The Arches Heritage Inventory and Management System: a standards-based approach to the management of cultural heritage information

P. K. Carlisle,[a] I. Avramides,[b] A. Dalgity,[c] D. Myers[c]

[a] English Heritage, The Engine House, Fire Fly Avenue, Swindon, SN2 2EH UK
`philip.carlisle@english-heritage.org.uk`
[b] World Monuments Fund, Empire State Building, 350 Fifth Avenue, Suite 2412, New York, NY 10118
`iavramides@wmf.org`
[c] Getty Conservation Institute, 1200 Getty Center Drive, Suite 700 Los Angeles, CA 90049-1684
`adalgity@getty.edu, dmyers@getty.edu`

Abstract. The Arches Heritage Inventory and Management System was developed by the Getty Conservation Institute and World Monuments Fund as an open source web-based geospatially enabled information system to help inventory and manage immovable cultural heritage. The system incorporates two CIDOC standards, the CIDOC-CRM (Conceptual Reference Model [ISO 21127: 2006]), and the draft of the CIDOC International Core Data Standard for Archaeological and Architectural Heritage. The implementation of the CDS in a modern information system, as well as the first mapping of the CDS to the CRM, have helped to further develop the CDS itself. This paper documents and shares the experiences of that process and the potential benefits of incorporating the two CIDOC standards into the design of Arches. Arches was created in response to the persistent need for a system that fits the inventory requirements of the heritage field without requiring onerous investment of time and resources by heritage organizations individually. Arches represents a unique initiative undertaken for the benefit of the cultural heritage field at large, with the long-term goal of improving management of immoveable heritage worldwide through the use of more effective inventory systems. It combines state-of-the-art software development with the insights of many heritage professionals from around the world.

Keywords. GIS, heritage management, inventories, CIDOC CRM, semantic technologies, ontologies

1   Introduction

Arches is a new, open source software system that provides a solution for compiling and managing inventories of immovable cultural heritage at a range of scales (national, regional, city, or site). The Arches project was initiated through a collaboration of the Getty Conservation Institute and World Monuments Fund, with substantial technical assistance from English Heritage and the Flanders Heritage Agency. Arches is purpose-built for the cultural heritage field, using international standards. It is freely available and has a growing international community of Information Technology specialists and heritage professionals. This paper describes the way that various international heritage standards have been incorporated in the architecture of the system, and offers guidance for future customizations by adopters.

Arches was designed as a generic platform, specifically built for managing information about immovable cultural heritage, and can be configured and customized to serve the requirements and needs of a variety of implementing organizations. The project was conceived as a response to the persistent need for a system that fits the inventory requirements of the heritage field without requiring onerous investment of time and resources by organizations to individually create systems to meet very similar needs. Equally important were the goals of promoting long-term preservation and interoperability of datasets that are maintained in Arches. These needs translated into the extensive use of existing standards. As a result, Arches is underpinned by two CIDOC standards: the revised International Core Data Standard for Archaeological and Architectural Heritage (CDS), which has been used as a basis to define the data fields in the generic version of the system, and the CIDOC Conceptual Reference Model (CRM), which has been used to provide the semantic framework.

In addition, Arches employs open data standards 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, as well as with common web browsers and online satellite imagery and map services.

1.1   The International Core Data Standard for Archaeological and Architectural Heritage

The Core Data Standard for Archaeological Sites and Monuments[1] was the result of an original collaboration between CIDOC, the International Committee for Documentation of the International Council of Museums, and the archaeology documentation group of the Council of Europe. The standard has its origins in an international conference of representatives of national archaeological records held in Copenhagen in 1991, and the decision to develop a core data standard for archaeological sites and monuments was made at the 1992 CIDOC meeting in Quebec.

To undertake the project the Archaeological Sites Working Group was established and shortly after work had begun the European Plan for Archaeology was launched under the aegis of the Council of Europe's Cultural Heritage Committee.

The launch of the Plan, in April 1993, followed the signing of the revised European Convention on the Protection of the Archaeological Heritage at Valletta in January 1992, and one of the four main elements of the Plan was a programme focusing on inventory and documentation techniques and standards with regard to the archaeological heritage including the preparation of a core data standard for records of archaeological sites and monuments, intended to complement the Council's Core Data Index to Historic Buildings and Monuments of the Architectural Heritage (CDI).[2] When the Council of Europe working party became aware of the CIDOC working group and its work on the CDS it decided that the most practical course of action was to adopt the CIDOC data standard as the basis of its own standard, subject to minor adjustments reflecting the narrower geographical focus of the Council of Europe.

