

# An Interactive Multimedia Framework for Digital Heritage Narratives

Neeharika Adabala<sup>1</sup>, Naren Datha<sup>1</sup>, Joseph Joy<sup>1</sup>, Chinmay Kulkarni<sup>1</sup>  
Ajay Manchepalli<sup>1</sup>, Aditya Sankar<sup>2†</sup>, Rebecca Walton<sup>2†</sup>

<sup>1</sup>Microsoft Research India  
“Scientia”, 196/36 2nd Main  
Sadashivnagar, Bangalore 560 080  
India

<sup>2</sup>University of Washington  
Box 352350  
Seattle, WA 98195-2350  
USA

## ABSTRACT

The cultural heritage of a region is conveyed by both tangible physical artifacts and intangible aspects in the form of stories, dance styles, rituals, etc. Hitherto, the task of creating digital representations for each of these aspects has been addressed in isolation, i.e. using specific media most suited to the artifact such as video, audio, three-dimensional (3D) models, scanning, etc. The challenge of bringing together these separate elements to create a coherent story, however, has remained unaddressed until recently. In this paper we present a unified digital framework that enables the integration of disparate representations of heritage elements into a holistic entity. Our approach results in a compelling and engaging narration that affords a unified user experience. Our solution supports both active (user-controlled explorations) and passive (watching pre-orchestrated narrations) user interactions.

We demonstrate the capabilities of our framework through a qualitative user study based on two rich interactive narratives built using our framework: (1) history and folklore surrounding a temple in South India, and (2) a historical account of an educational institution also in South India.

## Categories and Subject Descriptors

H.5. [INFORMATION INTERFACES AND PRESENTATION]:  
H5.1 [Multimedia Information Systems]: *Artificial, augmented,  
and virtual realities*; H.5.2 [User Interfaces];

## General Terms

Design, Human Factors, Measurement.

## Keywords

Multimedia, Storytelling, Heritage, Narratives, User Study

## 1. INTRODUCTION

Over the past few years, the creation of digital representations of cultural heritage artifacts has gained considerable attention from both private and public organizations such as museums, communities and governments [1]. This recent thrust is a reflection of attempts to digitally preserve cultural heritage, provide widespread access to information through technology like

† This work was done during an internship at MSR India

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'10, October 25–29, 2010, Firenze, Italy.

Copyright 2010 ACM 978-1-60558-933-6/10/10...\$10.00.

the internet, and create documentation that helps advance research and aid understanding. The success of these initiatives is largely dependent on the richness of the digital representation utilized.

A variety of technologies may be employed to create a digital representation of tangible artifacts like sculptures, buildings, and cities or intangible artifacts like music, dance, and folklore, depending on the core aspect of the heritage to be digitized. For example, audio recordings are most suitable for capturing music or spoken language; rituals and dances are best captured in videos; physical artifacts are best captured by 3D modeling, 3D scanning, or photography. However, the complete significance of the artifacts in a digital depiction of heritage is brought out only if the artifacts are presented in the context of each other. The ability to bring together diverse representation media to create a holistic cultural depiction remains elusive. We address this gap in the digital representation of cultural heritage by providing a *framework* that enables composing diverse aspects of culture (captured with the most appropriate technology) into a holistic, coherent representation. The main contributions of our work are (1) Mechanisms to build rich, immersive narratives; (2) Support for user interaction; (3) Modular composition of multimedia experiences; and (4) Evaluation through qualitative user studies.

The framework enables the weaving together of *multiple* underlying media from diverse technologies into a single compelling narrative and also enables a guided narration to coexist with interaction and exploration. Details about the platform and the process of creating the narrative can be found at the Digital Heritage web site at <http://research.microsoft.com/rin/> and [2].

This paper is organized as follows: the next section outlines our interactive multimedia framework. Section 3 presents an evaluation with a qualitative user study, and Section 4 presents conclusions and ongoing work.

## 2. FRAMEWORK

The framework has two main functions:

1. To enable a creator/author of a cultural heritage resource to compose an interactive narrative that conveys the richness of the information while preserving the subtleties of relationships among different artifacts that make up the underlying story.
2. To enable the user/viewer of the resource to experience the narrative and underlying environment as a guided tour while allowing for deeper, context-specific exploration at any time.

