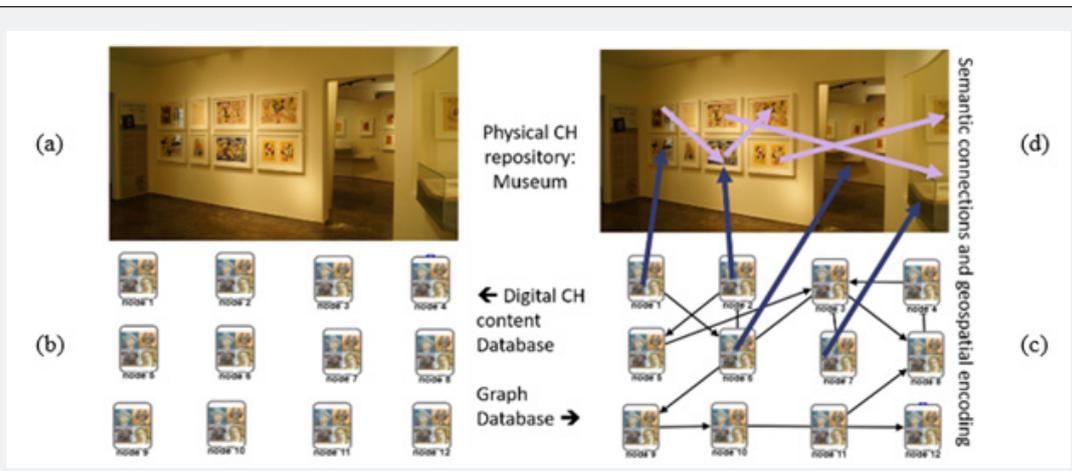


Figure 1: Transitioning from the physical cultural space to the digital cultural space and back to the physical cultural space through augmented digitization and digital curation.

In order to implement the advanced digitization and digital curation (ADDC) outcome, we will discuss next the use on the three platforms in TRILOGY so as to arrive at the desired outcome of the ADDC process. The description of the three platforms follows the sequence that are being used in the ADDC process: SYNTHESIS first, followed by NARRATION (AFIGISSI) and ended with COSMOS for the visualization of the digital curation outcome.

Overview of the SYNTHESIS platform

SYNTHESIS is a platform enabling the creation of interactive and educational 3D immersive environments of the highest

quality, the georeferencing of digital assets in the virtual or physical space, and the navigation in the geocoded digital assets space, either in the form of a virtual tour in a virtual space or physical one depending on the space the digital assets toured are georeferenced to. The SYNTHESIS platform consists of three main components: (a) a set of tools that allow the creation of the 3D virtual space and the associated digital assets; (b) the Waygoo⁴ platform for georeferencing 3D/2D virtual spaces and assets in absolute or relative coordinate systems; and (c) the Waygoo mobile application to experience the virtual tours using iOS and Android smart phones.

