Zenith Outdoor/Indoor Experience

May 2025 – April 2027 (24M)

Project 2024-1-CY01-KA220-ADU-000256958

D3.2 Guided AR Tour

Authors	TUW
Abstract	This deliverable documents the implementation of the guided outdoor AR tour, the central prototype of work package 3. Building on the technical core established in A3.1 and the spatial calibration implemented in A3.3, it brings location-based cultural-heritage content, georeferenced visualisation and in-situ interaction together into a single, deployable application for the Apple Vision Pro. Using the authoring tool developed in work package 4, we created location-based content for Vienna covering three guided-tour use cases: interaction with a large landmark, proximity-based interaction with small monuments, and interaction with the facades around a town square. The cultural experience logic was designed as a sequence of points of interest and stored in the database of the project platform, which forms part of the same authoring tool. These use cases have already undergone mock-up testing, and we guided AUTH in providing comparable content for Thessaloniki to ensure a seamless cross-city experience. All use cases for both cities were implemented in Unity by extending the core of A3.1 and embedding the calibration of A3.3. The application dynamically retrieves cultural-heritage content from the project database through dedicated API endpoints, places it accurately in the real

	environment using the calibration package, and reveals it through selection- and proximity-based interactions. It also includes the calibration procedure users have to perform at the very beginning of an experience. Upon successful calibration, they are free to move around and can interact with the environment and retrieve different types of accurately georeferenced content (text, pictures, video). Gamification elements are prepared on the project platform and will be implemented as part of the user interaction task.

THE CONSORTIUM

No.	Abbreviation	Partner Name	Logo
1	**GEO**	Geoimaging ltd	
2	**AUTH**	Aristotle University of Thessaloniki	
3	**TUW**	Technische Universität Wien	
4	**VIR**	Virtual Solutions	

DISCLAIMER

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the

European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Co-funded by the European Union

Co-funded by the European Union

CONTENTS

The consortium 2
Disclaimer 2
Contents 4
Glossary 5
1. Introduction 6
2. Conceptual Framework of Guided AR Tour 6
2.1 Use Case Types 7
2.1.1 Large Landmark 7
2.1.2 Small monuments on a square 7
2.1.3 Facades around a Town Square 7
2.2 Content / Media 8
2.2.1 Text 8
2.2.2 Audio 8
2.2.3 Image 8
2.3 Level-of-Detail / Hierarchical Content Structure 8
2.4 Gamification elements 9
3. Preparation of Guided AR Tour 9
3.1 Guided tour locations / sites 9
3.1.1 Large Landmark: Karlskirche 9
3.1.2 Small monuments on a square: Resselpark 10
3.1.3 Facades around a town square: Graben 11
3.1.4 Use cases in Thessaloniki, Greece 12
3.2 Content acquisition 14
3.3 Experience Setup on Authoring Tool 14
4. Implementation of Guided AR Tour 15
4.1 Before the tour 16
4.1.1 Handling of POIs 18
4.1.2 Handling of Media 18
4.1.3 ExperienceManager script 18
4.2 During the tour 18
4.2.1 Selection of POIs 19

4.2.2 Proximity-based interaction 19
4.2.3 POI prefabs 20
5. Summary and Conclusions 21

GLOSSARY

A	Activity
API	Application Programming Interface
AR	Augmented Reality
AUTH	Aristotle University of Thessaloniki
CC BY 4.0	Creative Commons Attribution 4.0 International licence
CRS	Coordinate Reference System
DLL	Dynamic Link Library
GDPR	General Data Protection Regulation
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GRF	Global Reference Frame
HMD	Head-Mounted Display
HTTP	Hypertext Transfer Protocol
JSON	JavaScript Object Notation
LRF	Local Reference Frame
POI	Point of Interest
UI	User Interface
WP	Work Package

1. INTRODUCTION

This deliverable is part of work package 3, which bundles the outdoor AR implementation activities of the ZOE project. This activity (A3.2) builds directly on two preceding deliverables of this work package: the technical infrastructure and core implementation established in A3.1, and the spatial calibration of the Apple Vision Pro implemented in A3.3. Whereas those activities prepared the framework, hardware, software and georeferencing required for outdoor AR, the present activity uses them to realise a complete, content-filled guided AR tour prototype.

