



Developing a Heritage Database for the Middle East and North Africa

Andrea Zerbini

To cite this article: Andrea Zerbini (2018) Developing a Heritage Database for the Middle East and North Africa, Journal of Field Archaeology, 43:sup1, S9-S18, DOI: [10.1080/00934690.2018.1514722](https://doi.org/10.1080/00934690.2018.1514722)

To link to this article: <https://doi.org/10.1080/00934690.2018.1514722>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 08 Nov 2018.



Submit your article to this journal [↗](#)



Article views: 16



View Crossmark data [↗](#)

Developing a Heritage Database for the Middle East and North Africa

Andrea Zerbinì 

University of Oxford, Oxford, UK

ABSTRACT

The Endangered Archaeology in the Middle East and North Africa (EAMENA) project based at the Universities of Oxford, Leicester, and Durham uses remote sensing to record archaeological and cultural heritage landscapes across twenty countries from Morocco to Iran and from Syria to Yemen. The project has developed an online heritage database built by adapting the open-source Arches heritage inventory platform developed by the Getty Conservation Institute and the World Monuments Fund. This paper discusses the process of customization of Arches v.3 adopted by EAMENA, particularly with regard to the development of new CIDOC CRM resource data models, reference data, and modifications to the Arches codebase.

KEYWORDS

Arches; web mapping; open source; CIDOC CRM; remote sensing

Introduction

The Endangered Archaeology in the Middle East and North Africa (EAMENA) project was launched at the beginning of 2015 with the aim of recording archaeological and cultural heritage landscapes across twenty countries of the Middle East and North Africa (Bewley et al. 2016). Through the interpretation of freely-available satellite imagery and historical aerial photography, as well as a range of other sources and methods, the project has collected, to date, evidence for hundreds of thousands of features of cultural heritage significance spanning all cultural periods from the Palaeolithic to the Second World War. Moreover, EAMENA—in collaboration with the partner project Aerial Photographic Archive for Archaeology in the Middle East (APAAME)—has heavily invested in the retrieval and digitization of tens of thousands of historical aerial photographs. All of these resources are in the process of being entered on a heritage inventory platform, the architecture of which is the main topic of this paper.

The EAMENA team is made up of archaeologists and heritage practitioners with field experience in several of the MENA countries, including Egypt, Jordan, Lebanon, Libya, Morocco, Syria, and Tunisia. The project is a partnership between the Universities of Oxford, Leicester, and Durham, and is funded by the Arcadia Fund until at least 2020. Since December 2016, a British Council's Cultural Protection Fund (CPF) grant has allowed the project to develop training packages for heritage professionals in the MENA region: by building on the data gathered by the EAMENA project and on the region-wide platform that it has designed, these training courses aim to introduce participants to the principles of remote sensing and its uses for archaeological investigation and heritage management. Ultimately, the goal of the training is to support the introduction of national heritage inventories or—in the rare cases where such inventories already exist—substantially update them.


Methodology and Datasets

The EAMENA project's methodology focuses on the rapid collection and presentation of large amounts of digital data as a means to preserve a record of endangered heritage landscapes and, where possible, ensure their protection. In doing so, the project sets itself within the field traditionally occupied by national heritage inventories, with their emphasis on the importance of comprehensive documentation for the enactment of solid management practices (Myers 2016). However, while heritage inventories (sometimes also known as 'Sites and Monuments Records' or, more recently, 'Historic Environment Records,' see Carlisle and Lee 2016) are mostly built from the ground up, through the systematic collection of field-acquired data, the EAMENA project adopts a different approach. Over 95% of EAMENA heritage data is acquired via the analysis of satellite imagery and aerial photography (Rayne et al. 2017: 4–8).

This determines the project's methodology for data recording; there are three basic questions to be addressed for each record: "Where?" Where is a site physically located? How accurate are the coordinates and what is the shape and size of the polygon needed to define the boundaries of the site? "What?" What makes up a given site (e.g. a number of tombs that make up a necropolis, a structure that may be interpreted as a temple, etc.)? What does the site look like in terms of site morphology, shape and features arrangement? What cultural period(s) can be assigned to the site? What might the function(s) of the site be? "How?" How does a site's condition change through time as affected by disturbances of different kinds?

The bulk of the project's datasets is acquired via the systematic investigation of very-high-resolution (VHR, varying from 0.4 m to 1 m), RGB satellite imagery freely available via Google Earth and Bing Imagery (e.g. for an application to Libya, see Rayne et al. 2017). These platforms provide numerous sets of dated imagery, thus allowing the project's image analysts to not only identify potential sites of heritage

CONTACT Andrea Zerbinì  andrea.zerbinì@arch.ox.ac.uk,  CBRL British Institute in Amman, P.O. Box 519, Jubaiha 11941, Amman, Jordan.

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/00934690.2018.1514722>

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

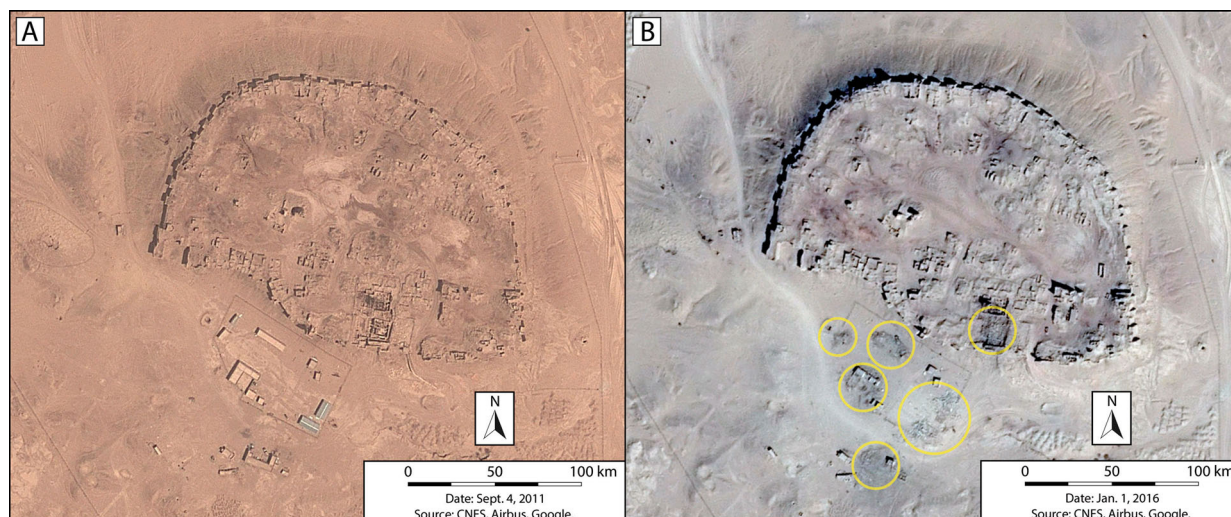


Figure 1. The pre-Islamic site of Baraqish in Yemen, in two satellite images showing the damage wrought by bombing in the summer and early-autumn of 2015 (B). Imagery: © CNES/Airbus via Google Earth.

significance, but also record their changing condition through time (FIGURE 1).