Creation and browsing a 3D virtual environment

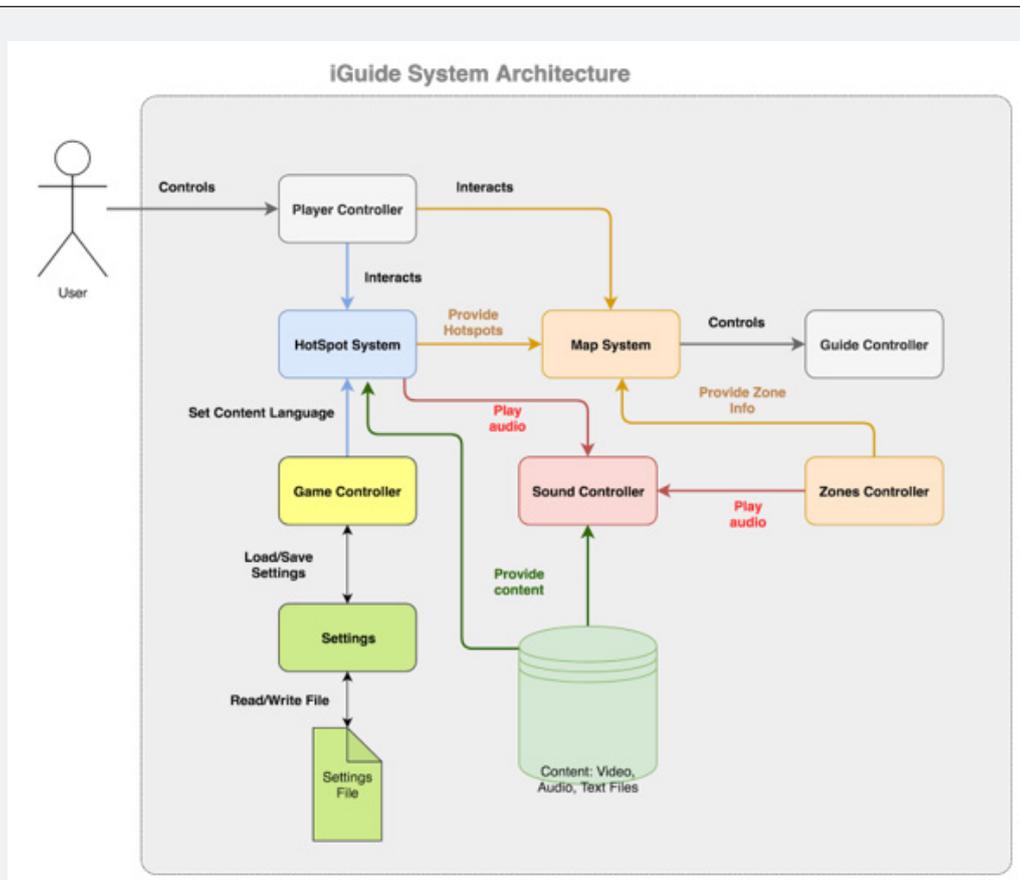


Figure 2: iGuide System Architecture [7].

With the use of Virtual Reality (VR) technologies, the experience of the end user is enriched as a “digital” visitor in places and monuments of cultural and touristic interest. Through a detailed three-dimensional representation/reconstruction of the CH infrastructure, the visitors are guided and navigated in the digital twin of the CH monument and/or in the space of the digital CH assets with no spatial or temporal constraints. The

three-dimensional (3D) digital representations / reconstructions of CH monuments and assets allow the visitor to either browse the space in its current state, or in a three-dimensional model that presents the space reconstructed based on valid data. The virtual space that is constructed seeks to create a digital cultural reference point for the expansion and transmission of cultural elements and infrastructures to a wider audience based on age

⁴Waygoo: <https://claret.gr/waygoo>

criteria and preferences. It combines entertainment and education in an accessible and fun way for wide audiences while developing at the same time an innovative detailed browsing tool. Browsing high-quality digital environments can be achieved either through virtual reality (VR) in the virtual space or through augmented reality (AR) in the physical space. Both ways are supported by the SYNTHESIS platform. The creation and browsing of a 3D virtual environment in the SYNTHESIS platform is based on the iGuide [5] methodology and processes, and include: (i) a 3D graphics back-end subsystem based on a 3D graphics engine, such as Unity, Unreal, etc, that is used to create the necessary the 3D models; (ii) a front-end subsystem that includes a desktop browser interface for geotagging and georeferencing digital assets and content in the 3D model; and (iii) a game engine for gamifying the 3D model and the embedded content as an interactive game that can be experienced using desktops, portable and mobile devices, in a variety of different means including VR/AR (Virtual Reality/Augmented Reality) tours in the virtual and physical space associated with the digital cultural heritage (DCH) assets. The aim of SYNTHESIS is to upgrade the user experience as a visitor to cultural sites and monuments in both virtual and physical spaces. The users have the option of personalized tours of the spaces they choose, including personalized narratives, animations and educational interactive games. The iGuide system architecture block diagram is shown in Figure 2.

The other two components of SYNTHESIS are related to the Waygoo geocoding platform [6].

i. Geolocating digital assets on the 3D physical space with Waygoo^{5,6}

When digital CH assets are associated with physical CH artefacts residing in a physical space that may be accessible as a museum, an archaeological site or a cultural space, it might be desirable to provide visitors of CH spaces with an augmented experience by associating the physical artefacts with their digital replicas (“digital twins”) by geocoding the digital replicas within the same space that their physical counterparts reside. On another circumstance, one may want to curate a collection of digital assets and provide a museum or exhibition-like experience related to a narrative experienced within a physical space. In either case, these digital assets must be associated with a physical space by geocoding them appropriately. In the context of the SYNTHESIS platform, this is achieved through the Waygoo platform.

The Waygoo platform allows to geocode any 2D or 3D physical space on a digital map (in absolute or relative Mercator coordinates). Using the geocoded digital space, the Waygoo platform allows in turn to geocode each digital asset with the

appropriate coordinates the curator wants to associate with (“place it in”) the physical space. This way, the digital assets (content) are embedded in the physical space and become part of it allowing to be experienced in two different ways: (a) either by navigating in the virtual replica of the physical space; or (b) by being projected (the digital assets) in the physical space and retrieved while touring the actual physical space. In either case, the Waygoo platform supports two different automatic navigation modes: one on-demand based on origin-destination demarcation, and another based on a curated (and possibly personalized) narrative involving a succession of preselected geocoded space points according to the story-telling of the curated narrative. The outcome of a Waygoo geocoded and curated collection of digital assets can be experienced as both an iOS and Android mobile application and downloaded from either the AppStore or PlayStore. An example of the archaeological site of the Palace of Phaistos in Crete, Greece, that has been geocoded using Waygoo can be found in [2] and by downloading the Waygoo app from Play and App Stores. Overview of the NARRATION (AFIGISS) platform The NARRATION (AFIGISS) platform is a tool targeted at CH and creative industry professionals for creating, interconnecting, managing, and curating diverse types of digital assets, enabling them to produce narratives and experiences that are presented in an immersive way to end users and audiences through a multiplicity of different user interfaces.