The purpose of this deliverable is to document the conceptual design and technical implementation of the guided AR tour. It describes how the three guided-tour use cases identified in earlier work packages are translated into location-based cultural-heritage content, how this content is authored and stored, and how the Unity application loads, georeferences and reveals it on the Apple Vision Pro. As with the other deliverables of this work package, the content herein reflects the collaborative efforts of the project consortium and contributes to the overall goal of promoting accessible, inclusive and technology-driven lifelong learning experiences.

This document is structured as follows. Section 2 presents the conceptual framework of the guided AR tour, describing the three use-case types, the supported media types, the hierarchical level-of-detail structure of the content, and the gamification elements foreseen for the final prototype. Section 3 documents the preparation of the guided tour, including the selection of the heritage sites, the acquisition of content, and the setup of each experience on the project authoring tool. Section 4 describes the technical implementation in the Unity application, extending the core of A3.1 and embedding the calibration of A3.3, and is organised into the steps performed before and during the tour. Finally, Section 5 summarises the outcomes and highlights how they enable further activities within and beyond this work package.

2. CONCEPTUAL FRAMEWORK OF GUIDED AR TOUR

The guided AR tour is built around three different types of use case. Although all three rely on the same core functionality, they place different requirements on the experience regarding the position of the user, the position and elevation of the

content, and the way the user interacts with it. These use-case types were derived together with all project partners and were tested both in the requirements questionnaire and in subsequent mock-up tests (WP2, activities 2-4). Within each use case, the content can be enriched with different media types and is organised in a hierarchical level-of-detail structure. Gamification elements are prepared on the project platform and are planned for the final prototype.

2.1 Use Case Types

As outlined below, the three use-case types differ mainly in how the user is positioned relative to the points of interest (POIs), how high the POIs are placed, and whether the user can see all of them simultaneously.

2.1.1 Large Landmark

In the large-landmark use case, the user stands at some distance from a single, prominent structure that can be overlooked as a whole, such as a church or a monument. The individual sub-elements of the landmark – for example a tower, a portal or a sculpture – are represented as separate POIs. As these elements are distributed across a tall structure, the corresponding POIs are potentially elevated above the ground.

2.1.2 Small monuments on a square

In the small-monuments use case, several smaller objects – typically statues – are distributed across an open space such as a park or a square. The user might not be able to see all of them at the same time and is expected to walk between them. The POIs are usually placed between zero and three metres in height. Each monument is represented by a single top-level POI, to which different media can be attached.

2.1.3 Facades around a Town Square

In the facades use case, the user is surrounded by the building facades that enclose a town square; here the facades themselves are the POIs. Their annotations are placed at a height of roughly three to eight metres, corresponding to the visible parts of the buildings. As the user typically stands in the middle of the square, all facades are most likely visible at the same time.

2.2 Content / Media

Each POI can be filled with different media types so that digital content can be experienced in situ, next to the real cultural-heritage object. Three media types are supported: text, audio and images. A single POI, or a single monument, may combine several of them.

2.2.1 Text

Textual content carries the descriptive, "hidden" knowledge about a POI. It is displayed in a content box that appears below the POI icon when the content is revealed. The size of the text is adjusted such that it can easily be read in the environment.

2.2.2 Audio

Audio content is streamed to the POI and played back when the POI is revealed or selected. It is used both for narration and for characteristic sounds – for example, the peal of a church bell or a piece of music associated with a building (or a former or current resident of it).

2.2.3 Image

Image content is downloaded and shown on a placeholder surface inside the POI content box. Images are typically historical photographs or depictions that reveal how a site looked in the past.