Where the quality or coverage of freely-available imagery is not satisfactory (as is the case for Palestine [Zerbini and Fradley 2018]), EAMENA has been purchasing VHR, multispectral imagery from a wide range of international retailers.

To add further chronological depth to EAMENA's condition assessments, the project has been actively engaged in the acquisition and digitization of a wide range of different collections of historical aerial imagery. Among them, it is worth recalling the Hunting Aerial Survey, a collection of 4,000 vertical photographs taken in 1953 and covering about half of Jordan (Bewley and Kennedy 2012: 225–227), which was digitized by the APAAME project and has been used by EAMENA to produce long-term conditions assessments, e.g. in the Madaba region of Jordan (Banks and Zerbini 2015). Another important collection currently being digitized comprises ca. 39,500 prints of vertical aerial photographs produced by the Royal Air Force (between 1952 and 1973) over the former British protectorate of Aden and parts of the former Yemen Arab Republic: this is currently being scanned by EAMENA in collaboration with the

Bodleian Library of the University of Oxford; this collection is essential for understanding landscape change in Yemen, particularly in areas where agricultural exploitation has vastly expanded since the 1950s (e.g. in the wadi Hadramaut drainage system). A recent open call for historical aerial photographs issued by EAMENA (Bewley and Fradley 2017) has resulted in the project being contacted by a number of private collectors and institutions, and promises to further expand our ever-growing dataset of digitized historical aerial imagery. Where copyright restrictions do not prevent it, these collections are being made available on EAMENA's online database or on the APAAME Flickr page (<https://www.flickr.com/people/apaame/>).

Through partnerships, EAMENA has also been working with other digital archaeology projects, such as Durham University's Fragile Crescent Project (Galiatsatos et al. 2009; Lawrence et al. 2012) and Leicester University's Trans-Sahara Project (Mattingly et al. 2017), to integrate existing datasets which together comprise over 16,000 sites across Syria, Iraq, and North Africa. At the time of writing, the EAMENA database contained 200,500 resources (including heritage site records, metadata for satellite imagery and aerial photographs, and bibliographic information (FIGURE 2).

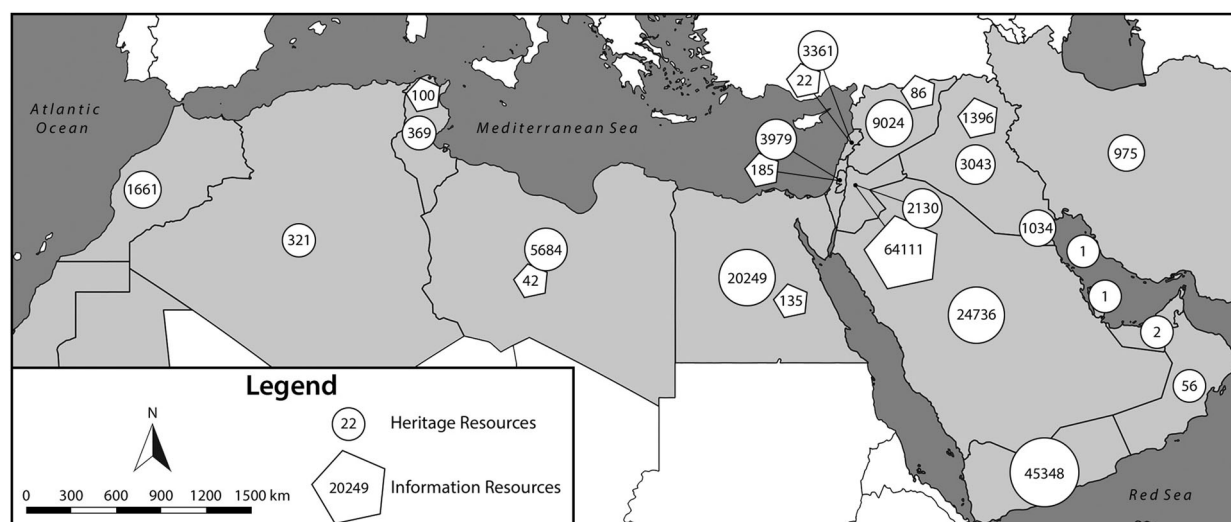


Figure 2. Distribution of geolocated resources within the EAMENA database. Centroids are heritage resources while pentagons represent information resources.

The Database

To store these ever-growing datasets, the EAMENA project has developed an online database (<http://eamenadatabase.arch.ox.ac.uk>). This database is a deployment of the Arches heritage inventory package—an award-winning open-source web app developed by the Getty Conservation Institute and the World Monuments Fund (<https://www.archesproject.org/2017/10/24/arches-wins-2017-cgia-gis-award/>).

Arches was launched in 2013 with the aim to provide an easy-to-use geospatial inventory to heritage authorities across the world (Myers et al. 2012; Myers et al. 2016). This heritage platform (Arches) particularly appealed to the EAMENA team, for in addition to allowing us to record heritage data, it could also be readily customized for the rapid and large-scale documentation of archaeological landscapes identified via remote sensing across the MENA region. Online access and the ability for records to be updated by local heritage practitioners, at little to no cost for the end user (i.e., most often, stakeholders from the local departments of antiquities), were also major benefits.

Another important consideration in the decision to adopt Arches was its native support of the CIDOC Conceptual Reference Model (CRM) for data modeling. CIDOC CRM (<http://www.cidoc-crm.org/>) is the ISO standard for describing cultural heritage data and it is widely adopted by museums and universities worldwide. It is a semantic framework designed to facilitate standardization across diverse datasets and to provide the ability to search across a number of them. In recent years, numerous archaeological projects have invested resources in an effort to map their project-specific data models to CIDOC CRM terminology, with a view to ensuring the long-term survival of these datasets and their semantic interoperability (Binding et al. 2008; Felicetti et al. 2013; Henninger 2017: 10–12).

Arches is gaining momentum as the standard web-based heritage inventory platform. Its growing user community comprises a mix of university researchers, software engineers, NGOs and, increasingly, the English historic environment records (HER) community (Sivak 2017). This is encouraging, as it ensures the continued development of the package beyond the remit of the original project and the capabilities of the core Arches team. By way of example, developers working at the Center for Virtualization and Applied Spatial Technologies (CVAST), University of South Florida, have recently extended Arches to support Docker (<http://www.docker.com>), the world's leading container software, thus making the installation of Arches easier and quicker (see this Github pull request for details: <https://github.com/archesproject/arches/pull/1748>). As of January 2018, the Arches codebase on the coding platform Github had been forked 38 times, suggesting the existence of at least as many on-going implementations. Among the projects that have adopted Arches, Zbiva (<http://zbiva.zrc-sazu.si/>) is worth mentioning as it represents the first publicly-available Arches deployment of a traditional archaeological relational database (of early Medieval funerary contexts from the Eastern Alps and northern Balkans). This project's database, first developed in the 1980s, was recently remodeled using the CIDOC CRM ontology, with the help of the European-funded project Ariadne (<http://portal.ariadne-infrastructure.eu/page/24054304>).