NARRATION (AFIGISS: Greek for NARRATION) [4], [7] is a platform designed to support, manage, curate and promote digital CH content of cultural and tourist infrastructures through the creation of personalized individual and collaborative narratives. The CH content can come entirely from a digital repository, or it can be the “digital twin” of a “physically existing” CH content in cultural and tourist infrastructures, in which case, the platform allows the association and mapping of the digital curated result onto the physical georeferenced space the “physical twin” resides with the aid of the embedded *WayGoo* platform⁷. Access to narratives is possible either during a physical visit or remotely through a virtual visit. NARRATION (AFIGISS) reveals cultural content by implementing new narrative structures by overturning time sequences and by offering a variety of querying options in order to respond to modern perceptions of versatile objects and their multiple interpretations. It allows specialized and non-specialized users to produce narratives based on the available digital content and to share them with targeted audiences in a multitude of experiences. The various steps in the digital curation process supported by NARRATION (AFIGISS) are shown in Figure 3 and are summarized below.

⁵Waygoo: <https://play.google.com/store/apps/details?id=com.demokritos.waygoo&hl=el>

⁶Waygoo: <https://itunes.apple.com/gr/app/waygoo/id1041460914>

⁷See <http://claret.gr/waygoo>

The NARRATION (AFIGISSI) platform⁸ is currently being tested with the 2D CH artifacts (paintings) from the Teriade Museum in Lesvos (<http://www.museumteriade.gr/en/>). The NARRATION (AFIGISSI) process is highlighted in Figures 4, 5 & 6. The visual on Figure 4 demonstrates the concept behind the creation of an advanced digitization outcome in the context of an Enriched Cultural Experience of Analog & Digital Content through Semantic & Topographic Linkage (Teriade Museum Paintings Collection & Digitised Replicas). This conceptual process is digitally encoded in the NARRATION (AFIGISSI) platform as demonstrated in the sequence of the four snapshots of the platform's UI. The sequence of the four visuals in Figure 4 demonstrates the steps of the creation of an advanced digitization outcome through the (manually) supported curation and narration processes of NARRATION (AFIGISSI) and the resulting experience. From top left and

clockwise in Figure 4, describes the asset representation, curation of digital assets, narration creation, and user experience. Figure 5 demonstrates the mapping (transitioning) of a "narration" created in the NARRATION (AFIGISSI) platform from the virtual content graph used for curating the specific content in the NARRATION (AFIGISSI) platform into a physical space that is associated with the content that has been curated in the platform. This mapping is achieved through the Waygoo platform. Figure 6 corresponds to the Visualization of the outcome of processing of the digital content in the NARRATION (AFIGISSI) platform and corresponds to the fourth stage of processing in the platform, namely the User Experience. In the particular example in Figure 6, visualization is achieved by converting the Advanced Digitization outcome created using the ADDC Platform to an AR and VR experience in the left and right segments of the Figure 6 respectively.

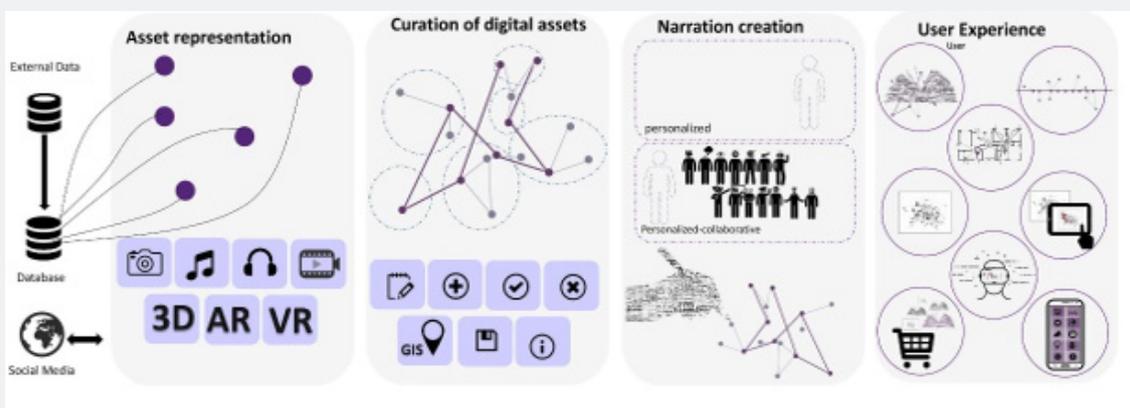


Figure 3: The four stages of processing of digital content in NARRATION (AFIGISSI) include (from left to right): Asset representation, Curation of digital assets, Narration creation, and User Experience.



