During the past decade, the Middle East and North Africa have witnessed a surfeit of remote, geospatial data collection projects, resulting in big databases with powerful deductive capacities. Despite the valuable insights and expansive evidentiary record offered by those databases, emphasis on anthropogenic threats to cultural heritage combined with a limited integration of local perspectives, have raised important questions on the ethical and epistemological dimensions of big data. This paper contextualises maritime cultural heritage in those debates through the lens of the Maritime Endangered Archaeology in the Middle East and North Africa project (MarEA). MarEA is developing a unique for the region large database for maritime archaeological resources, associated environment, and threats. This database is designed to amalgamate a baseline record with emphasis on spatial location, state of preservation and vulnerability of maritime cultural landscapes. This record will form a steppingstone toward finer grained research on maritime cultural heritage and its interdisciplinary intersections. It is also developed as an information resource that will facilitate local collaborators to prioritise site monitoring and develop documentation, management and mitigation strategies.

Suggested Reviewers:

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author Comments:</td>
<td></td>
</tr>
<tr>
<td>Suggested Reviewers:</td>
<td>Jesse Casana Dartmouth College</td>
</tr>
</tbody>
</table>
Jesse Casana has worked and published extensively on geospatial remote sensing in the Middle East.

Andrew Dufton
The University of Edinburgh

Andrew Dufton co-edited the 2020 special issue on Big Data and his area of research is North Africa.

Damian Robinson
Oxford University: University of Oxford

Damian Robinson is a maritime archaeologist working in the Eastern Mediterranean.

**Funding Information:**

<table>
<thead>
<tr>
<th>Arcadia Fund</th>
<th>Lucy Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Society for Libyan Studies</td>
<td>Julia Nikolaus</td>
</tr>
</tbody>
</table>
Big Data in Maritime Archaeology: Challenges and Prospects from the Middle East and North Africa

Andreou, G.M.\textsuperscript{a,c} J. Nikolaus\textsuperscript{b}, K. Westley\textsuperscript{b}, C. El Safadi\textsuperscript{a}, L. Blue\textsuperscript{a}, A. Smith\textsuperscript{a}, C. Breen\textsuperscript{b}

\textsuperscript{a}University of Southampton, Southampton, United Kingdom

\textsuperscript{b}Ulster University, Coleraine, United Kingdom

\textsuperscript{c}University College London, London, United Kingdom
Big Data in Maritime Archaeology: Challenges and Prospects from the Middle East and North Africa
Abstract

During the past decade, the Middle East and North Africa have witnessed a surfeit of remote, geospatial data collection projects, resulting in big databases with powerful deductive capacities. Despite the valuable insights and expansive evidentiary record offered by those databases, emphasis on anthropogenic threats to cultural heritage combined with a limited integration of local perspectives, have raised important questions on the ethical and epistemological dimensions of big data.

This paper contextualises maritime cultural heritage in those debates through the lens of the Maritime Endangered Archaeology in the Middle East and North Africa project (MarEA). MarEA is developing a unique for the region large database for maritime archaeological resources, associated environment, and threats. This database is designed to amalgamate a baseline record with emphasis on spatial location, state of preservation and vulnerability of maritime cultural landscapes. This record will form a steppingstone toward finer-grained research on maritime cultural heritage and its interdisciplinary intersections. It is also developed as an information resource that will facilitate local collaborators to prioritise site monitoring and develop documentation, management, and mitigation strategies.

Keywords: big data, maritime archaeology, cultural heritage, Middle East, North Africa, Arabia
1. Introduction

**Big data** is a term increasingly used since the 1990s to describe large volumes of data managed and processed through IT architecture. Due to their volume, big data are considered to have a higher resulting performance and variety, as demonstrated in information technologies and corporate environments (Tang 2016). Widely used to inform decision-making and forecasting in business analysis, marketing, transportation and, increasingly, environmental monitoring (e.g., Gorelick et al. 2017), the emergence of big data in archaeological practice in the past decade has been slower and received with both enthusiasm and scepticism (Kitchin 2014; Gattiglia 2015; Green 2020; *Journal of Field Archaeology* 45, S1).

In a recent special issue in the *Journal of Field Archaeology*, Van Valkenburgh and Dufton (2020) offer a thoughtful definition of big data, with emphasis on the adjective “big” used not solely to refer to the scale or dimension of data, but primarily on the extensive possibilities they afford for archaeological analysis. Contributions to that issue generally agree that big data have changed the scope of archaeological research, aiming for complete datasets and accelerating research output. Ongoing conversations on big data in archaeology tend to discuss geospatial technological and methodological advances (satellite and airborne remote sensing, machine learning) that have opened the way for data collection beyond the scale of regional surveys and have enabled more nuanced interregional perspectives. The ensuing collection of unprecedented amounts of data has led to critical questions surrounding the meaning, quality and ethics in the remote collection and analysis of archaeological information (Fisher et al. 2021).