In the late 2000s the CIDOC Archaeological Sites Working Group made the decision to revise the Core Data Standard to bring it up to date. Both documents had been developed before the introduction of mass computing and as such reflected their paper-based, card index origins. With the growth of connectivity and the Internet, the development and widespread adoption of Geographic Information Systems, and the advent of cloud-based computing, it was recognized that a revision was needed. As they were so closely linked, the group decided to combine the two standards. The resulting, revised International Core Data Standard for Archaeological and Architectural Heritage is intended for use in the creation of inventories for both the built and buried heritage. In September 2013, CIPA, the International Committee for Documentation of Cultural Heritage of the International Council on Monuments and Sites was invited to collaborate on the revision.

The Core Data Standard provided the Arches project team with a starting point for the development of the data structure. By identifying the types of data which users would be able to record it was possible to define the data fields which would be used within the system. Once identified these could then be mapped to the entity classes within the CIDOC CRM.

1.2   CIDOC Conceptual Reference Model (CRM)

The CIDOC Conceptual Reference Model (CRM) grew out of the work of the CIDOC Documentation Standards Working Group (DSWG), which focused on the creation of a general data model for museums, with a particular focus on information interchange.[3]

Until 1994 its activities had centered around the development of the CIDOC Relational Data Model but in 1996 the CIDOC committee decided to move away from the relational data model and adopt an object-oriented approach in order to benefit from its expressive power and extensibility for dealing with the necessary diversity and complexity of data structures in the domain. This resulted in an initiative to create the conceptual reference model and in 1999 the first complete edition of the "CIDOC Conceptual Reference Model" (CRM) appeared. In order to fully exploit the potential of the CRM as a means of enabling information interchange and integration in the museum community, and beyond, CIDOC decided to submit the CRM to the International Standards Organization (ISO). ISO, in contrast to CIDOC, had the procedures and authority to create and declare well-defined, valid editions of international recommendations. An ISO working group was established and in September 2000 the CIDOC

CRM was accepted as a working draft by ISO/TC46/SC4. In 2006 it became an official ISO standard *Information and Documentation: A Reference Ontology for the Interchange of Cultural Heritage Information* (ISO 21127:2006).[4]

The CIDOC CRM provides the definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation. It is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to.

In addition to the aforementioned standards the development and structure of Arches has been informed by MIDAS Heritage, the UK historic environment data standard, and by the international conventions on archaeological, architectural, landscape and underwater heritage.

2   Arches Application Architecture

2.1   How Arches is put together

As cultural heritage inventory software, Arches combines a geospatially-aware system with customizable definitions of heritage information, including the ability to manage controlled vocabularies, as well as searching, reporting, and user management capabilities. To allow for more flexibility on the part of individual adopters, versions 2.0 (March 2014) and higher consist of two main components: 1) the Arches Server, the core of the system, which provides system administration services and essential services for managing cultural heritage information, and 2) Data Management Packages, which the Arches Server has the ability to implement.

The initial package, which is based on the CIDOC CDS, is the "Core Data Standard" Package (CDS Package). This paper focuses on the development of the CDS Package, described in the following section. Other packages are also under development independently. An envisioned role for packages would be to apply standards and criteria that are found in different jurisdictions, such as a package complying with the United States National Register of Historic Places standards. Organizations implementing Arches are free to develop packages in compliance with standards in their own jurisdictions. Package developers have latitude in deciding how a package influences the behaviour of Arches.

Arches uses a graph data structure, and can accommodate user-defined graphs. Each Resource Type in an Arches Package corresponds to a graph that defines the database schema for a particular class of resources in that package. In the CDS Package, as well as in other packages, the graphs identify the CRM classes (nodes) and CRM properties (edges) that create the framework for storing information. While Arches is by default aware of the CIDOC CRM and contains an acknowledgment of all CRM classes and properties, package developers are not required to create graphs that represent a network of relationships between CRM classes and properties. Nevertheless, conformance with the CIDOC CRM is strongly advised for users that want to take advantage of the benefits of semantic awareness, now or in the future.