2.3 Level-of-Detail / Hierarchical Content Structure

To avoid overloading the scene and overstimulating the user, content is organised in a hierarchical level-of-detail structure: a top-level POI can reveal one or more sub-POIs, which can in turn reveal further sub-sub-POIs. At the highest level the user only sees a virtual icon marking a real-world POI; the more detailed content of the lower layers is revealed on demand. A POI reveals its children in one of two ways: either by selection (the user looks at the POI and confirms it) or by proximity (the user approaches the POI). This structure lets the tour designer guide the depth and order in which information is discovered.

2.4 Gamification elements

A subset of the gamification mechanics offered by the indoor 3D platform that is relevant and plausible for outdoor environments is foreseen for the guided AR tour and will be implemented as we finalize the interaction module (A3.4). The gamification submodule and its mechanics are defined by the project partner VIR as part of the authoring platform and are documented in deliverable D4.1. For the outdoor AR tour these mechanics include searching for hidden POIs in a scene, collecting georeferenced objects, displaying a progress indicator of the game status, and awarding points for specific interactions with AR objects. The proximity-aware and selection-based interactions implemented in this activity already provide the technical basis on which these elements will be built.

3. PREPARATION OF GUIDED AR TOUR

Preparing a guided AR tour involves three steps: selecting the heritage sites, acquiring the content to be presented at each site, and setting up the corresponding experience on the ZOE authoring tool. These steps are described in the following subsections.

3.1 Guided tour locations / sites

For the first prototype, six sites were prepared across two cities, with one site for each use-case type per city. The three Vienna sites were selected and filled with content by TUW and are described in detail below. The three Thessaloniki sites were selected and filled with comparable content by AUTH, ensuring a seamless cross-city experience, and are summarised more briefly.

3.1.1 Large Landmark: Karlskirche

The Karlskirche (St. Charles's Church) is a Baroque church on Karlsplatz in the fourth district of Vienna, built between 1716 and 1723 by Johann Bernhard Fischer von Erlach and completed by his son Joseph. Its front side is fully visible from a distance of about 80 metres, which makes it a representative large-landmark use case. Three top-level POIs were defined: the central information POI on the church itself, the left triumphal column, and the right (southwest) tower that houses the Karlsglocke bell. These POIs reveal sub-content such as historical photographs from the Wien Museum collection,

explanatory texts on Fischer von Erlach and on the reliefs of the triumphal columns, and an audio clip of the church bell.

Figure 1: Karlskirche from the user's perspective on the Apple Vision Pro.

3.1.2 Small monuments on a square: Resselpark

Resselpark contains several statues of Austrian inventors of the past. Two statues were selected as top-level POIs: Josef Ressel, the forester and inventor best known for the ship propeller, and Josef Madersperger, the inventor of an early sewing machine. Each statue is represented by one POI that reveals an image, an audio narration and an explanatory text about the respective inventor.

Figure 2: Resselpark with the statue of the inventor of the ship propeller, Josef Ressel, from the user's perspective on the Apple Vision Pro.

3.1.3 Facades around a town square: Graben

The Graben is a busy pedestrian street and square in the historic first district of Vienna, lined with historic and architecturally significant buildings. Three top-level POIs were selected: the building at Graben 14–15 (the "Grabenhof", built in 1874–1876 by Otto Wagner and Otto Thienemann), the building at Graben 26–28, and the Plague Column (Pestsäule, also called the Trinity Column). The POIs reveal texts on the architecture and history of the buildings, historical images, and an audio clip of Schubert's Serenade, who was a frequent visitor of the pub 'Zur Schnecke' which used to be located at Graben 15.

Figure 3: Graben from the user's perspective on the Apple Vision Pro.

3.1.4 Use cases in Thessaloniki, Greece

Three comparable sites were prepared for Thessaloniki by AUTH, one per use-case type. Kamara, the Arch of Galerius, is a large Roman triumphal arch and serves as the large-landmark use case. The Church of the Transfiguration of Christ provides the facades use case. The Makedonomahi monument, which commemorates the fighters of the Macedonian Struggle, provides the small-monuments use case, with individual POIs dedicated to the historical figures it honours.