Figure 4: Narration process: Clockwise: AFIGISSI Portal, Content Curation, Creation of a CH graph (advanced digitization), Experience Creation.

⁸<https://www.afigissi.gr/>



Figure 5: Geocoding NARRATION (AFIGISSI): Georeferencing Digital Narration (Advanced Digitization – left image) to the physical Teriade Museum space is achieved using Waygoo (right image) (<https://www.facebook.com/wayGoo.ISL/>).



Figure 6: Visualisation: Converting an Advanced Digitisation outcome to an AR (left) and VR (right) experience respectively through curation in AFIGISSI.

Overview of the COSMOS platform

The COSMOS platform uses a variety of user interfaces and devices to better promote the understanding of cultural heritage. The AD platform enables creative professionals to retrieve diversified information in innovative and intuitive new ways. Practically, this is achieved by using state-of-the-art graph database software e.g. Neo4j⁹, TigerGraph¹⁰ which allows deep interlinking of digital assets as well as the inclusion of temporal and spatial information. Innovative Natural Language Processing (NLP) methods are used for the extraction of concepts, entities, and topics from song lyrics belonging to significant regional traditions of Europe. The enrichment of interrelated digital objects associated with CH assets will enable the ADDC platform users with more functionalities in exploring and studying cultural

artifacts from these endangered cultural traditions. The ADDC platform exposes a user-friendly interface offering sophisticated and novel query mechanisms that allow multi-modal information retrieval using a variety of different techniques ranging from simple text search to advanced relationship querying. The platform provides support for a wide range of audio and music specific media types (wav, AIFF, mp3, AAC), as well as a broad range of other multimedia formats, including text, images, video, CAD and 3D files.

COSMOS¹¹, as a backbone system for the creation of semantic associations and their contextual visualization, has been developed so far with focus on mythology-oriented stories, characters featured in them and places these stories are set, including the representations of myths in ancient artworks,

⁹<https://neo4j.com>

¹⁰<https://www.tigergraph.com>

¹¹S. C.A Thomopoulos, P. Tsimpiridis, E.-I. Theodorou, & X. Kristopoulou, "COSMOS: Πολιτισμική Ωσμωση – Μυθολογία & Τέχνη (Cultural Osmosis – Mythology & Art)," Proceedings of the 3rd Panhellenic Congress of Digitisation of Cultural Heritage, EUROMED 2019 - Culture, Education, Research, Innovation, Digital Technologies, Tourism, University of West Attica Conference Center - Campus 2, Athens, September 25-27, 2019.

highlighting the correlations with the depicted characters, the places where the art was produced, as well as their present location, offering the opportunity to record the Greek mythology, both in written and visual forms. To that extent, the outcome of COSMOS is, in the context of the definition of AD adopted in this paper, an “advanced digitization” tool for CH digital content related to stories (mythological or not), characters, events, places and associated and/or inspired by artwork. A visualization of the semantic associations in COSMOS is shown in the sequence of frames below. The result of the process constitutes an example of archetype of advanced digitization. In addition to managing contextual associations among stories, characters, events, places and associated and/or inspired by artwork, the tools of COSMOS are being extended in the ADDC platform to cover music, lyrics, photos, videos, 2D/3D animations, as well as intangible CH content. Furthermore, COSMOS is being augmented with a set of linguistic and aural semantic annotation and AI tools, as well as deep learning algorithms, to allow automatic, yet human supervised, semantic association and knowledge discovery, that will allow advanced digitization and the creation of complex advanced digital objects.