The EAMENA deployment of Arches is based on version 3.1.1, which was released in the autumn of 2015. This is a web

application written in Python 2.7 using Django (version 1.6), one of the most popular open source web frameworks, powering the websites of large companies such as the Washington Post and Pinterest (<https://djangobook.com/tutorials/why-django/>). The database engine used is PostgreSQL (version 9.3), with PostGIS (2.2) to handle the storage of geospatial data. Arches uses Elasticsearch, a scalable full-text search engine built on top of the Apache Lucene library (Gromley and Tong 2015), which makes live searching across complex datasets quick and easy. The front-end of Arches employs a fully responsive layout that uses the Bootstrap framework (<https://getbootstrap.com/>) and Javascript libraries jQuery (<https://jquery.com/>), Knockout (<http://knockoutjs.com/>) and Backbone (<http://backbonejs.org/>). The mapping component of Arches v.3 draws on the OpenLayers library (version 3.1, <https://openlayers.org/>) to display spatial data in the app's front-end interface. Arches v.3 comes with two separate apps: the Arches core and the Arches Heritage Inventory Package (HIP), which includes six CIDOC CRM template resource graphs (or models) and Simple Knowledge Organisation System (SKOS) reference data (also known as controlled vocabularies or thesauri), most of which created to populate the atomic database nodes of HistoricPlacesLA (historicplacesla.org), the online database of the Los Angeles Historic Resources Survey (SurveyLA), which pioneered the deployment of Arches v.3 (Bernstein and Hansen 2016: 89). On top of these two apps, the Arches installation process creates the new project's app folder structure, where customized templates, code files, and style sheets may be added without the need to modify the underlying Arches code.

At the time of writing, Arches had reached version 4.0.1 (released in October 2017). The launch of the first stable version 4 in July 2017 marked a complete re-write of the codebase as well as a major dependency upgrade, with new versions of PostgreSQL, Django, and Elasticsearch, and the replacement of OpenLayers with MapBox (<https://www.mapbox.com/>) for front-end map display. Among the many new features of Arches v.4 one should note its inclusion of a tile server (an integration of TileStache, <http://tilestache.org/>), which allows the user to import custom-made base-maps and overlays into Arches, and the Card and Graphs Designer, which make it possible for a project's data architect to modify the database resource models and data entry forms on the fly, without needing to engage with the code.

Discussions are on-going to establish a timeline to port the EAMENA database to the latest Arches version, though the difficulties inherent in the data migration process (particularly when dealing with large quantities of data, as it is the case for our project) as well as the amount of project-specific customization applied by EAMENA to its Arches v.3 deployment are likely to delay this.

Laying the Foundations: Developing Resource Graphs and Reference Data for EAMENA

The first significant step in the Arches implementation workflow pertains to the creation and upload of the resource graphs and reference data. The Arches resource graphs function as data schemata, although they are not used to produce the PostgreSQL tabular structure, as it would be the case in a conventional relational database (RDB). Instead, the Arches RDB is divided up into four schemata, which complement

the default Django ‘public’ schema. “Data”: comprising tables for files, geometries, dates, numbers, strings and concepts (i.e., entities which are instances of controlled vocabulary terms), as well as the edit log table, where changes to database entities are recorded. This schema hosts the user-created business data (i.e., the actual database records). “Ontology” is the schema used to store the resource graphs, loaded programmatically by the user as CSV files at the time of deployment. “Concepts” is used to store reference data (or thesauri), loaded programmatically by the user as CSV files at the time of deployment. And the “aux” schema is used to store geospatial data pertaining to cadastral zones.

This database architecture allows for the flexible deployment of any kind of user-created data model without the need to alter the underlying RDB tables. The six template data models (or resource graphs) included with the Arches HIP app are designed using CIDOC CRM classes and properties. These are: heritage resources (described as “culturally significant objects such as buildings and monuments”); heritage resource groups (spatial groupings of heritage resources, such as “districts”); activities (such as surveys and excavations); historic events (chronologically defined episodes of historical import, e.g. “The battle of Actium”); actors (i.e., people, groups of people, or institutions); and information objects (anything from photographs to bibliographic items) (Carlisle et al. 2014). To each of these graphs corresponds a dedicated data entry interface, available to the user via the “Resource Manager” tool.

The “Heritage Resource” graph, is structured around the CIDOC CRM class E18 (“Physical Thing,” which is defined as “all persistent physical items with a relatively stable form, man-made or natural,” <http://www.cidoc-crm.org/Entity/e18-physical-thing/version-6.2.1>). Following the criteria for data recording established by the CIDOC International Core Data Standard (CDS) for Archaeological and Architectural Heritage, this resource graph captures information concerning a heritage resource’s physical components, its measurements, condition, location, and chronology (FIGURE 3).

Many projects which have adopted Arches v.3 have tended to adapt this graph to represent virtually all their heritage resource types. For example, the Zbiva project mentioned above has developed three customized versions of the Arches E18 graph to represent sites, graves and objects. Zbiva “Site.E18” graph (https://github.com/bojankastelic/zbiva/blob/master/zbiva/source_data/resource_graphs/SITE.E18.pdf) is essentially a slimmed down version of the Arches “Heritage Resource” graph, with only the minimal addition of E55 (“Type,” <http://www.cidoc-crm.org/Entity/e55-type/version-6.2.1>) classes such as “Finding Type.E55,” where archaeological interpretations (e.g. cemetery, pit, fortification) are recorded. Other branches, such as that describing a site’s location (i.e., Place.E53 and its sub-nodes) have been left entirely untouched. A similar approach has been adopted by a university project aimed at creating a heritage inventory in Taiwan (Jihn-Fa 2016).

Early on in the life of the project, EAMENA adopted the same strategy, making only minimal changes to the template graphs provided with Arches HIP: the project had to develop a recording system very quickly and given the expertise of the team at that time, CRM data modeling was not identified as a priority. Given EAMENA’s emphasis on the use of remote sensing to identify heritage resources, the project decided to concentrate on adapting the “Heritage Resource Group”

Arches graph. This graph had the advantage of being modeled on the CIDOC CRM class E27, which “describes constellations of matter on the surface of the Earth” (<http://www.cidoc-crm.org/Entity/e27-site/version-6.2.1>). In contrast to E18, the centrality of spatial extent in the definition of the E27 class seemed poised to better capture the nature of what EAMENA identifies as sites, namely a wide range of morphologically different features or feature groups, which only have in common the fact of being visible, or being represented on satellite imagery, photographs and maps and being believed to be of heritage value (by image interpreters affiliated to the EAMENA project). During this first phase of data modeling, only small changes, mostly at the periphery of the template resources graphs were made, particularly pertaining to the addition of E55 nodes.