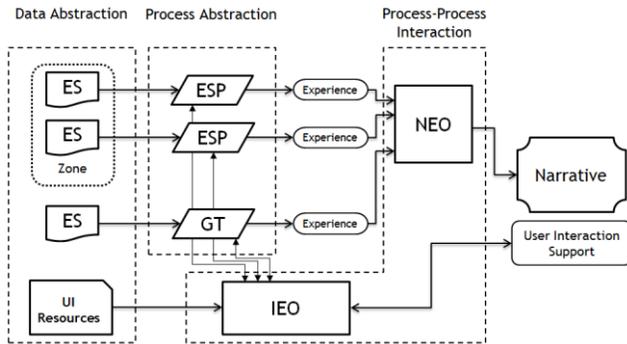
The framework contains the following core aspects:

- **Data Abstraction:** Categorizing and abstracting heritage data (content) into a small but powerful set of generalized data types that may be combined to create rich interactive narratives.

- **Process Abstraction:** Encapsulating the diverse technologies that render artifacts into modules within a uniform interface that enables flexible composition of current and future rendering and visualization techniques.
- **Process-Process Interaction:** Enabling orchestration of the elements created in the framework to provide the user a holistic experience that allows both interactive exploration and a guided narration of heritage artifacts.

The framework incorporates semantic units of representation in the form of modular *experiences* – compositions of data into individual user-perceivable streams. Experiences are in turn composed into *narratives* – mostly linear compositions of experiences in a story or episode in history. The goal of the framework is to enable composition of experiences into a seamless and holistic experience.

## 2.1 Architecture Overview



**Figure 1: Architecture of the framework depicting interaction between entities**

The following are the entities that comprise the framework:

**Data Abstraction:** The digital data associated with cultural heritage artifacts are factored into the following fundamental data types:

- **Experience Stream (ES)** is a “flythrough” of an abstract “world.” Each ES represents a particular visualization of some aspect of a heritage artifact. Examples of ES’s include (1) a predefined path along a map and (2) a predefined flythrough of a 3D model of a building.
- A **Zone** is a grouping of cultural artifacts based on spatial proximity. Ambient artifacts (such as ambient sound effects) can be triggered based on the viewer’s proximity to particular Zones.
- **User Interface Resources** are audio/visual resources and other user interface objects that are tied to the spatial embedding of the narrative.

**Process Abstraction:**

- **Experience Stream Provider (ESP)** represents an abstraction of the audio and graphical rendering mechanisms that translate the data in an ES into a form that is perceivable to the user.
- **Generalized Trajectory (GT)** represents an abstraction of mechanisms that implement various styles of guided flythroughs or navigation paths. Regardless of the type of ES, the GT abstraction provides a standard interface to

generate a smooth flythrough experience (with support for pause, explore, and resume) for the viewer, constructed from time-stamped key frame location data in the ES.

**Process-Process Interaction**

- **Narrative Experience Orchestrator (NEO)** coordinates the activities of multiple Experience Streams to provide a holistic, interactive narrative experience.
- **Interactive Experience Orchestrator (IEO)** manages ambient artifacts (e.g. location-specific sound effects) and sends events to ESPs when artifacts associated with a particular Zone come in or out of scope.

The user-visible (rendered) outputs of the framework are below:

- **Experience** is the visual/aural output of an ESP process and provides the building blocks to create a holistic representation of the cultural heritage.
- **Narrative** is the final product of the framework that is generated by the NEO. It comprises multiple Experiences, which play out over time while still allowing the user to explore and interact.
- **User Interaction Support** transforms the user experiences from “passive observational” to “active user controlled” by leveraging User Interaction Resources that enable user-controlled interaction with the holistic heritage representation. This allows the user to pause the narrative at any time and “plunge into” the worlds of underlying Experience Streams.