Using a variety of Natural Language Processing (NLP) modules for extracting information regarding entities, concepts and topics available in CH content, the COSMOS platform allows automatic, yet supervised, semantic association and knowledge

discovery. Named Entity Recognition (NER) is a critical pre-processing step in NLP for a wide range of information extraction tasks applications. A variety of techniques and tools [9] has been applied for NER, including machine learning-based, rule-based, and hybrid solutions. Neural network architectures that employ character-based word representations, learned from small annotated and large unannotated resources, improve state-of-the-art performance in different languages without making use of language-specific, handcrafted rules or gazetteers. More recently, researchers have achieved state of the art results by enhancing NER-specific neural architectures with pre-trained contextual word representations [10]. Sentiment Analysis has also exploded in recent years mainly because of the importance assigned by businesses to the vast amount of opinionated text generated on social media by their customers towards specific (aspects of) their products [11]. Fine-tuning a large, pre-trained, general-purpose model on a downstream task has recently achieved state of the art results [12]. COSMOS uses the spaCy framework for the model development. spaCy is an open-source, industry-strength NLP framework which is very fast while offering excellent performance (<https://spacy.io/usage/facts-figures>). Furthermore, as of version v2.1, spaCy supports transfer learning methods which accelerate model development by starting from a pretrained model (recently enabled the usage of Google’s BERT pretrained model) (Figures 7 & 8).



Figure 7: From left to right: Digital assets visualization in COSMOS: Left: Digitized myths content interlinked (“Digital Curation”)with respect to key actors in the myths, interlinks within actors, and geographic location of the occurrence of the myths; Center: A particular form of “Advanced Digitization” is created once a specific sequence of myths is selected and “Digitally Curated” according to a key event/god/hero/actor: Three different visualizations of the “Advanced Digitization – Digital Curation” outcome is shown in event space, actors space and location space; Right: The “Advanced Digitization & Digitally Curated” outcome interlinked with other digital assets (objects) in the database.

The integration of the trilogy of the three platforms in one coherent ADDC platform, alongside with the content geocoding WayGoo platform, provide the necessary tools to support, manage, curate and promote digital CH content of cultural and tourist infrastructures through the creation of personalized individual and collaborative narratives. The CH content can emerge entirely from a digital repository, or it can be the “digital twin” of a “physically existing” CH content in cultural and tourist infrastructures. In the latter case, the ADDC platform allows the

association and mapping of the digital curated result onto the physical georeferenced space the “physical twin” resides with the aid of the embedded *WayGoo* platform¹². Access to narratives is possible either during a physical visit or remotely through a “virtual” visit. The ADDC platform reveals the cultural content by implementing new narrative structures by overturning time sequences and by offering a variety of querying options in order to respond to modern perceptions of versatile objects and their multiple interpretations. It allows specialized and non-

specialized users to produce narratives based on the available digital content and to share them with targeted audiences in a multitude of experiences. The AD Back - End platform consists of service and storage layers. The service layer provides API's to be consumed from Front - End applications such as web and mobile applications. The storage layer is responsible for interfacing with the various data stores. Graph Databases will be used as a data store in the platform as they are well suited for big data storage along with interrelation information between digital artefacts while ensuring querying performance for data retrieval. These

properties fit the goals of the AD platform and provide a good foundation upon which the platform can expand new functionality. This technology is already part of the AFIGISSI platform and will be further enhanced in the AD Platform with a number of features such as: multiple new relationship types between tangible or intangible CH digital artefacts, temporal information storage along with object history and novel querying capabilities through advanced relationship-based information retrieval. Semantic interconnection of graphs (i.e. outcomes of advanced digitization) is one of the research issues beyond SoA that will be investigated.

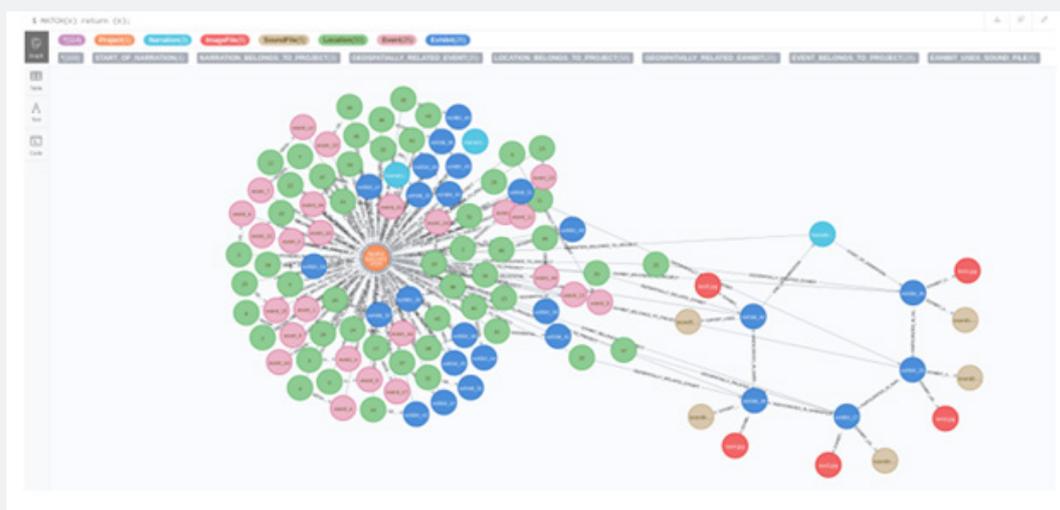


Figure 8: The AD Platform Graph database visualization. The nodes represent digital assets along with their annotations (including semantics), whereas the links represent contextual and narrational interlinks. The entire graph or part of it represent the outcome of advanced digitization in accordance to the definition of AD given in this paper. The visual is courtesy of NARRATION (AFIGISSI).