This reflected the nature of the EAMENA database as a thesauri-heavy database: thesauri correspond to E32 (“Authority Document,” <http://www.cidoc-crm.org/Entity/e32-authority-document/version-6.2.1>) classes in the CRM, and are used to populate E55 classes. For example, since we intended to record causes and effects of a range of disturbances at a site, we considerably expanded the E3 (“Condition State,” <http://www.cidoc-crm.org/Entity/e3-condition-state/version-6.2.1>) branch of the default Arches “Heritage Resource Group” graph, by creating a child E3 (which we termed “Disturbance State”) node, further characterized by a series of additional E55 nodes (and their corresponding E32); we did the same for our threat assessment, by creating a child E3 node (which we termed “Threat State”), once again supported by a set of E55 nodes. An example of this data modeling strategy is shown by Supplemental Material 1, where the structure of the EAMENA condition assessment branch (b) is compared with the standard Arches E18 condition assessment branch (a).

Further E55 nodes were added to several other components of the resource graph (among them: site function; site location; disturbance causes, and effects) in order to capture differing levels of belief in the accuracy (or “certainty”) of the data recorded. The reason for this is apparent: data collected via remote sensing is inherently subject to a high degree of error, which may be due to poor imagery resolution, ambiguous features, and the experience of the interpreters (Rayne et al. 2017).

The only other Arches resource graph which was substantially modified by the EAMENA team is the “Information Resource.E73” graph (built on the CRM class E73 “Information Object,” <http://www.cidoc-crm.org/Entity/e73-information-object/version-6.2.1>), which was adapted, once again mostly via the inclusion of new E55 nodes, to describe the four main information resource types used by the project: satellite and aerial imagery, cartography, bibliography, and shared datasets. As for the remaining three default resource graphs provided with the Arches HIP, only minor changes were made to the “Actor.E39” graph, while the project decided not to adopt the Activity and Historic Event graphs in its Arches deployment at this initial stage. The reasons for this choice are to be found in the fact that the recording of historic events does not currently feature in the project’s data recording strategy; a similar point may be made for activities, with the exception of condition assessment activities, which are embedded directly within the “Heritage Resource Group” graph and its corresponding data entry interface.

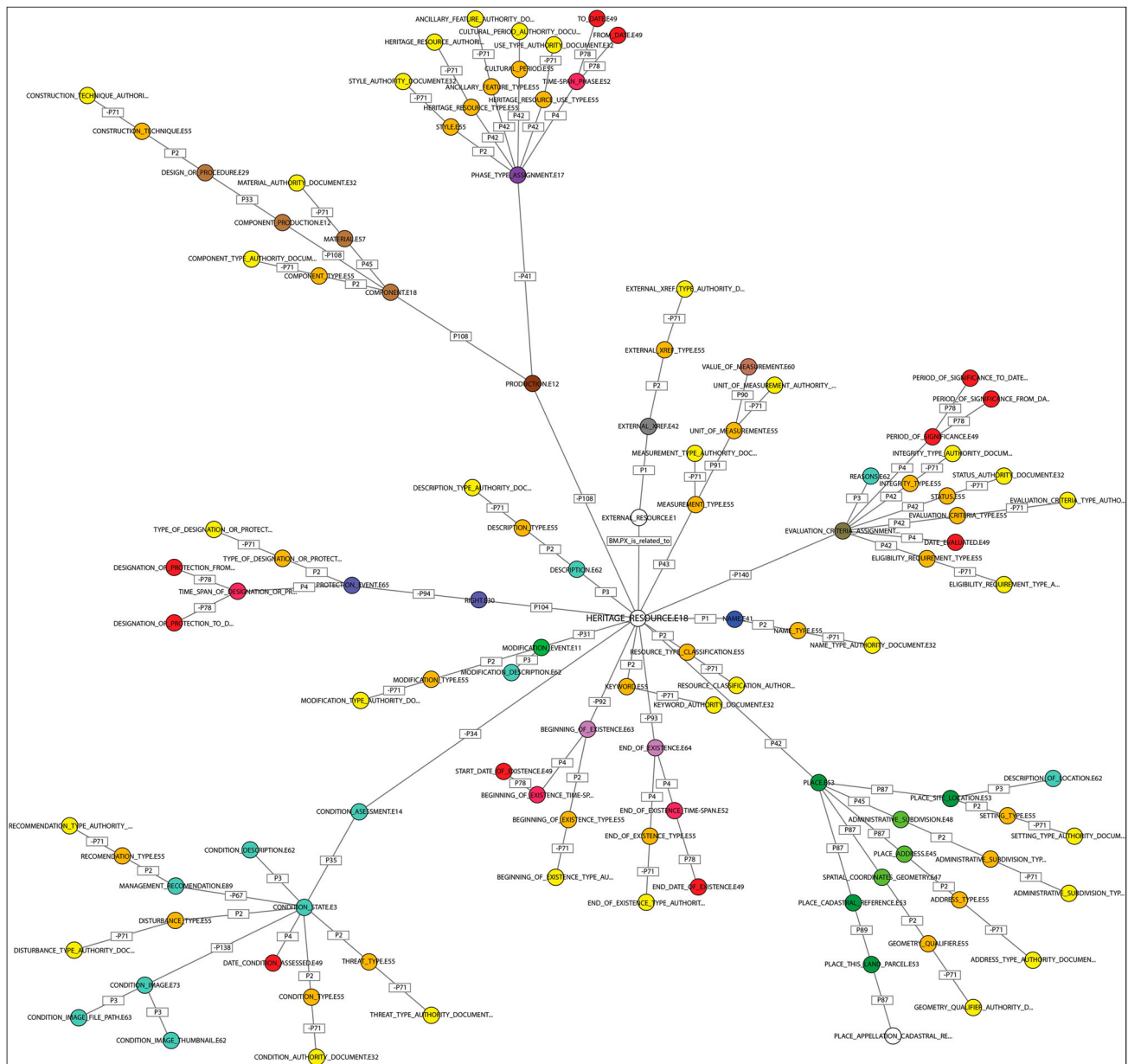


Figure 3. The Arches v.3 default “Heritage Resource” E18 graph © Getty Conservation Institute.

Starting in late-2016, a decision was made to restructure the EAMENA database with a view to incorporating more complex research datasets and, especially, to promoting its adoption as a heritage management platform in several of the countries where the project was set to deliver its CPF-funded training workshops. Among the key requirements that emerged was the need to develop a workflow to record field-based data to sit alongside the remotely-sensed data. In particular, the new workflow would have to allow the user to capture the changing function of site features through time (e.g. a Roman temple which is turned into a church in the Byzantine period and, eventually, into a mosque in the Umayyad period). As for condition assessments, field data presented the opportunity to narrow the assessment down to the level of components: in the case of a Roman temple, field-based assessments may lead to identify disturbances affecting its *propylaeum* (e.g. wind action causing erosion), which may be different from those affecting its *cella* (e.g. earthquake leading to collapse).