2.2   The Core Data Standard Package

The CDS Package was developed to respond generically to the need for recording information relating to various aspects of the cultural heritage as well as the history, activities, and actors associated with a heritage asset. It contains a flexible implementation of 13 "Resource Types," organized into four categories. These 13 Resource Types are described briefly here:[*]

Heritage Resources.
Each resource of these five types allows the user to record information about a specific site or monument. The Resource Types that make up the Heritage Resources category are defined according to international conventions relating to cultural heritage:

- Archaeological Heritage (site)
- Archaeological Heritage (element)
- Architectural Heritage

---

[*] Work is currently underway on version 3.0 of the Arches Server. In conjunction with this work, a new package is being developed with similar, but more streamlined Resource Types.

- Landscape Heritage
- Maritime Heritage

Activities.

The Core Data Standard only provides cross-references to activities but as Arches is intended not only as an inventory but also a management tool it was necessary to identify another standard which could be used to define the data fields required. MIDAS Heritage, the UK Monument Inventory Data Standard, is based on an extended version of the Core Data Standard which includes information schemes for various activity types. Activities are divided into four types:

- Investigation Activity
- Management Activity
- Designation and Protection Activity
- Historical Event

Documents.

This category comprises any file, document or image, whether digital or analog, which refers or relates to another Arches resource. Information sources are divided into:

- Document
- Image

Actors.

Actor resources allow the user to record information about those people or groups who have an association with a site or monument. Actors are divided into two types:

- Person
- Organization

3 Arches Graphs in Practice

3.1 Graphs in the Core Data Standard Package

In the CDS Package, the nodes that make up each graph correspond to an item of required information as specified in the International Core Data Standard for Archaeological and Architectural Heritage. As described in section 2.1, those nodes correspond to a CRM class, while edges correspond to a CRM property. Within the database each node can store a value, and the graph also defines the data type that each node can store. An example of a branch that specifies how the assignment of an appellation to a resource is represented in the database is illustrated in Fig. 1. In this branch, property P1 ("is identified by") represents the assignment of an E41 Appellation to an Architectural Heritage resource, represented as an instance of class E18 Physical Thing.
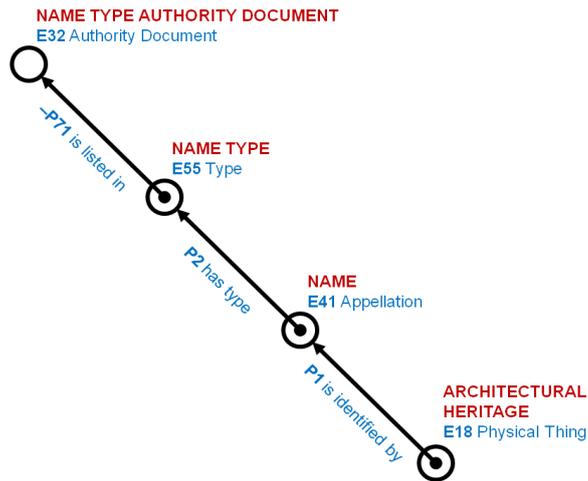
Fig. 1. This branch of a graph in the Core Data Standard Package demonstrates how the assignment of an appellation to a resource is represented in the underlying database.

As seen in Fig. 1, in addition to providing structure for the Arches database, graphs specify the integration of authority documents into Arches. For instance, the name assigned to the resource is characterized with property P2 ("has type") as a name that belongs to a particular E55 Type, which, in turn, is characterized with the inverse of property P71 ("is listed in") as an item that can be found in a particular authority document. The authority document in question is one of a suite of documents contained within the CDS Package. It contains values such as "Primary, Former, Alternative," or other concepts that are useful for characterizing the name of a resource. The pattern that represents that the value of a node is listed in, and is therefore constrained by, a particular authority document is common in Arches graphs.

