CHANGING THE HERITAGE

The Arches Open Source System

BY DAVID MYERS, YIANNIS AVRAMIDES, AND ALISON DALGITY

Whether to help guide construction of a highway or gas pipeline, evaluate the condition of cultural heritage in the aftermath of a natural disaster, or determine if a demolition permit should be issued for a significant building, inventories are a necessary tool for heritage management. They inform authorities, scholars, and the public of essential information about heritage resources including their size, location, and significance. They also enable comparison of sites, aiding in categorization, appraisal of authenticity and integrity, and determination of relative significance—assessments that can assist in prioritizing management interventions. Legislation in many jurisdictions mandates the use of inventories as a means of heritage protection. In addition to their role in public administration, inventories are valuable for research, heritage tourism, and general public interpretation and understanding, because they organize information about cultural heritage.

In recent decades, the development of new digital information technologies, particularly geographic information systems (GIS), has substantially improved the effectiveness of heritage inventories. Rapidly growing global access to the Internet has

Aerial view of Ghent, capital of East Flanders in Belgium. Early in the development of Arches, the Flanders Heritage Agency provided test data and valuable advice on a number of development issues. Photo: Edelseider, courtesy Wikimedia, licensed under the Creative Commons Attribution–Share Alike 3.0 Unported license.

INVENTORY PARADIGM

made possible broadly accessible, web-based information systems for the inventory and management of immovable heritage. However, the development of an effective system can be a costly and substantial undertaking well beyond the reach of many heritage agencies.

To address this challenge, the Getty Conservation Institute (GCI) and World Monuments Fund (WMF) partnered to develop for the international heritage field a modern information system specifically designed to create and manage heritage inventories. A major result of this partnership is the just-released version 1.0 of Arches, a new open source geospatial web application for cultural heritage inventory and management, which allows organizations to view, create, edit, and query data.[1] As an open source product, the Arches software is available at no cost, and individual users may modify it to meet their specific needs. Incorporating internationally adopted inventory standards, Arches will give heritage organizations most of what they require to create high-quality inventory and management systems.

BACKGROUND AND RATIONALE

In 2004, a year after the invasion of Iraq, the GCI and WMF formed the Iraq Cultural Heritage Conservation Initiative to support the Iraq State Board of Antiquities and Heritage (SBAH) in its role as national steward of Iraq's wealth of archaeological and architectural heritage. One component of the initiative was to create a national heritage information system for the SBAH to help it protect Iraq's heritage from looting, development, and other threats. Unfortunately, the political and security conditions in Iraq in the years following the invasion made progress slow and intermittent.

To overcome these difficulties, the GCI and WMF collaborated with the Jordanian Department of Antiquities to establish a similar system in Jordan with the aim of adapting that system for use in Iraq when the situation there improved. In June 2010, the GCI and WMF completed development of the Middle Eastern Geodatabase for Antiquities (MEGA)-Jordan—a web-based, bilingual (Arabic and English), geospatial information system designed to serve as an archaeological site inventory and management system for the Department of Antiquities of Jordan. The system was deployed nationwide in December 2010 and remains in use. The Department of Antiquities has allowed public access to the system for viewing and searching purposes at www.megajordan.org.

In April 2011, the GCI and WMF made available to the Iraq SBAH a prototype of MEGA-Iraq for inventory and management of archaeological sites. The GCI and WMF planned to expand the capabilities of MEGA-Iraq to document other types of immovable heritage, but this work has been delayed by the situation in Iraq.

In the process of creating MEGA, the GCI and WMF quickly discovered that the options available to heritage agencies for creating and managing inventories left much to be desired. Proprietary software rarely fits the needs of the heritage field. Therefore, cultural heritage organizations are spending precious resources to create custom information systems from scratch, in addition to committing to high-priced proprietary software licenses and upgrades. The use of proprietary and nonstandard data formats often leads to the inability to combine or compare datasets or share data with other systems, frequently because of obsolescence. As a result, organizations often struggle to make the best use of the data they have amassed.

The successful deployment of MEGA in Jordan prompted many institutions worldwide to express interest in using a customized version of that system for their own inventories. Based on the needs observed and the interest expressed, the GCI and WMF decided to pool resources to create a purpose-built information system that would be ready for organizations to download, customize, and deploy independently. Arches is intended to reduce the necessity for heritage institutions to spend scarce resources on creating systems from the ground up, and also to alleviate the need for them to engage the complex and constantly changing world of software development.