Blockchain Tool

The ADDC platform benefits from using a Distributed Ledger Technology (DLT), also known as blockchain. The main role of blockchain in the context of the AD platform will be to track the usage of CH artifacts that the platform manages. In general, Blockchain allows the creation and the maintenance of an immutable transaction ledger which the nodes of a distributed network share, maintain, and update by committing transactions. There are two types of DLTs, permissionless and permissioned DLTs. Permissionless DLTs allow anyone with an internet connection to join them and participate in their management. Permissioned DLTs on the other hand, allow only authorized entities to join them. There are two subcategories of permissioned DLTs. The first subcategory is consortium DLT that allows a set of authorized entities to manage and participate in the blockchain network. The second subcategories is private DLT, their only difference from consortium blockchains lies in the fact that a single entity manages the DLT network.

As mentioned earlier, in the context of the ADDC Platform, the DLT will act as means to verify the integrity of digital CH artifacts.

Maintaining and verifying the integrity of a CH digital assets is of the essence since the purpose of the platform is preserving digital artifacts. For example, in the case of TM digital assets, blockchain will act as a record that will hold metadata about each traditional musical artifact in a language independent specification (e.g. json) format. This metadata will be stored on the blockchain as transactions. Metadata will include but be not limited to; name of the composer, author, year, duration, country and genre. Every time a third party wishes to create derivative digital content using a musical artifact managed by the ADDC Platform, the following sequence of actions will occur. First, the ADDC Platform will record an entry that correlates the musical artifact and its metadata to the derivative content. Then the blockchain will store that entry as a transaction to the blockchain. Similar to the case of metadata storage these transactions will record the entry in a language independent specification format. In particular, the blockchain will record information such as the identity of the third party that wishes to use the digital asset, details about the derivative digital content. As aforementioned, it is one of the ADDC platform goals, and consequently one of the DLT's as well, to preserve the integrity of endangered music artifacts. To contribute in achieving

¹²See <http://claret.gr/waygoo>

this goal the blockchain will hold a checksum of the digital artifact. This will facilitate the detection of accidental corruption or unauthorized alterations of the digital artifact. Moreover, the blockchain tool will store metadata with regards to the editing of nodes of information. In particular, this metadata will contain information about who edited which node, timestamp and duration of editing as well as the nature of the added content. Additionally, it will contain information about each information node in the ADDC Platform.

Intelligent Semantic Indexing and Enabling Technologies

i. Using and indexing Non – Perfect data at knowledge representation using graph databases

Imperfect data express their meanings incompletely and the Theory of Fuzzy Sets arises mathematical support to the interpretation of those data. The union of these concepts describes a new data type, called fuzzy data. Nevertheless, the use of fuzzy data in Graph Databases is not allowed and although we can define fuzzy queries on Graph Databases the data stored inside has to be regular and perfect. A potential solution would be the extension of the Graph Database concept for allowing the storage of fuzzy data, however such an extension would violate one of the key properties of Graph databases, that integrity constraints enforce data consistency. An indirect way of using fuzzy data without compromising one of the basic features of graph database would be the introduction of a “parallel structure” for storing the incomplete pieces of information considering that a piece of information to be stored, may be crisp (perfect), totally fuzzy, or partially crisp (partially fuzzy), but without isolating this structure from the graph database itself. In this case the database indexing which is the primary data retrieval mechanism for effective data retrieval should be common to both structures however it would violate the database constraints and therefore the integrity constraints which enforce data consistency. In order to override these limitations, we introduce a fuzzy graph database indexing mechanism for bridging these two structures. As a matter of fact, there is a single storage space where all data is being stored, however through fuzzy indexing, the two virtual storages (perfect and imperfect storages) can be separated quickly and efficiently in just one index scan. K-D trees is a very efficient semantic indexing methodology that could be also parallelized (using spatial decomposition) [13] for exploiting the power of today’s multi – kernel CRVUs or even by performing GPU based massively parallelization. Using some special purpose KD-tree implementations, some restrictions regarding key violations such as duplicate keys [14] could be overcome whereas some structural defects of standard KD-tree indexes such as scalability [15] and updatability [16-20] can be improved, Figure 9. The TRILOGY ADDC platform design methodology for the creation and curation of AD CH assets is modular, yet evolutionary and integrated, so as to best demonstrate: (a) the capabilities of the proposed AD platform to create advanced digitization content by contextual