The process of applying these entities to create a representation of the heritage is briefly described here. The data representations of various artifacts are categorized and packaged into ES’s, which may be grouped into Zones. Each ES is acted on (rendered) by an ESP that translates the data into an audio-visual experience. These experiences are encapsulations in the form of building block (“componentized”) walkthroughs into different kinds of “worlds.” The experiences are compiled by the NEO to generate a narrative. The Ambient Experience Orchestrator (AEO) is a process that helps to support the user interaction with the system. It tracks the status users in the context of the narration and also supports user interaction throughout the narrative. This is achieved by enabling the AEO to poll the user control events from the interface and pass them to the appropriate ESP which effects the resulting modification in the experience provided to the NEO. Figure 1 diagrammatically summarizes the above description the framework’s architectural entities.

We have applied this architectural framework to create a platform for digitally representing heritage, and have implemented this platform as a Windows application, using ES technologies to create compelling narratives. These technologies included mapping, Photosynths [3], gigapixel imaging [4], audio, video, and text annotations. However, the framework is not tailored for current technologies alone. The platform is scalable to accommodate both additional technologies and techniques.

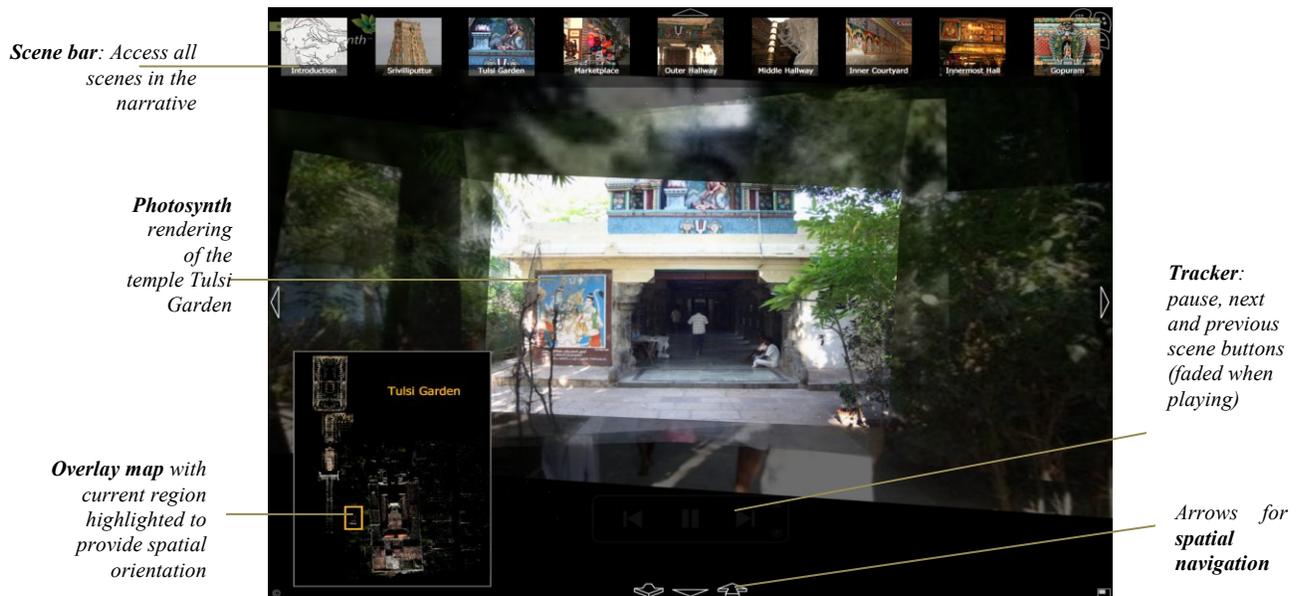


Figure 2: Screenshot from the Andal Narrative

### 3. EVALUATION

We evaluated the effectiveness of our framework with a qualitative user study. The narratives employ the interactive narrative framework, and users can explore the information in any order. In addition, they can pause, explore the environment, and resume the narrative. Our study exposed users to two digital heritage narratives, as described in the following section.

#### 3.1 Outline of narratives

##### *Narrative 1: The Legend and History of Andal*

This narrative describes the history and folklore surrounding a temple in South India.