SYSTEM DEVELOPMENT

In the summer of 2011, the GCI and WMF began work on Arches as an open source system. To develop Arches, the GCI and WMF selected Farallon Geographics, which had proven its

This screen shot of Arches depicts a user–defined polygon search for heritage resources in the area of a simulated construction project. Heritage resources within the polygon that would be impacted appear as icons. When many heritage assets are in close proximity in a map view, a cluster icon appears and displays the number of heritage assets that cluster represents. Image: © 2013 Google Map Data.

The Roman ruins at Jarash in Jordan, with the modern city looming in the background. The Arches system grew out of GCI–WMF development of MEGA–Jordan—a geospatial information system designed to serve as an archaeological site inventory and management system for Jordan's Department of Antiquities. Photo: © Samer Muscati.

expertise and had worked effectively with the GCI-WMF team during the development of MEGA. The first step was to define the requirements of a generic system that would support the needs of heritage organizations internationally to create and maintain inventories of heritage resources and would help manage them. The GCI and WMF consulted international best practices and standards, engaging nearly twenty national, regional, and local government heritage authorities from the United States, England, Belgium, France, and several Middle Eastern countries, as well as American and European information technology experts.

As a result of these efforts—and from lessons learned during the development of MEGA—the project team determined that Arches should serve several purposes fundamental to the understanding, appreciation, and protection of heritage places, including identification and inventory, research, condition assessment, determination of management priorities, and raising of public awareness of these resources.

SYSTEM DESIGN AND FUNCTIONALITY

Overriding principles for the design of Arches include:

- *Standards-based:* Arches incorporates internationally adopted standards for heritage inventories, as well as international standards for heritage data and information

technology, to promote data sharing and longevity regardless of technological advances.
- *Accessible:* Web-based to allow broad access, Arches is designed to be user friendly and purposely avoids complex analytic tools in favor of an easy-to-learn user interface.
- *Customizable:* The system's code is open and is capable of being configured and extended to meet the particular needs of adopters.
- *Economical:* Released under an open source license, Arches is available at no cost, and adopters may pool resources to pay for customizations and maintenance.

Based on the requirements gathered, the project team specified the features that the core version of Arches would need to contain. Because it was developed to serve a variety of organizations with different needs, it was designed to maximize flexibility. For example, adopters may control the degree of data privacy that their system contains. Arches will allow each organization deploying it to implement an individual information access policy; organizations may choose to have their systems and data open to online access or to limit data accessibility. The system does enforce and promote the standardization of data through validation and controlled vocabularies, including thesauri.

Data within Arches are structured into four primary theme

categories. *Heritage Resources* includes all types of immovable heritage, such as archaeological, architectural, landscape, urban, and maritime heritage. *Activities* comprises historical events as well as contemporary activities, such as those related to investigation, designation and protection, and management. *Actors* includes persons as well as organizations, both historical and contemporary, and *Documents* contains all types of documents and images. Arches manages relationships among data organized under these themes, so that a particular *Actor* (for instance, Mahatma Gandhi) may be related to multiple *Activities* (for example, historical events such as political protests), *Heritage Resources* (such as protest sites, judicial courts, and prisons), and *Documents* (such as letters written by Gandhi). This structure can aid in the discovery of previously unknown relationships among people, places, events, and documents, which can, in turn, lead to new knowledge.

HERITAGE INVENTORY STANDARDS

Growing out of documentation practices in a number of countries, international standards have been developed for the inventory of archaeological, architectural, and movable cultural heritage. These standards identify "core," or essential, items of information that should be part of any cultural heritage inventory. These standards were also created to facilitate data sharing across political boundaries and to serve as a reference for heritage organizations, which, as they create inventories, often grapple with identifying the optimal set of inventory data to meet the practical requirements of heritage stewardship. One standard for inventory of architectural heritage, the *Core Data Index to Historic Buildings*

and Monuments of the Architectural Heritage, was adopted by the Council of Europe in 1992. A second standard for inventory of archaeological heritage, the *Core Data Standard for Archaeological Sites and Monuments*, was adopted by the International Committee for Documentation (CIDOC) of the International Council of Museums in 1995.

CIDOC is now finalizing a combined standard for the inventory of both archaeological and architectural heritage known as the *International Core Data Standard for Archaeological and Architectural Heritage.* For the Arches system, this combined standard was used to identify the data fields of version 1.0. Organizations that deploy Arches can customize those data fields to meet their specific requirements.