interlinking digital CH artefacts with the know-how of a CH expert; (b) the ability of the proposed AD platform AI tools to generate advanced digitization content automatically through the discovery of semantic similarities across semantically annotated digital objects from digital content databases and the *semantic web* (the resulting advanced digitization outcome shall be subjected to the validation by a CH expert for its final acceptance as an augmented digital CH artefact); (c) the ability of the AD Platform to manage the outcome of advanced digitization, maintain it, update it over time and use it to create new advanced digitization outcomes; and (d) the ability of the AD Platform to curate advanced digitizations into narratives that lead to experiences that can benefit CH institutions, the creative industries and CH experience tourism.

Conclusion

Digital curation technologies contribution to CH study and promotion has brought forth significant results in creating efficient methods of digitizing multi-disciplinary resources including textual descriptions, multimedia assets and 3D reconstructions. Moreover, advanced information processing and visualization techniques have been tested in order to detect important cultural aspects, enhance the available resources and apply sophisticated representation methods in order to effectively increase public awareness about cultural heritage, including. The proposed TRILOGY ADDC platform advances the state of the art by developing and implementing a fuzzy semantic indexing system leading to a common semantic framework capable of modelling the semantically similar digital assets, including spatial and temporal information, even in cases where these are linguistic divergent. In this regard, the proposed ADDC platform advances the studying and understanding of cultural assets, characterized by linguistic diverges and differences. Furthermore, TRILOGY implements multi-modal semantic analysis with the capability of processing textual and audio textual assets and extracting common disciplines, semantic representation of the available resources, providing the capability to infer potential relations, affinities, similarities and differences among them. In this regard, TRILOGY employs new methods for aural and symbolic musical data analysis using advanced Machine Learning/AI techniques. Audio facts, be it aural or symbolic (e.g. transcriptions), along with textual facts (lyrics, background stories, comments, translations), related photographic material of varying content, spatial facts, instrument blueprints, and others, forming sets of tangible digital information, will be analyzed and correlated in a comprehensive manner resulting in both the preservation and augmentation of our understanding of cultural history. By incorporating advanced curation capabilities, TRILOGY promotes multi-faceted CH; not just as a static curation flow, but as dynamic storylines and narratives which can be modified or enriched over time, exploiting assets that can be used in several story flows based on their semantic representations. Tangible and intangible digital information will be effectively interrelated through an intuitive user-friendly interface, promoting both the preservation and augmentation of

our understanding of the worldwide cultural history through the TRILOGY ADDC case studies. Moreover, TRILOGY leverages the implementation of a fuzzy semantic indexing system providing querying tolerance in cases of dispersed or destroyed resources in order to create ADDC outcomes. Finally, the TRILOGY ADDC platform extends the state of the art by offering sophisticated and novel query mechanisms that allow multi-modal information retrieval by using a variety of different techniques ranging from simple text search to advanced relationship querying. Incorporating a user - friendly interface will increase the Platform's usability and accessibility. TRILOGY supports a wide range of audio and music specific media types (wav, AIFF, mp3, AAC), as well as a broad range of other multimedia formats, including text, images, video, CAD and 3D files, making it possible to bring into contact various types and formats of content.

Acknowledgements

The three platforms described in this paper have been supported by a number of EU- and nationally funded projects, namely:

i. SYNTELEISIS "Innovative Technologies and Applications based on the Internet of Things (IoT) and the Cloud Computing" with grant number MIS 5002521/ 2016-2020.

ii. AFIGISSI: *Narration* is co-funded by Greece and the European Union- European Regional Development Fund (ERDF) within the operational program Competitiveness-Entrepreneurship-Innovation (EPAnEK), <https://www.afigissi.gr/>, and

iii. COSMOS is co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship & Innovation (EPANEK), in the framework of the RESEARCH – CREATE – INNOVATE Action ((project code T1EDK-04283), <https://cosmosapp.gr/>.