As this second phase of data modeling began, in response to the direction in which the project was going, we also

became increasingly aware of the need to review and replace some of the semantics of the Arches v.3 default graphs (drawn entirely from the CRM core) with more semantically pertinent classes and properties, specifically relevant for the needs of EAMENA, and outlined in the CRM’s approved specialist extensions. A case in point is that which pertains to “certainty” nodes, for which we had initially used E55 classes. Inasmuch as a level of certainty represents the belief of the assessor in the accuracy of the data that he or she is entering, “certainty” nodes may be characterized in terms of belief systems, a feature covered by the classes I2 (“Belief”) and its subclasses, which were introduced by the extension CRMInf (Stead and Doerr 2015).

Discussions with members of the CIDOC CRM community also led us to identify another problem area in our modelling of threat assessment, which we had borrowed from the original Arches v.3 E27 resource graph, where “Threat Types” are a E55 specialization of the site’s “Condition State” (E3). However, the very notion of condition state cannot be reconciled with an assessment of threats/risks to heritage resources: the latter is necessarily speculative and concerns the potential

state of a resource in the future, a dimension which current CIDOC classes and properties do not sufficiently cover.

Consequently, new ways in which to model threat, possibly via new classes and properties are now being considered. In addition, three new resource graphs have also been developed and are currently being tested, under the provisional names of “Heritage Place.E27”; “Heritage Feature.E24” and “Heritage Component.B2” (the semantics and details of these three graphs will be presented elsewhere in a joint paper with Jennie Bradbury and Azadeh Vafadari).

Heritage Place.E27 bears a number of similarities with its predecessor, but crucially, it replaces several E55 nodes with more semantically pertinent ones drawn from the CRM extensions CRMsci (for science and scientific metadata [Doerr et al. 2015]) and CRMinf (developed to describe inference making). EAMENA “Heritage Place” resources may be essentially characterized as collections of features of heritage significance located at a given place on the surface of Earth. Assessments of their chronology, morphology, archaeological interpretation, condition, and potential threats reflect, consequently, their nature as groupings: no effort is made, for example, to link a specific chronological phase present at a given “Heritage Place” with any of its features.

This latter job is accomplished by Heritage Feature.E24, a brand new resource graph. This graph is centered on class E24 (“Physical man-made thing,” <http://www.cidoc-crm.org/Entity/e24-physical-man-made-thing/version-6.2.1>), representing the site feature under investigation (e.g. a temple within a larger town; a record for the latter would have been created, in turn, using the Heritage Place.E27 resource graph). Instances of Heritage Feature.E24 are characterized by multiple phases of production (using the CRM class E12 “Production,” <http://www.cidoc-crm.org/Entity/e12-production/version-6.2.1>), each of which defined by a specific cultural period (an E55 class) and time span (E52, <http://www.cidoc-crm.org/Entity/e52-time-span/version-6.2.1>), and resulting in the assignment of a specific interpretation (a specialization of E14 “Type Assignment,” <http://www.cidoc-crm.org/Entity/e14-condition-assessment/version-6.2.1>).

Instances of Heritage Feature.E24 may also be linked to instances of the third EAMENA graph, “Heritage Component.B2,” via the CRM property P46 (“is composed of,” <http://www.cidoc-crm.org/Property/p46-is-composed-of/version-6.2.1>). Drawing on the CRMba extension for buildings archaeology (Ronzino et al. 2016), this final graph puts the specialized class B2 (“Morphological built section,” comprising “instances of man-made things that are considered functional units for the whole building,” Ronzino et al. 2016: 9–10) at its center and adopts its various subclasses to break heritage features (e.g. a temple) down into its components (e.g. the *propylaeum*, the *cella*, etc.) and their respective condition assessments.

Alongside the customization of resource graphs, a significant amount of work has gone into defining the terminology for use as the project’s reference data. The default Arches reference data is formatted using the Simple Knowledge Organisation System (SKOS), a W3C recommendation particularly suited to represent hierarchical thesauri (<https://www.w3.org/TR/2009/CR-skos-reference-20090317/skos.html>), with headings and subheadings, synonyms, and definitions. Arches HIP comes with a set of 42 pre-loaded thesauri (or controlled vocabularies, corresponding to CIDOC CRM E32 “Authority Document” classes), covering

domains such as cultural periods, disturbance and threat types, component types, eligibility requirements, and so on. Most of these were produced by the Getty Conservation Institute (31), often directly for the purposes of the Los Angeles Historic Resources Survey; five thesauri were supplied by the SurveyLA team, whilst English Heritage provided two thesauri (for cultural periods and measurement types) and the Dublic Core Metadata Initiative (DCMI) was drawn upon to provide the reference data for information resources.

Only a small part of this reference data was retained in the EAMENA database. The large remit of the project, which encompasses 20 archaeologically diverse countries from Morocco to Iran, meant that thesauri dealing with aspects such as cultural periods and archaeological interpretations had to be drastically expanded: by way of example, the standard Arches cultural period thesaurus (45 terms), designed by English Heritage and consequently British-specific, was replaced by a 450-strong thesaurus encompassing a wide spectrum of local and regional cultural periods from the Upper Palaeolithic to the Contemporary Period. As for the controlled vocabularies used to identify feature morphology and disturbance causes and effects, the project largely borrowed and expanded upon the terminology already developed by the Fragile Crescent Project mentioned above, a remote sensing project based at the University of Durham which concentrated on Syria and Iraq. Overall, the EAMENA database contains 1,770 controlled vocabulary terms and their definitions, most of which fully translated in Modern Standard Arabic (Mahdy and Zerbin 2016), with local Arabic variants also in the process of being recorded. This reference data may be exported as RDF by accredited users of the EAMENA database and re-deployed in another Arches environment. We plan to make all the full EAMENA reference data available on our website as RDF in the coming months (Supplemental Material 2).

Modifying the Arches Codebase: Front- and Back-End Development

Alongside data modeling and the extension of reference data, the EAMENA project has been modifying the Arches codebase in response to the needs of the project and, increasingly, of its partners in the MENA region. In compliance with Arches’ GNU Affero General Public License (<https://www.gnu.org/licenses/agpl-3.0.en.html>), all changes to the codebase implemented by the EAMENA staff are published online and are available for download on the project’s Github repository (https://github.com/azerbini/eamena_v3), to which the reader should refer for a detailed breakdown of the coding changes.

At the beginning of our implementation, changes were mostly limited to front-end design, particularly focusing on the HTML templates for the data entry forms, which needed to be adapted to the new resource graphs. From this process arose the need to ensure that the entire user interface could be displayed in multiple languages and particularly in Arabic, which is the main language spoken across the region in which the EAMENA project operates. Although the Django web framework natively supports the internationalization of any static text present in its front-end templates and media files, translations of dynamic text served via the Django database models (such as the EAMENA thesauri) could be stored, but not displayed by the default Arches HIP package. To

Perhaps the most significant change to the Arches codebase implemented thus far by the EAMENA team pertains to the development of a more advanced search toolkit (<https://github.com/azerbini/eamena> v3/pull/59). The default Arches



v.3 search, which is powered by Elasticsearch, allows only for the concatenation of tags in search queries bound by AND operators. OR searches are not possible, nor is it possible to combine AND-bound and OR-bound tags. Moreover, search queries are performed across the whole resource tree: it is impossible to restrict the scope of a query to, for example, only the data recorded as part of a resource's condition assessment. Because of this limitation, for instance, if one wished to search for all EAMENA sites where "Clearance (bulldozing/levelling)" (a term recorded in the "disturbance causes" thesaurus) has produced a "Loss of archaeological material" (part of the "disturbance effects" thesaurus), the resulting search query would yield not only the expected results, but also any site in which the disturbance cause "Clearance (bulldozing/levelling)" and the disturbance effect "Loss of archaeological material" appear, even though they may not be necessarily linked one to the other (a "false positive" is displayed below, see [FIGURE 5](#)).