The narrative comprises ten scenes (Figure 2 shows a screenshot). The first scene describes the general history of temples in India, followed by a scene that offers the user a panoramic view of the town. Successive scenes walk the user through different parts of the temple, starting with the Outer Courtyard and ending with the Innermost Sanctum. Interspersed in this walk is a brief account of the history of the temple and its deity. All scenes in the narrative use multiple media streams, including animated slideshows, Photosynth, audio, video and gigapixel images, and clickable annotations. Scenes range in duration from 33 seconds to 1 minute 26 seconds (mean= 51.5 seconds, median=51 seconds). All scenes (except ones that are outdoors) also include an overlay map that highlights the room that contains the scene.

##### *Narrative 2: A Walk through IISc's Heritage*

Our second narrative deals with the history and heritage of the Indian Institute of Science (IISc). This institution was chosen because it offers different subject matter than the first narrative and is easily accessible to subjects in our user study (see Section 3.2.1 below).

This narrative comprises seven scenes, with duration ranging between 28 seconds and 1 minute 36 seconds (mean: 1 minute, median: 36 seconds). The first scene provides a general history. Successive scenes discuss important areas and rooms in the institution. All scenes employ multiple media elements: Photosynth or HDView (gigapixel) background, ambient audio, voiceover, overlaid images and videos. The second and the fifth scenes include clickable annotation elements.

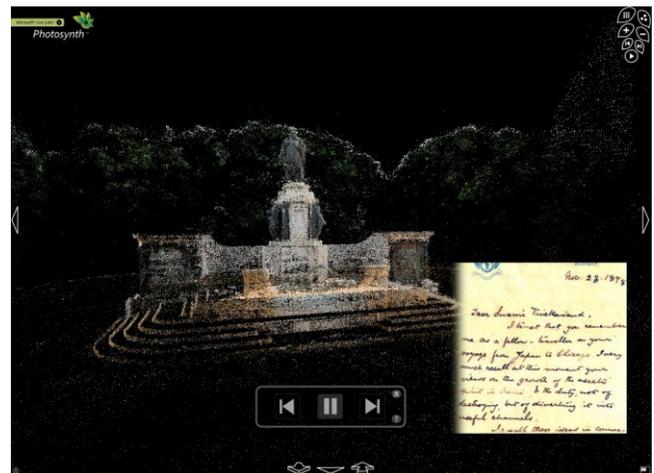


Figure 3: Screenshot from the IISc narrative, showing a Photosynth point-cloud and an image overlay

#### 3.2 Methodology of User Study

The first narrative was tested with untrained users who had never seen the platform before the test. The second narrative was tested with the same users after a five-minute training session. Testing with two narratives allowed us to both gauge initial user impressions and also understand how users interacted with narratives with some prior exposure.

Our evaluation sought to qualitatively understand the effectiveness of our framework. Our evaluation addresses three questions:

**RQ1:** Do users holistically experience digital heritage narratives that leverage multiple streams?

**RQ2:** Are users able to successfully combine exploring and experiencing narration?

**RQ3:** Does the digital heritage platform present users with a logical and discoverable means to use its features?

We performed our evaluation with 8 subjects (4 male, 4 female) between the ages of 20 and 28 (mean=21.7, median=20). This demographic was chosen based on prior studies that showed their familiarity with Internet and multimedia [5]. The subjects were

recruited among interns in our own organization and students from a local engineering college and a local design school.

Each subject participated in an hour-long session that consisted of five parts: (1) an exploration task that asked subjects to try using the digital heritage platform with the IISc narrative loaded but no instructions; (2) a five-minute training session that demonstrated how the platform should be used for tasks such as playing, pausing, and interacting with the Photosynth background; (3) a set of post-training tasks (stopping and zooming at a particular painting during a walkthrough; finding and clicking an annotation that brought up a video); (4) an exploration period where users were free to interact with either narrative of their choosing; and (5) an oral interview and written questionnaire about the overall impressions of the program. Users were encouraged to think aloud during the first, third, and fourth parts of the session.

### 3.2.1 Results and Discussion

#### RQ1: Do users holistically experience the narratives?

Overwhelmingly, users tended to look at a digital heritage narrative as a monolithic media element with multiple scenes that could be paused and explored at will. P2 remarked at the start of the session, “Oh, it is like a movie!” and was surprised to understand that one could “zoom and move [around] in a movie.” Participants also said they liked the ambient music and tied it strongly with the other media elements onscreen. P5 remarked, “I like this changing music...it’s as if I’m in a temple.” While immersion was not a direct objective of our platform, we observed several other participants also reported immersive experiences. P2 told us, “when you click around, you feel like... ooh... I need to go there, I need to visit.” Such remarks indicate that the platform could also be used for highly immersive experiences.