Co-funded by the European Union

Figure 4: Kamara from the user's perspective on the Apple Vision Pro.

Figure 5: Makedonomahi from the user's perspective on the Apple Vision Pro.

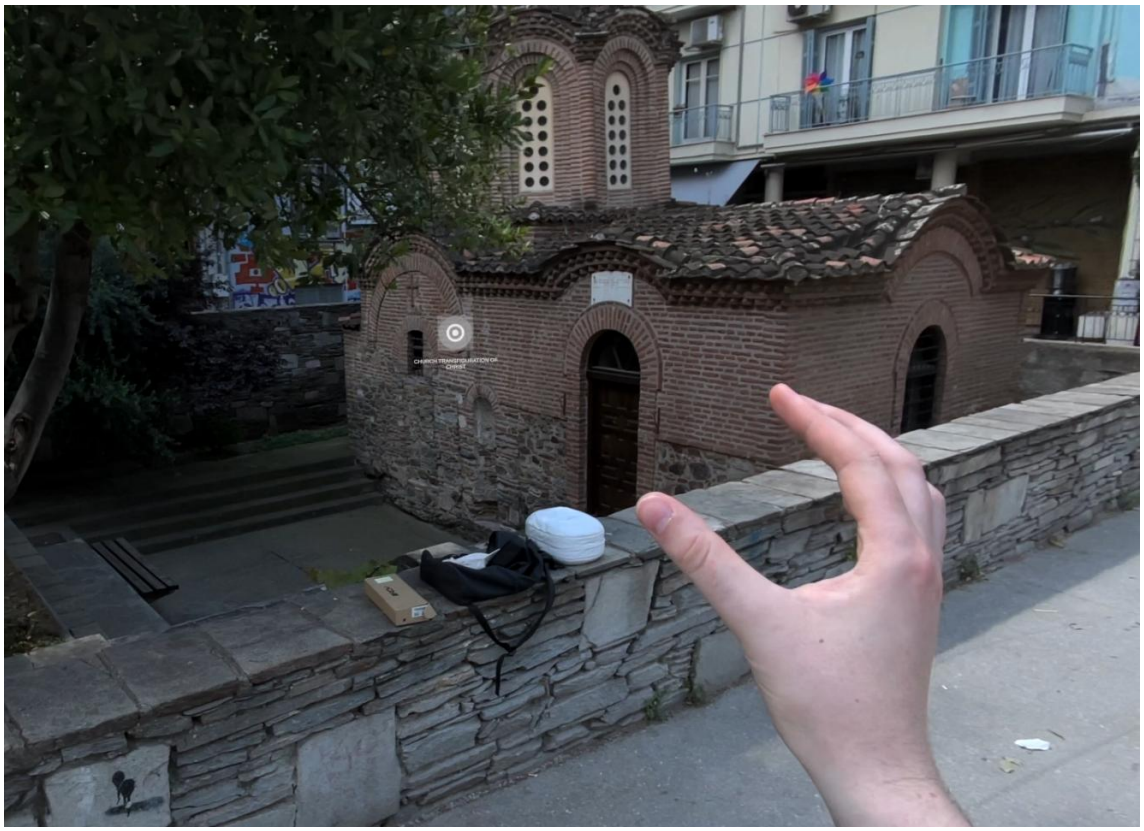

Figure 6: Church Transfiguration of Christ from the user's perspective with a POI on the Apple Vision Pro.

3.2 Content acquisition

The content focuses on "hidden" cultural-heritage knowledge that is not immediately apparent from the physical site and that can be revealed to the user through the AR tour. It was compiled from publicly available sources, such as the online collection of the Wien Museum (Vienna History Museum). Only material that may be freely reused was included; image content is released under the Creative Commons Attribution 4.0 International licence (CC BY 4.0) or is in the public domain, and the respective sources are stated within the experience itself.