In addition to the key role that they play in the function of the software, Arches graphs serve as user-friendly visualizations of the relationships that exist within Arches. Fig. 2 is a complete view of the graph for Architectural Heritage resources in the CDS Package.
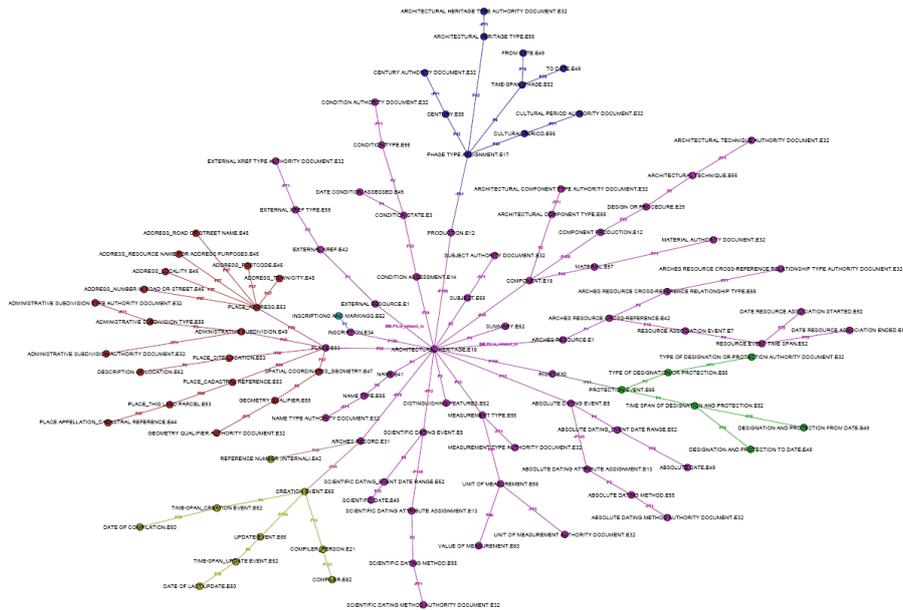
Fig. 2. Overview of the graph that corresponds to Architectural Heritage resources in Arches' "Core Data Standard" Package. Each node in a graph in Arches corresponds to a CIDOC CRM Class, and each edge refers to a CIDOC CRM Property. Visualizations were created using Gephi, an open-source network visualization and analysis software package.

3.2 Mapping Process

As described in section 3.1, the creation of the graphs in the CDS Package necessitated representing the CIDOC Core Data Standard using classes and properties defined by the CIDOC CRM. This section outlines the steps taken during that process. The starting point for the mapping process was the spreadsheet template provided by Stephen Stead of the CRM-SIG (Fig. 3).

Fig. 3. Detail of spreadsheet mapping the Core Data Standard to the CIDOC CRM, based on a mapping of MIDAS Heritage.

The provided mapping of the CDS to the CRM was based on a mapping of MIDAS Heritage, the UK historic environment data standard, first published by the Royal Commission on the Historical Monuments of England in 1998 and revised since then. The mapping provides various chains of CRM classes and properties with which items of information in MIDAS Heritage can be represented using the CRM. MIDAS Heritage represents a greater level of complexity than the Core Data Standard, and its existing mapping to the CRM provided an ideal starting point for Arches.

To create machine-readable graphs, a simplified spreadsheet was created, identifying the CDS field name represented, the CRM class used to represent it, and the path leading to it from the central node (Fig. 4).

Fig. 4. Simplified mapping of the CDS to the CRM for Maritime Heritage resources in the Arches Core Data Standard Package.

Gephi (version 0.8.2-beta),[†] an open-source network visualization and analysis software package, provided a solution for visualizing the mappings while maintaining the simple, machine-readable format. (Nodes and edges files exported from Gephi are ready for importing into Arches without further manipulation). The visualization method also helped pin-point flaws in the design of the mappings, especially where different nodes in the graph were represented using the same CRM class (for example, E32 Authority Document, as described in section 3.1).

At times, the need arose for a general, less semantically-rich representation of an arbitrary relationship between different resources in the database, whether they belong to the same or to different Resource Types. (One such example is forming a general association between two monuments, for instance, two barrows contained within the same cemetery). To create these relationships, the property "BM.PX is related to" (from the British Museum's extension to the CRM) was used.[‡]

4 The Application of CIDOC Standards in Arches

The Arches project represents the first time that the CIDOC International Core Data Standard for Archaeological and Architectural Heritage and the CIDOC Conceptual Reference Model have been brought together in the same software implementation. Users of Arches will be able to take advantage of a software system that has been designed to record an internationally sanctioned optimal set of data on immovable cultural heritage, while also encoding this information in a recognized semantic standard for cultural heritage information. In this way, Arches is a practical means to furthering the aims of the Archaeological Sites Working Group in undertaking the Core Data Standard, which include:[5]

- To facilitate communication between national and international bodies responsible for the recording and protection of the archaeological heritage,
- To assist countries at an early stage in developing systems for the recording and protection of the archaeological heritage,
- To facilitate research utilising archaeological core data where this has an international dimension.