The new advanced search toolkit designed by EAMENA addresses these issues, by introducing the ability to combine multiple search bars functioning as query subsets bound by a combination of AND/OR operators. Each search bar provides the option to restrict the scope of a query subset to a specific branch of a given resource tree. In this way, highly sophisticated search queries can be built, such as the one presented in [Figure 6](#).

Many other features are currently being developed, particularly in the field of bulk data management, with a view to facilitating the manipulation of large datasets (e.g. bulk resource creation and editing via shapefile, MS Excel workbook, etc.; group deletion of resource and reference data). Moreover, in order to better integrate field-acquired data, EAMENA is in the process of developing interoperability tools that will allow it to acquire data from mobile recording apps such as AMAL, a tool designed by the Global Heritage Fund to collect rapid impact assessments of damaged heritage sites (Rouhani 2016). The release of a mobile version of

Arches v.4, scheduled for 2018, will also dramatically increase our ability to integrate the desk-based, remotely-sensed assessments of heritage sites conducted by EAMENA with ground-checking and more detailed data collection on the ground.

Conclusion: EAMENA and National Inventories in the MENA region

In three years, the EAMENA project has grown in scope and ambition far beyond the remits devised at the time of its inception. In early 2015, the emphasis was placed on rapidly building a documentation base on the endangered cultural heritage of the MENA region, drawing on the expertise of EAMENA's archaeologists using satellite imagery as their source material. The wide range of tools provided by Arches v.3 out of the box, its user-friendly interface and support of CIDOC CRM resource graphs made selecting it as our platform a natural choice. Similarly, during the first phase of data modeling, our decision to implement only one heritage resource graph (the expanded "Heritage Resource Group.E27") may be explained in the light of the project's desire to quickly create a large amount of relatively thin-layered summary records rather than a smaller number of very detailed site reports.

The goals of the project have since broadened, as EAMENA began to establish contacts in the countries in which it operates and, crucially, since the receipt of a Cultural Protection Fund grant, which will enable the project to train up to 140 MENA-based heritage managers in remote sensing and in using the EAMENA database. With a team of twenty researchers and heritage specialists, EAMENA is now seeking to build on the momentum of its training workshops in order to create the conditions for the establishment of bespoke national heritage inventories (or HERs).

Many of the aspects of data modeling and codebase development discussed in this paper are essential to strengthen the case for the adoption of the EAMENA platform as the baseline

Disturbance state defined records:

- ✖ Clearance (Bulldozing/Levelling) (cause), High (cause certainty), Covered (effect 1), High (effect 1 certainty), Artefact Displacement (effect 2), High (effect 2 certainty), (effect 3), (effect 3 certainty), (effect 4), (effect 4 certainty), (effect 5), (effect 5 certainty), Development (type) Occurred between (date type), 2003-11-18 (start), 2004-08-14 (end)
- ✖ Vegetation/Crops (cause), High (cause certainty), Covered (effect 1), High (effect 1 certainty), Collapse/Structural Damage (effect 2), Low (effect 2 certainty), (effect 3), (effect 3 certainty), (effect 4), (effect 4 certainty), (effect 5), (effect 5 certainty), Agricultural/Pastoral (type) Occurred before (date type), 2003-09-25 (start), (end)
- ✖ Excavation (Hand) (cause), Definite (cause certainty), Relocation of Archaeological Features (effect 1), Definite (effect 1 certainty), (effect 2), (effect 2 certainty), (effect 3), (effect 3 certainty), (effect 4), (effect 4 certainty), (effect 5), (effect 5 certainty), Archaeological (type) Occurred before (date type), 2003-09-25 (start), (end)
- ✖ Construction (cause), High (cause certainty), Covered (effect 1), High (effect 1 certainty), Earth Displacement (effect 2), High (effect 2 certainty), Loss of Archaeological Material (effect 3), High (effect 3 certainty), Alteration of Terrain (effect 4), High (effect 4 certainty), (effect 5), (effect 5 certainty), Development (type) Occurred between (date type), 2005-07-30 (start), 2007-03-29 (end)
- ✖ Construction (cause), High (cause certainty), Alteration of Terrain (effect 1), High (effect 1 certainty), Covered (effect 2), High (effect 2 certainty), Earth Displacement (effect 3), High (effect 3 certainty), Loss of Archaeological Material (effect 4), Medium (effect 4 certainty), (effect 5), (effect 5 certainty), Development (type) Occurred between (date type), 2005-01-13 (start), 2005-05-06 (end)
- ✖ Vegetation/Crops (cause), High (cause certainty), Covered (effect 1), High (effect 1 certainty), Collapse/Structural Damage (effect 2), Low (effect 2 certainty), (effect 3), (effect 3 certainty), (effect 4), (effect 4 certainty), (effect 5), (effect 5 certainty), Agricultural/Pastoral (type) Occurred between (date type), 2003-09-25 (start), 2016-01-07 (end)

Figure 5. The EAMENA report for Khirbet Yajuz in Jordan, where "Loss of archaeological material" is associated with a "Construction" disturbance cause rather than with "Clearance (bulldozing/levelling)."