3.3 Experience Setup on Authoring Tool

Each guided tour is set up on the ZOE authoring tool, the web-based platform developed by the project partner VIR and documented in deliverable D4.1. The workflow is the same for every site: a new experience is created from a base experience; all POIs are added and positioned (see Figure 7 for an example); for each

POI the media types are selected and the media are uploaded; the OnSelect actions of the main POIs are defined, linking them to their sub-POIs; and the proximity values of the main POIs are set. Once an experience has been saved, it is stored in the project database and can be retrieved by the Unity application through a dedicated API endpoint, addressed by the experience identifier in its URL.

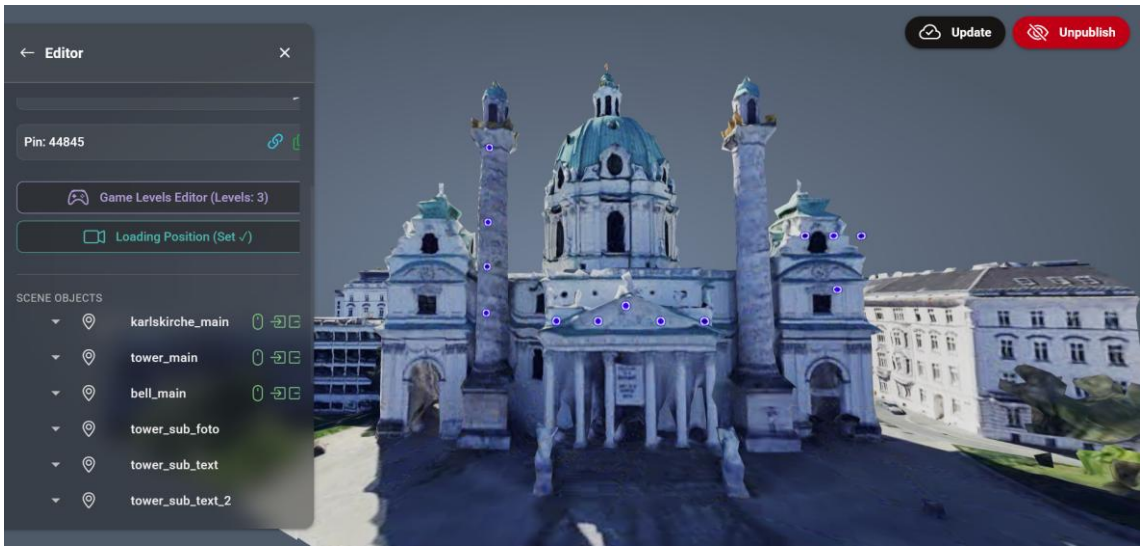

Figure 7: Set-up of the Karlskirche experience on the ZOE platform (WP4). POIs (represented as blue dots) are positioned and automatically geo-referenced.

4. Implementation of Guided AR Tour

The guided AR tour was implemented using the Unity game engine (Version 6.3 LTS) for VisionOS development by extending the core implementation of A3.1 and embedding the spatial calibration of A3.3. The application logic is organised into two phases: the steps performed before the tour – selecting and loading an experience and calibrating the device – and the behaviour during the tour, when POIs are revealed and the user interacts with them. The implementation rests on two main building blocks: the ExperienceManager script, which orchestrates the overall flow, and the POI prefabs, which represent the individual points of interest. Both are described below. Figure 8 gives a visual impression of what the POIs look like from a user perspective and how different media types are visually distinguished.

Figure 8: All POIs presented at their real-world locations on the Karlskirche from the user's perspective on the Apple Vision Pro. The different icon types reflect different media types.

4.1 Before the tour

A single application serves all experiences of both cities. On the start screen the user first chooses a location (Vienna or Thessaloniki) and then a specific site. Selecting a site triggers an API call that retrieves the corresponding experience from the project database, based on a per-experience identifier appended to the data endpoint.