4.1 Feedback on the CIDOC Core Data Standard

Some observations about the Core Data Standard yielded during the course of the mapping and software development effort have informed the current process of revision of the standard. In particular, the following areas were identified for clarification or improvement:

1. As a descriptive document, the CDS avoids making interpretive information a core data requirement. Nevertheless, information on significance is commonly recorded on heritage inventories, and is preferentially sought by members of the public when accessing publicly available information systems. In addition, in the case of statutory designations of heritage resources ("listing"), many jurisdictions mandate transparency about the reasons why properties have received a particular designation.
2. The "Cartographic reference" section of the CDS does not fully reflect the capabilities inherent in modern geospatial software systems such as Arches, which can automate most of the recording process.
3. Some scope notes in the CDS are unclear (for example, the scope note for "Subject," "a description of the subject of the item or group.")

Feedback on the practical application of the CDS based on the experience of users of Arches may prove useful in informing future development of the standard. It should be remembered that the inventory standards on which the CDS is based (*Core Data Standard for Archaeological Sites and Monuments* and *Core Data Index to Historic Buildings and Monuments of the Architectural Heritage*) were created in the early 1990s, and were based on an early digital inventory paradigm. Their integration within Arches is the first implementation of the complete Core Data Standard using a state-of-the-art

---

[†] Available at https://gephi.github.io.

[‡] "BM.PX is related to" is "used to indicate there is some relationship between the two resources." An extension to the CIDOC CRM was created by the British Museum during Stage 2 of the ResearchSpace project. It is available online at http://crm.rkbexplorer.com.

software system. Its use in Arches may contribute to its continued relevance for the documentation of immovable cultural heritage.

4.2   Integration of the CIDOC Conceptual Reference Model

In addition to being an implementation of the Core Data Standard, the Arches CDS Package provides a built-in representation of all information in the database using the CIDOC CRM. Whereas the learning curve and inherent complexity of the CRM are commonly voiced concerns on the part of lay users who would like to take advantage of this standard ontology, Arches has been designed to conceal the CRM from the end user. End users of Arches instead interact with the database through familiar data entry forms, similar to those that are found in modern relational database systems. As a result, the sometimes bewildering complexity of the Conceptual Reference Model remains hidden behind the scenes.

5   Conclusion

Arches represents a unique initiative undertaken for the benefit of the cultural heritage field at large, with the long-term goal of improving the management of heritage worldwide through the use of more effective inventory systems. Aside from providing ease of use and reliability, Arches addresses problems that have plagued the heritage field for a long time: the dearth of standardization in recording information about sites, the long-term loss of data due to technological obsolescence, and the challenge of accessing and sharing adequate information with stakeholders in time to mitigate risks to heritage. The incorporation of the international standards for heritage information, as discussed in this paper, will serve to promote long-term preservation of data and interoperability of the datasets that are created using the software. Available at no cost as open source software, Arches gives heritage organizations most of what they require to create high-quality inventory and management systems for immovable cultural heritage.

References

[1] International Core Data Standard for Archaeological Sites and Monuments. In: Thornes, R. and J. Bold, eds.: Documenting the Cultural Heritage, pp. 33-44. Getty Information Institute, Los Angeles (1998), http://archives.icom.museum/object-id/heritage/int.html

[2] Core Data Index to Historic Buildings and Monuments of the Architectural Heritage. In: Thornes, R. and J. Bold, eds.: Documenting the Cultural Heritage, pp. 27-32. Getty Information Institute, Los Angeles (1998), http://archives.icom.museum/object-id/heritage/core.html

[3] Crofts, N., M. Doerr, T. Gill, S. Stead, and M. Stiff, eds.: Definition of the CIDOC Conceptual Reference Model (December 2011), http://www.cidoc-crm.org/definition_cidoc.html

[4] Information and Documentation: A Reference Ontology for the Interchange of Cultural Heritage Information. ISO 21127:2006. Geneva, Switzerland: ISO, http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=34424

[5] International Core Data Standard for Archaeological Sites and Monuments. In: Thornes, R. and J. Bold, eds.: Documenting the Cultural Heritage, pp. 33-44. Getty Information Institute, Los Angeles (1998), http://archives.icom.museum/object-id/heritage/int.html