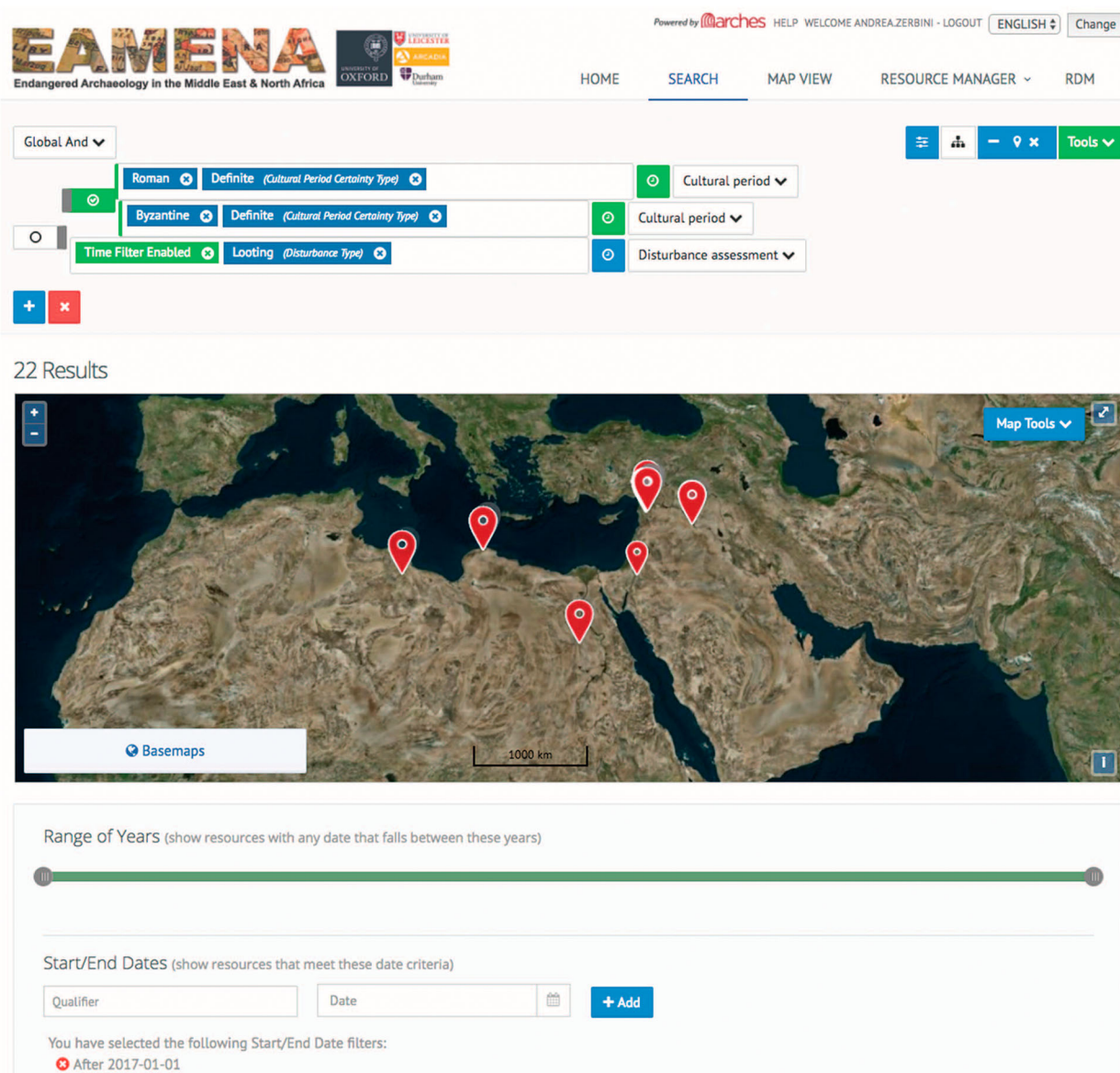


Figure 6. An example of a search query using the EAMENA advanced search toolkit. The query exposes sites with a Roman or Byzantine phase which have been looted after 1st January 2017.

framework for these national HERs. On this front, substantial progress has already been achieved in Yemen where, under the overall coordination of UNESCO (Doha office), EAMENA has been teaming up with the General Organisation for Antiquities and Museums based in Sanaa and with the international archaeological community to create a heritage inventory for the country. Drawing upon the large dataset collected by the EAMENA team via remote sensing (ca. 40,000 sites) as well as field-collected data provided by archaeologists, a stand-alone platform provisionally named YHMP (Yemen Heritage Management Platform) has been developed. Currently, neither the reference data nor the codebase of this platform present any differences with EAMENA's. However, in a meeting held in Amman in August 2017, a technical team made up of IT experts and archaeologists from Yemen has drawn up detailed feedback, highlighting areas where coding and database structural changes will need to be made in order to make this platform more suited to the needs of the Yemeni authorities. Requested changes range from the deletion of redundant reference data (e.g. cultural periods deemed irrelevant for Yemen) to improvements to the Arabic-language user interface, from the ability to create and download versioned site reports (timestamped

whenever a change is made to a site record) to that of adding a data entry form for interventions and recommendations.

The establishment of national HERs based on the EAMENA methodology of recording may prove to be a cornerstone in shifting heritage management practices in the MENA region, and a unique opportunity for the project to achieve lasting impact: it would not only ensure that inexpensive remote sensing techniques such as those promoted by EAMENA are embedded in the daily workflow of heritage managers in the Middle East and North Africa, but also guarantee that the data collated and created by EAMENA staff during the five years of the project continues to be updated and improved on.

Acknowledgments

An earlier version of this paper has appeared in the *Annals of the International Society of Photogrammetry and Remote Sensing* (Sheldrick and Zerbini 2017). The author wishes to thank all of the members of the EAMENA team (<http://eamena.arch.ox.ac.uk/meet-the-team/>) for their work and contributions, and particularly Jennie Bradbury and Azadeh Vafadari (who have been leading the data modeling work), Andrew Haith, Markos Ntounpanakis, and Richard Jennings, who have been

involved in the codebase changes described in this paper. Thanks are also due to the School of Archaeology of the University of Oxford. Any errors remain my own.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

The EAMENA project is based at the Universities of Oxford, Leicester, and Durham, and is generously supported by the Arcadia Fund (<https://www.arcadiahfund.org.uk/>) and the British Council's Cultural Protection Fund (<https://www.britishcouncil.org/arts/culturedevelopment/cultural-protection-fund>).

Notes on Contributors

Andrea Zerbin (PhD 2013, University of London) is an IT/GIS Officer and a Research Associate on the Endangered Archaeology in the Middle East and North Africa (EAMENA) project based at the University of Oxford. Prior to this, he held a CBRL Visiting Fellowship at the British Institute in Amman and a Fondation Fyssen Postdoctoral Fellowship affiliated with the CNRS team 'Archéologies et Sciences de l'Antiquité' based at the Université Paris X - Nanterre. From May 2018, he has assumed the role of Assistant Director of CBRL - The British Institute in Amman.