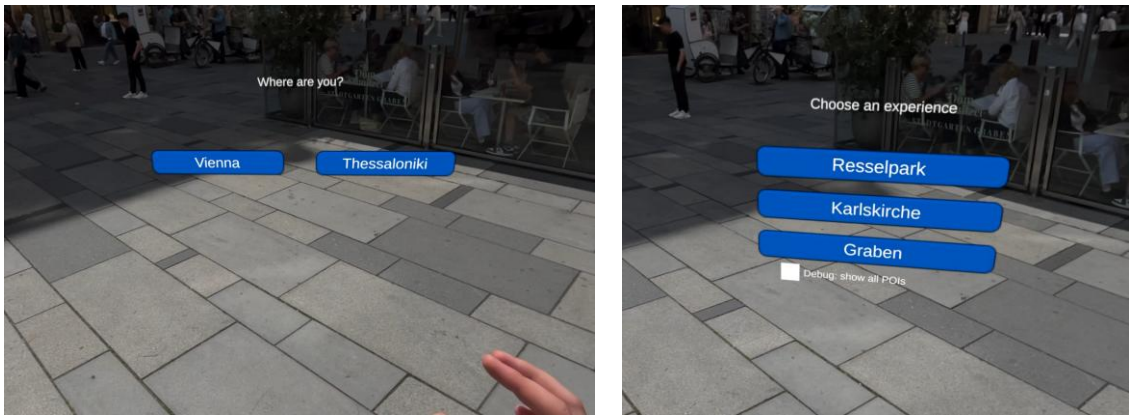

Figure 9: Experience selection dialogue on the Apple Vision Pro from the user's perspective in sequential order. Left: Users can choose a city. Right: The three experiences defined for Vianna. If Thessaloniki is selected, the corresponding experiences are shown instead.

The experience is then loaded in the background – the POIs are created but kept hidden and their media are downloaded – while the user is handed over to the spatial calibration dialogue implemented in A3.3. Once the calibration is complete, the POIs are moved to their real-world positions and the user is offered a "Start guided tour" screen that activates the experience.

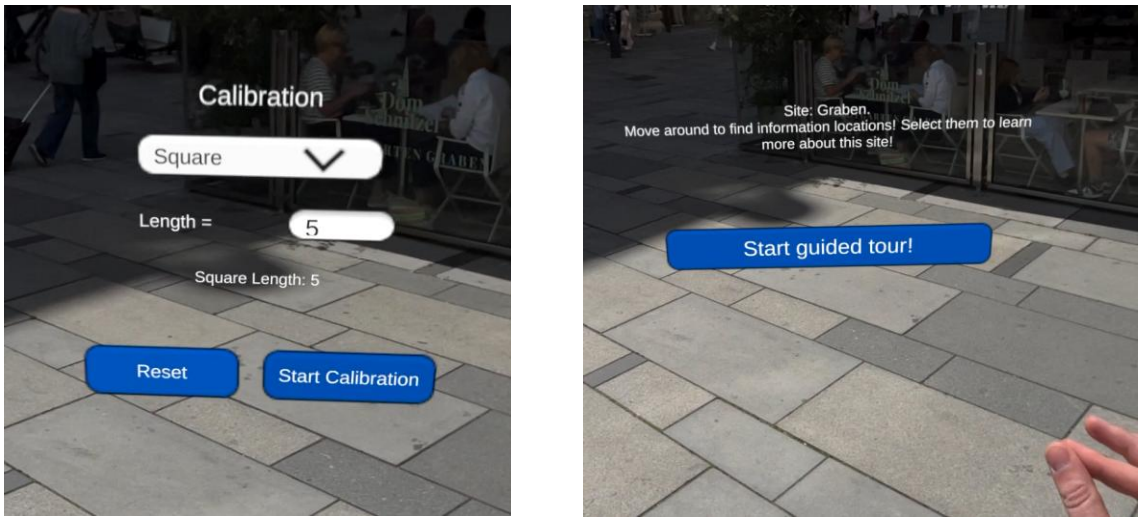

Figure 10: Calibration (left) and Start tour (right) dialogue on the Apple Vision Pro from a user's perspective in sequential order.