HERITAGE DATA STANDARDS

A widespread problem with heritage data over the long term is that data loses its meaning if it was inadequately documented and the individuals who originally understood it are no longer available. To encode and preserve the meaning of information managed by the software, Arches uses the CIDOC Conceptual Reference Model (CRM), adopted by the International Organization for Standardization (ISO); this model is a carefully constructed ontology that defines the possible relationships among heritage concepts.[2] Use of the CRM keeps the data independent of conventions that are particular to the design of Arches. It also contributes to powerfully effective searches within, as well as across, data sets. It will facilitate data migration to newer systems and aid in the preservation of data over time.

The integration of the CRM into Arches reflects one of

A GCI–WMF open source community workshop on Arches in England in June 2013. Twenty heritage professionals from government agencies and academic institutions in Britain and Europe worked with the project team to provide feedback in advance of release of version 1.0 of the system. Photo: Martin Newman.

the most exciting recent developments in information technology—namely, semantic awareness and the potential to publish and disseminate information as linked open data (LOD).[3] This is a significant achievement because it allows heritage agencies and others to share data and establish new relationships among that data. For example, a historian somewhere in the world might document the birthplace of an artist, which happens to be a building now under the jurisdiction of a local heritage agency elsewhere in the world. If the heritage agency is unaware of this connection, a search using LOD will yield the new information from the historian, and this knowledge may help to justify the building's legal protection by the agency. Because of this kind of capability, there is growing interest on the part of libraries, museums, and archives in publishing structured data as LOD. Arches will greatly facilitate the publication of data on immovable cultural heritage in this format.

INFORMATION TECHNOLOGY COMPONENTS AND STANDARDS

Arches has been developed with several modern open source components. The application employs open data standards[4] and is designed to access and process geospatial data based on the standards and specifications published by the Open Geospatial Consortium (OGC). Compliance with OGC standards makes Arches compatible with desktop GIS applications widely used in the heritage sector, including Esri ArcGIS, Google Earth, and Quantum GIS, as well as with common web browsers and online satellite imagery and map services (e.g., OpenStreetMap, Google, and Microsoft). Arches implements application programming interfaces (APIs) providing modern and standards-based methods for integrating multiple information management systems.

SYSTEM RELEASE AND THE CREATION OF COMMUNITY

Version 1.0 of the software was completed and released in September 2013, and it is freely available for download from www.archesproject.org. This core version was unveiled during the 2013 symposium of the ICOMOS International Committee for Documentation of Cultural Heritage (CIPA) in Strasbourg, France, and discussed in an information session and a keynote presentation. An evolving software road map lists and prioritizes many additional features that the Arches open source community may now create to enhance the core version. These features include more advanced options to import and export data and to interact with other systems, and a mobile app for use in field data collection.
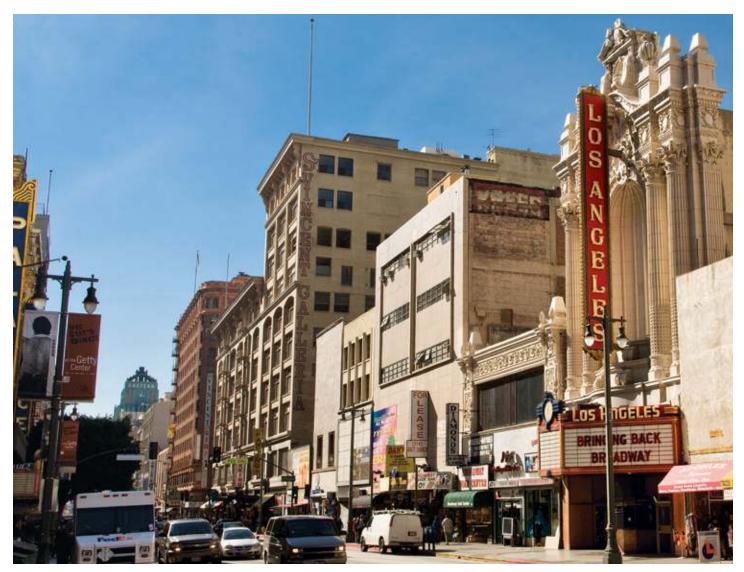
Arches was designed as a common platform for heritage data management that is easy to customize and extend depending on users' existing or future needs. To meet this ambitious goal, it was important to create and nurture an environment of ongoing collaboration by a diverse community of heritage orga-

nizations and professionals, as well as IT specialists. During the development of version 1.0 of Arches, the contributions of heritage institutions played a critical role. Early on, the Flanders Heritage Agency gave test data and valuable advice on a number of development issues. English Heritage contributed substantially by providing additional data for system development, testing, and demonstration, by offering guidance on controlled vocabularies and the incorporation of the CIDOC CRM, and by leading the system documentation effort.