ORCID

Andrea Zerbin  <http://orcid.org/0000-0003-0043-6691>

References

- Banks, R., and A. Zerbin. 2015. "The Madaba Ring Road, Jordan. Evidence of Cultural Heritage Assets from Remote Imagery in the Madaba Hinterland." http://eamena.arch.ox.ac.uk/wp-content/uploads/EAMENA_Madaba_Report.pdf.
- Bernstein, K., and J. Hansen. 2016. "SurveyLA: Linking Historic Resources Surveys to Local Planning." *Journal of the American Planning Association* 82 (2): 88–91. doi:10.1080/01944363.2015.1137199.
- Bewley, R., and M. Fradley. 2017. "The EAMENA Aerial Photograph Appeal." <http://eamena.arch.ox.ac.uk/the-eamena-aerial-photograph-appeal/>.
- Bewley, R., and D. Kennedy. 2012. "Historical Aerial Imagery in Jordan and the Wider Middle East." In *Archaeology from Historical, Aerial and Satellite Archives*, edited by W. S. Hanson and I. A. Oltean, 221–242. New York: Springer.
- Bewley, R., A. I. Wilson, D. Kennedy, D. Mattingly, R. Banks, M. Bishop, J. Bradbury, et al. 2016. "Endangered Archaeology in the Middle East and North Africa: Introducing the EAMENA Project." In *CAA2015. Keep the Revolution Going: Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology*, edited by S. Campana, R. Scopigno, G. Carpentiero, and M. Cirillo, 919–932. Oxford: Archaeopress.
- Binding, C., K. May, and D. Tadhope. 2008. "Semantic Interoperability in Archaeological Datasets: Data Mapping and Extraction Via the CIDOC CRM." In *ECDL 2008: Research and Advanced Technology for Digital Libraries*, edited by B. Christensen-Dalsgaard, D. Castelli, B. Ammitzbøll Jurik, and J. Lippincott, 280–290. New York: Springer.
- Carlisle, P., I. Avramides, A. Dalgity, and D. Myers. 2014. "The Arches Heritage Inventory and Management System: A Standards-based Approach to the Management of Cultural Heritage Information." http://www.cidoc-crm.org/sites/default/files/I-1_Carlisle_Dalgity_et-al_paper.pdf.
- Carlisle, P., and E. Lee. 2016. "Recording the Past: Heritage Inventories in England." *Journal of Cultural Heritage Management and Sustainable Development* 6: 128–137. doi:10.1108/JCHMSD-02-2016-0013.
- Doerr, M., A. Kritsotaki, Y. Rousakis, G. Hiebel, and M. Theodoridou. 2015. "Definition of the CRMsci An Extension of CIDOC-CRM to Support Scientific Observation." <http://www.ics.forth.gr/isl/CRMext/CRMsci/docs/CRMsci1.2.3.pdf>
- Felicetti, A., T. Scarselli, M. L. Mancinelli, and F. Niccolucci. 2013. "Mapping ICCD Archaeological Data to CIDOC-CRM: the RA Schema." In *Practical Experiences with CIDOC CRM and its Extensions (CRMEX 2013) Workshop, 17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013), Valetta, Malta, September 26, 2013*, edited by V. Alexiev, V. Ivanov, and M. Grinberg, 11–22. <http://ceur-ws.org/Vol-1117/paper2.pdf>.
- Galiatsatos, N., T. J. Wilkinson, D. N. M. Donoghue, and G. Philip. 2009. "The Fragile Crescent Project (FCP): Analysis of Settlement Landscapes Using Satellite Imagery." In *Making History Interactive. Computer Applications and Quantitative Methods in Archaeology (CAA). Proceedings of the 37th International Conference, Williamsburg, Virginia, United States of America, March 22–26*, edited by B. Frischer, J. Webb Crawford, and D. Keller, 1–15. Oxford: Archaeopress. http://archive.caaconference.org/2009/articles/Galiatsatos_Contribution365_c.pdf.
- Gromley, C., and Z. Tong. 2015. *Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine*. Sebastopol, CA: O'Reilly.
- Henninger, M. 2017. "From Mud to the Museum: Metadata Challenges in Archaeology." *Journal of Information Science* 1–14. doi:10.1177/0165551517741790.
- Jihn-Fa, Jan. 2016. "Digital Heritage Inventory using Open Source Geospatial Software." In *22nd International Conference on Virtual System & Multimedia (VSMM)*, edited by H. Thwaited, Lau Sian Lun and A. Addison, 229–236. Piscataway, NJ: Institute of Electrical and Electronics Engineers. doi: doi:10.1109/VSMM.2016.7863192.
- Lawrence, D., J. Bradbury, and R. Dunford. 2012. "Chronology, Uncertainty and GIS: A Methodology for Characterising and Understanding Landscapes of the Ancient Near East." In *Landscape Archaeology. Proceedings of the International Conference Held in Berlin, 6th–8th June 2012, Special Volume 3*, edited by W. Bebermeier, R. Hebenstreit, E. Kaiser, and J. Krause, 353–359. Berlin: Excellence Cluster Topoi.
- Mahdy, H., and A. Zerbin. 2016. "Translating EAMENA's platform into Arabic: Challenges and Solutions." <http://eamena.arch.ox.ac.uk/translating-eamenas-platform-into-arabic-challenges-and-solutions/>.
- Mattingly, D. J., V. Leitch, C. N. Duckworth, A. Cuenod, M. Sterry, and F. Cole. 2017. *Trade in the Ancient Sahara and Beyond*. Trans-Saharan Archaeology. Cambridge: Cambridge University Press.
- Myers, D. 2016. "Heritage Inventories: Promoting Effectiveness as a Vital Tool for Sustainable Heritage Management." *Journal of Cultural Heritage Management and Sustainable Development* 6: 102–112. doi: doi:10.1108/JCHMSD-02-2016-0009.
- Myers, D., A. Dalgity, and I. Avramides. 2016. "The Arches Heritage Inventory and Management System: A Platform for the Heritage Field." *Journal of Cultural Heritage Management and Sustainable Development* 6: 213–224. doi:10.1108/JCHMSD-02-2016-0010.
- Myers, D., A. Dalgity, I. Avramides, and D. Wuthrich. 2012. "Arches: An Open Source GIS for the Inventory and Management of Immovable Cultural Heritage." In *EuroMed 2012: Progress in Cultural Heritage Preservation*, edited by M. Ioannides, D. Fritsch, J. Leissner, R. Davies, F. Remondino and R. Caffo, 817–824. New York: Springer.
- Rayne, L., J. Bradbury, D. Mattingly, G. Philip, G. Bewley, and A. Wilson. 2017. "From Above and on the Ground: Geospatial Methods for Recording Endangered Archaeology in the Middle East and North Africa." *Geosciences* 7 (4). doi:10.3390/geosciences7040100.
- Rayne, L., N. Sheldrick, and J. Nikolaus. 2017. "Endangered Archaeology in Libya: Recording Damage and Destruction." *Libyan Studies* 48: 23–49. doi:10.1017/lis.2017.7.
- Ronzino, P., F. Nicolucci, A. Felicetti, and M. Doerr. 2016. "Definition of the CRMba An Extension of CIDOC CRM to Support Buildings Archaeology Documentation."
- Rouhani, B. 2016. "Conflict-affected Countries in MENA Region. Training and Emergency Management Tools." *Economia della Cultura* 2: 335–340. doi:10.1446/84510. http://www.ics.forth.gr/isl/CRMext/CRMba/docs/CRMba_v1.4.pdf.
- Sivak, A. 2017. "Using Open-Source Software to Protect English Heritage." *The Iris*, January 5. <http://blogs.getty.edu/iris/using-open-source-software-to-protect-english-heritage/>.
- Stead, S. and M. Doerr. 2015. "CRMinf: the Argumentation Model. An Extension of CIDOC-CRM to Support Argumentation." <http://www.ics.forth.gr/isl/CRMext/CRMinf/docs/CRMinf-0.7.pdf>.
- Zerbin, A., and M. Fradley. 2018. "Higher Resolution Satellite Imagery of Israel and Palestine: Re-assessing the Kyl-Bingaman Amendment." *Space Policy*.