4.1.1 Handling of POIs

Each POI carries its geographic coordinates in the coordinate reference system of the experience. When the calibration is complete, every POI is relocated to its real-world position by applying a Helmert transformation (see D3.3), which converts the georeferenced coordinates into the local coordinate system of the device. The hierarchical structure described in Section 2.3 is encoded directly in the loaded data: each POI knows which sub-POIs it reveals on selection and which of these it reveals based on proximity. As required by the Unity PolySpatial architecture, every POI is placed on the "Unbounded" layer so that it is rendered in the unbounded volume on the device, and every POI is spawned hidden until the tour is activated.

4.1.2 Handling of Media

The media attached to the POIs are loaded asynchronously while the user calibrates, so that the experience is ready by the time the tour starts and the application never blocks. Images are applied to a placeholder surface inside the POI content box. Audio clips are assigned to an audio source on the POI, ready to be played back. Text is inserted into a text field within the POI content box. Relative media paths are resolved against a configurable base media URL, and all media downloads run concurrently to minimise loading time.

4.1.3 ExperienceManager script

The ExperienceManager is the central orchestrator of the guided tour. In short, it ties together the experience loading, the calibration of A3.3, and the runtime behaviour of the POIs. It keeps track of the selected location and site, resolves the correct, local coordinate reference system (Vienna: EPSG31256 and Thessaloniki: UTM34N) and experience identifier, and starts the asynchronous loading of the experience. It then opens the calibration dialogue and listens for its completion; once the device is calibrated, it relocates all POIs to their real-world positions and presents the "Start guided tour" screen. Finally, it activates the experience, after which the per-POI logic described below takes over.

4.2 During the tour

Once the tour is active, each top-level POI continuously evaluates the 2D Euclidean distance between itself and the user; this is enabled by using local projections of

coordinates. This distance drives the proximity-based reveal described below. Both ways in which the user can reveal and interact with content – selection and proximity – are described in the following subsections.

4.2.1 Selection of POIs

Selection reuses the Gaze&Pinch interaction pattern that was implemented for the UI dialogues in A3.1 and extends it from flat UI elements to georeferenced 3D objects. Each POI is a 3D prefab that carries an interactable component and a collider that registers Gaze&Pinch events; the user looks at a POI icon and confirms it with a pinch gesture. Selecting a POI reveals its sub-POIs according to the hierarchical structure of Section 2.3 and, where applicable, plays or shows its media. Selecting a different branch collapses the previously revealed one, so the scene never becomes cluttered.

Figure 11: Hierarchical POI selection in sequential order from the user's perspective on the Apple Vision Pro during the Graben experience (use case "Facades around a town square). First, an information POI is selected by the user, revealing text and two image-type sub-POIs. Then, one image POI is selected, revealing the image.

4.2.2 Proximity-based interaction

Proximity-based interaction uses the continuous distance check of the top-level POIs. When the user comes within the outer proximity radius of a POI, its icon appears; when the user comes within the inner radius, the POI becomes selectable and its text is shown, and any proximity-triggered sub-POIs are revealed. Walking away again hides the POI and the sub-POIs it had revealed. This is conceptually identical to the

collision-based detection used for the calibration spheres in A3.3. In addition, the content of each POI is continuously billboarded: it is rotated to face the user while kept upright, and scaled with distance so that it keeps a roughly constant apparent size at any range, which keeps the icon, text and images legible whether the user is near or far. In combination, this results in a guided tour that invites users to come closer to the POIs without sacrificing readability of virtual content. Proximity control can be turned off at the beginning of the experience resulting in all POIs being constantly visible, which makes the experience inclusive to less mobile users.