The community is open to everyone and can include a wide range of individuals with varied interests, expertise, and skills.[5] Computer programmers may decide to try solving existing issues, or bugs, that have been posted publicly for review and resolution; others may prefer to create or participate in conversations about Arches; yet others may choose to help update the user manual, translate the user interface, or answer newcomers' questions. Many participants are drawn from organizations that have direct need for the software and interest in seeing it updated and maintained. It is hoped that as the software gains adopters, development will become increasingly robust and collaborative. Under the open source license, any improvements must be made available to everyone. The success of the open source community is a key factor for the success of this initiative, and the GCI and WMF are committed to providing resources to assist the community during the first years of its existence. Ultimately, it is the community's dedication to the evolving vision of Arches that will help it thrive and guarantee its long-term sustainability.

Work on the first implementation of Arches has already begun. In parallel to the development of the core version of Arches for the international heritage field, the GCI began customizing the system for the City of Los Angeles. For several years the GCI has offered technical advice—and the Getty Foundation has provided financial support—to an ongoing citywide survey of historic resources in Los Angeles called SurveyLA (an effort that grew out of an earlier GCI assessment of the city's need for a survey). Arches will be used to manage the records of the hundreds of thousands of properties documented through SurveyLA and to publish them online so that they are publicly accessible. Once implemented, it will serve as a tool for the city to help manage historic resources and as an aid to scholars and the public conducting research on the Los Angeles historic environment.

THE POTENTIAL AND THE ROAD AHEAD

Arches represents a groundbreaking effort to create for the cultural heritage field a purpose-built system incorporating the latest information technologies in data structuring, geospatial software, and semantics. It is hoped that Arches will help break the cycle of heritage organizations independently expending scarce funds to create their own custom-made systems from scratch—a

Historic Broadway in downtown Los Angeles. In parallel to the development of the core version of Arches, the GCI began customizing the system for the City of Los Angeles, where it will serve as a tool for the city to help manage historic resources. Photo: Emile Askey, GCI.

pattern that has long characterized the heritage field. To that end, the GCI and WMF have invested substantial resources in the development of a standards-based, ready-to-use system that will preclude multiple expenditures addressing identical needs.

Arches has the potential to become the profession's standard for the inventory of immovable heritage, with multiplying benefits for the entire field. Using Arches provides a way for organizations to benefit from customizations, upgrades, improvements, and maintenance undertaken by anyone within the community. The careful integration of standards into Arches encourages the creation and management of data using best practices. This capacity, in turn, facilitates the exchange and comparison of data among Arches and other information systems, within both the heritage community and related fields, and it will ultimately support the longevity of data.

Technology advances relentlessly, and Arches must evolve through the support of the community or eventually become obsolete. However, it is most important that the heritage field address the challenge of ensuring that the body of knowledge painstakingly assembled in information systems over many decades is well protected and continues to advance heritage management and protection into the future.

David Myers is a senior project specialist with the Getty Conservation Institute. Yiannis Avramides is a program associate at World Monuments Fund. Alison Dalgity is a senior project manager with the Getty Conservation Institute.

1. Open source refers to a computer program made available free of charge to the general public with its source programming code open—which means its original design may be modified. Customizations, upgrades, or improvements made to the software by anyone must remain freely available.
2. In information science, an ontology is a framework for organizing information. It lists concepts within a field of study and the possible relationships among those concepts. Data mapped to an ontology are more clearly understood by machines, and relationships or links among information items and groups can be inferred automatically. Using an ontology also helps protect data from becoming inaccessible and obsolete, because the data are more precisely described.
3. Linked open data is a method of using the web to expose, share, and connect openly available data. LOD enables data from different sources to be connected (linked) and queried. An ontology (see note 2) must be used to prepare data to be published as linked open data.
4. Open data standards ensure compatibility among diverse data sets.
5. For more information about how successful open source communities function, see Karl Fogel's *Producing Open Source Software: How to Run a Successful Free Software Project* (2013), producingoss.com/.