References

- Stelios CA Thomopoulos, Adam Doulgerakis, Maria Bessa, Konstantinos Dimitros, Giorgos Farazis, et al. (2016) DICE: Digital Immersive Cultural Environment. Springer, Cham, EUROMED 2016: Euro-Mediterranean Conference, pp. 758-777.
- S CA Thomopoulos, Farazis G, Scordili C, Kassianou K, Thomopoulou CP, Zacharakis D (2020) SYNTHESIS: A Platform for Generating Digital Cultural Heritage Assets and Creating Virtual or Physical Tours Through Geocoding. *Global Journal of Archaeology & Anthropology* Juniper Publishers, 11(4).
- Thomopoulos S CA, Tsimpiridis P, Theodorou E-I, Xristopoulou X (2019) COSMOS: Πολιτισμική Όσμωση – Μυθολογία & Τέχνη (Cultural Osmosis – Mythology & Art) Proceedings of the 3rd Panhellenic Congress of Digitisation of Cultural Heritage, EUROMED 2019 - Culture, Education, Research, Innovation, Digital Technologies, Tourism, University of West Attica Conference Center - Campus 2, Athens, pp. 25-27.
- Thomopoulos S CA, ΑΦΗΓΗΣΗ: Ολοκληρωμένο σύστημα διαχείρισης και επιμέλειας ψηφιακού περιεχομένου και παραγωγής προσωποποιημένων ατομικών και συνεργατικών αφηγήσεων (NARRATION: An integrated system of managing and editing digital content and producing personalized and collaborative narratives), Proceedings of the 3rd Panhellenic Congress of Digitisation of Cultural Heritage, EUROMED 2019 - Culture, Education, Research, Innovation, Digital Technologies, Tourism, University of West Attica Conference Center - Campus 2, Athens, September 25-27, 2019.
- George Farazis, Christina Thomopoulos, Christos Bourantas, Sofia Mitsigkola, Stelios CA Thomopoulos (2019) Digital approaches for public outreach in cultural heritage: The case study of iGuide Knossos and Ariadne's Journey Digital Applications in Archaeology and Cultural Heritage, Elsevier Publ, 15(e00126).
- Waygoo: <https://claret.gr/waygoo>. Also, waygoo Android application downloadable from the Play Store and waygoo iOS application downloadable from the App Store.
- Stelios CA Thomopoulos, Konstantinos Dimitros, Konstantinos Panou, Giorgos Farazis, Sofia Mitsigkola, et al. (2020) Narration: Integrated system for management and curation of digital content and production of personalized and collaborative narratives, EUROMED 2020 Conference, 2-5.
- Stelios CA Thomopoulos, Panagiotis Tsimpiridis, Eleni-Iso Theodorou, Christos Maroglou, Efstathios Georgiou, et al. (2020) COSMOS: Cultural Osmosis – Mythology & Art A data organization and visualization platform, with the use of AI algorithms," EUROMED 2020 Conference, 2-5.
- Nadeau D, Sekine S (2007) A survey of named entity recognition and classification.
- Straková J, Straka M, Hajic J (2019) Neural Architectures for Nested NER through Linearization.
- Liu B (2012) Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies.
- Yang Z, Dai Z, Yang Y, Carbonell JG, Salakhutdinov R, et al. (2019) XLNet: Generalized Autoregressive Pre-training for Language Understanding. NeurIPS.
- Flora Amato, Vincenzo Moscato, Francesco Gargiulo, Fabio Persia: AIPIS: Parallel Semantic Indexing System - Preliminary Experiments Algorithms and Architectures for Parallel Processing: 13th International Conference, ICA3PP 2013, Vietri sul Mare, Italy, December 18-20, 2013, Proceedings, Part II (pp.133-140)
- Meenakshi, Sumeet (2018) Gill k-dLst Tree: k-d Tree with Linked List to Handle Duplicate Keys. *Emerging Trends in Expert Applications and Security*, Springer, pp.167-175.
- Bkd-Tree: A Dynamic Scalable kd-Tree: Octavian Procopiuc, Pankaj K Agarwal, Department of Computer Science, Duke University, United States.
- Amalia Duch, Vladimir Estivill-Castro, Conrado Martinez (1998) Randomized k-dimensional binary search trees. In Chwa K-Y and Ibarra OH, (eds.), Proc. of the 9th Int. Symp. on Algorithms and Computation (ISAAC), volume 1533 of Lecture Notes in Computer Science, pages. Springer-Verlag, pp. 199–208.
- Fuzzy Classification based on adaptive networks and genetic algorithms Sun CT, Jang JS, Genetic Algorithms and Fuzzy Logic Systems: Soft Computing Perspectives.
- Learning fuzzy hyper-rectangles with instance and neural based methods Figueira LB, FOS Sa Lisboa, M do Carmo Nicoletti (2004) IEEE Annual Meeting of the Fuzzy Information, Processing NAFIPS '04.
- <https://en.wikipedia.org/wiki/Druze>
- <https://janusgraph.org/>



This work is licensed under Creative Commons Attribution 4.0 License
DOI: [10.19080/GJAA.2020.11.555820](https://doi.org/10.19080/GJAA.2020.11.555820)

Your next submission with Juniper Publishers will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
(Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission

<https://juniperpublishers.com/online-submission.php>