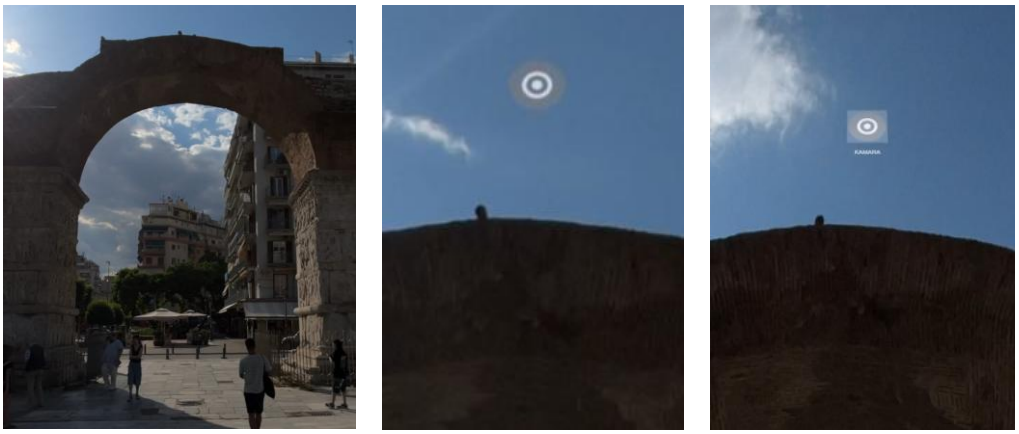

Figure 12: Proximity-based interaction with a POI from the user's perspective on the Apple Vision Pro within the Kamara experience (use case "Large Landmark"). First, no POI is visible (left). The user moves closer and reveals a POI at the top of the structure which is not interactible yet (middle). The user moves even closer, revealing the name of the POI (right). Now, it can be selected by the user.

4.2.3 POI prefabs

Each POI is instantiated from a prefab selected according to its media type – a default POI, a text (information) POI, an image POI, or an audio POI – through a prefab set configured on the ExperienceManager. A prefab bundles the billboarded icon, the content box for text and images, the selection collider and interactable, and the script implementing the visibility, distance and media behaviour described above. Visibility is toggled by enabling or disabling the relevant renderers and colliders depending on the current user location.

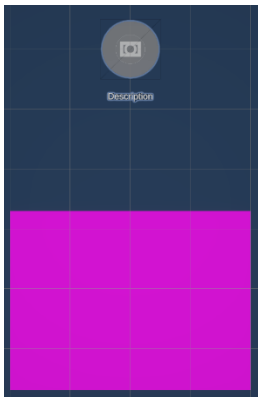

Figure 13: POI prefab of image POI with image placeholder in the Unity editor.

5. SUMMARY AND CONCLUSIONS

This deliverable has documented the conceptual design and technical implementation of the guided AR tour (A3.2). It defined three guided-tour use-case types – large landmarks, small monuments on a square, and facades around a town square – together with the supported media, the hierarchical level-of-detail content structure, and the gamification elements which will be implemented as part of the user interaction module (A3.4). It described how three Vienna sites and three comparable Thessaloniki sites were selected and filled with publicly available cultural-heritage content, and how each experience is set up on the project authoring tool. Finally, it described how the Unity application loads an experience using the project API, georeferences its POIs using the calibration of A3.3, and reveals their content through selection- and proximity-based interactions. The resulting prototype is content-complete and ready for deployment in the first study.

The achievements of this deliverable can be used in the AR-related further parts of work package 3 as discussed below:

- A3.4 Guide Recording Module: the running guided tour, with its georeferenced and revealed content, provides the scenes and interactions that users will be able to record and revisit.
- A3.5 User Interaction Module: the selection- and proximity-based interactions implemented here, together with the level-of-detail reveal logic, form the basis for the refined interaction dialogues designed in this activity.

- A3.6 Technical Evaluation: the complete, content-filled tour can be instrumented and logged to evaluate the spatial accuracy of the placed content and the usability of the interactions.

Furthermore, activities in the project are enabled as follows:

- Data collection of the first prototype testing (WP2): the deployable prototype makes it possible to collect user data during the first round of testing.
- Basis for the second prototype testing (WP2): the findings of the first study will inform the refinement of the prototype for the second round of testing.

The results outlined in this deliverable contribute directly to the broader objectives of ZOE – namely, to democratise access to culture, promote lifelong learning, and integrate digital innovation into adult education – by making it possible to experience virtually enhanced cultural sites through an outdoor guided AR tour.