#### RQ2: Can users combine exploration with narration?

Before the training session, we observed that participants engaged in only minimal interaction with the narrative.

However, after the training session, interaction rates increased considerably, and users explored the narrative with greater confidence. On average, participants paused the narrative every 20 seconds and explored the environment (moving around in Photosynth and zooming into specific regions in gigapixel imagery) at least once per scene. These explorations were more involved and longer towards the end of the session.

Participants also moved between scenes with greater frequency as the session progressed. At the start of the session, users were more concerned with understanding “the general idea” [P3] of the narrative. Towards the end, P5 said “I’m skipping between scenes because I want to see what’s in all the scenes.”

#### RQ3: Is the platform discoverable?

We asked users to rate the platform on several parameters. Where we asked users if they agreed with the statement “Learning to operate the system [is easy]”, (1: strongly agree, 7: strongly disagree) the median score was 2. However, for the question “Exploring new features by trial and error [is easy]”, the median score was 2.5. This suggests that while users can quickly learn to use the system, using the platform without any instructions is harder. We attribute this to two factors: 1) our prototype lacked in-program help and 2) most of our participants were unfamiliar with technologies like Photosynth and could not fully interact with these technologies before the training session.

This was also more directly seen in our own observations: while most participants undertook some exploration before the training session, they did not realize the platform’s full capabilities: P3

“did not know you could click on the [highlight] rectangle [at any time]” (to zoom into a Photosynth photo).

These observations suggest that while users could, in a limited sense, use our platform without instructions, in-program help and training would help users interact more efficiently.

#### Summary

Overall, we found that users liked the digital heritage concept. When asked to rate it, users gave it a median score of 2 on a 7-point Likert scale (1: strongly like; 7: strongly dislike) and preferred it to other representations: “Frankly speaking, I hate historical books. I’d prefer this more than [*sic*] books”[P2].

Participants interleaved narration and exploration with minimal training and perceived the narrations as single holistic experiences without breaking them down into their constituent parts. User experience could be improved with better instructions within the program itself and improving the interface to allay fears about losing one’s place in the narration.

## 4. CONCLUSIONS

In this paper, we present a unified framework that enables the integration of disparate representations of heritage elements into a holistic entity that enables immersive guided walkthroughs as well as active exploration.

We have used the framework as a platform to create narratives, proving that it is extensible and easily adapts to accommodate multiple representation technologies. These narratives weave both tangible and intangible heritage elements into a rich and immersive experience. Initial qualitative user feedback suggests that users find the unified experience more compelling and engaging than traditional media such as video.

As future work, we would like to encourage the definition of standard representation formats for rich interactive narratives in order to enable a diversity of authoring, transformation, content indexing and rendering tools. We believe that this will stimulate widespread creation and consumption of this new class of multimedia content.

## Acknowledgements

The authors would like to thank P. Anandan, Rick Szeliski and Curtis Wong for their guidance on this project. We would also like to extend our gratitude to the participants of the user study for volunteering their time towards the evaluation of the platform.

## 5. REFERENCES

- [1] Zorich, D.M., “A survey of digital cultural heritage initiatives and their sustainability concerns”, Council on Library and Information Resources, 2003, <http://www.clir.org/pubs/reports/pub118/contents.html>
- [2] Adabala, N., Datha, N., Joy, J., Kulkarni, C., Manchepalli, A., Sankar, A., Walton R., An Interactive Multimedia Framework for Digital Heritage Narratives. Microsoft Research Technical Report MSR-TR-2010-101, July 2010
- [3] Microsoft, Photosynth. <http://photosynth.net>
- [4] Kopf, J., Uyttendaele, M., Deussen, O., Cohen, M., Capturing and Viewing Gigapixel Images. In ACM SIGGRAPH, 2007
- [5] Jones, S., Fox, S., “Generations Online in 2009”. Pew Internet & American Life Project, Jan. 2009. <http://www.pewinternet.org/Reports/2009/Generations-Online-in-2009.aspx>