The Middle East and North Africa (MENA) region, an area with a longstanding presence in historical and archaeological discourse, has been at the forefront of analyses relying on remotely-sensed big data. Originally collected with the aim to rapidly and remotely monitor inaccessible archaeological sites, geospatial approaches in the MENA region have developed into a theoretically encompassing sub-discipline (Lawrence et al. 2020). The MENA region also offers a rapidly evolving methodological scope for the development of automated detection of sites (Casana 2014; Liss et al. 2017; Soroush et al. 2018), features (Brady et al. 2017; Orengo et al. 2020), and even potsherds (Orengo and Garcia-Molsosa 2019).

The proliferation of these methodologies in archaeological practice has resulted in an expansive corpus of cultural heritage databases (discussed below). The data captured vary
according to intended use, ranging from gazetteers of archaeological sites to databases that register threats impacting the archaeological record. In the latter, there is an apparent propensity to document threats relating to conflict and looting - a recent example being Nagorno Karabakh in Armenia (Khatchadourian et al. 2021). Less attention has been placed on producing databases that detail the impact of non-anthropogenic factors on cultural heritage in the MENA region, notably climate change. The benefits of such databases are obvious, especially to local communities, who will experience, and have to respond to, the detrimental impacts of climate change (Stahl 2020). In addition, disproportionately little attention has been placed on the large-scale documentation of maritime (underwater, nearshore and coastal) cultural heritage, even though the Mediterranean Sea, the Red Sea, the Gulf and the Arabian Sea figure conspicuously in major archaeological narratives from hominin migration to contemporary archaeology. Even though maritime cultural heritage (MCH) is equally if not more impacted by anthropogenic factors and more actively exposed to natural processes (Bennet et al. 2004; Galili and Rosen 2010; Poukermann et al. 2018; Reimann et al. 2018; Trakadas 2020; Brooks et al. 2020), it has been integrated to a lesser extent in ongoing discourses on endangered archaeology in the MENA region (e.g., Galili et al. 2018).

This paper examines the challenges and prospects of big data approaches in maritime archaeology through the lens of the *Maritime Endangered Archaeology of the Middle East and North Africa* project (MarEA), a 5-year project focusing on the collection of archaeological and environmental information on vulnerable maritime archaeological sites in the MENA region (FIGURE 1) (Andreou et al. 2020). Within this research scope, we use the broad definition of maritime archaeology as evidence for pre-modern human engagement with the sea, (exemplified by Westerdahl 1992; Ransley et al. 2013; Atkinson and Hale 2012), instead of more restrictive classifications, such as underwater or nautical archaeology. The evidence base, thus, can be found both on land and underwater (submerged) and could include anything from small objects to large vessels and structures, such as harbours.

We will first discuss the role and use of big data with respect to the archaeology of the MENA region, followed by some thoughts on associated ontological and epistemological challenges. Next, we will explore the ethical dimensions of collecting and managing data outside of their geographical origin. We subsequently expound on challenges particular to maritime archaeology, followed by a presentation of our data collection, documentation
practices, in-country partnerships and a demonstrative example of the applied use of the database in the context of disaster response.

2. Big data and the archaeology of the MENA region

Geospatial approaches have offered unparalleled insights into archaeological landscapes in the MENA region (Hammer and Ur 2019; Casana 2020), including documentation and monitoring of places exposed to damage and destruction, as well as locales with limited access by archaeologists. It is unsurprising then that an abundance of remote geospatial projects in this region have laid important theoretical and methodological groundwork for the documentation of endangered sites (e.g., Jakoby-Laugier and Casana 2017; Rayne et al. 2017; Rayne et al. 2020). Aerial and satellite imagery, most notably declassified Corona imagery has made a particularly important contribution, long serving as a baseline for archaeology in the Middle East. At a smaller scale, archaeologists have used satellite imagery and aerial photography to enhance the spatial characteristics of known features and sites (Parcak 2007; Hammer 2019) and to document and monitor site destruction (Fradley and Sheldrick 2017; Casana and Laugier 2017; Rayne 2017). At a broader scale, site documentation has also been conducted using automated classification (Ur 2013; Casana 2014; Soroush et al. 2018), offering promising results.

The above methodological developments resulted in the production of geographical and archaeological datasets, including the Corona Atlas of the Middle East (Casana and Cothren 2013), the Aerial Photographic Archive for the Archaeology in the Middle East (APAAAME) (Bewley et al. 2010), the Digital Archaeological Atlas of the Holy Land (DAAHL) (Savage and Levy 2014), the database of the Endangered Archaeology of the Middle East and North Africa (EAMENA) (Bewley et al. 2016), the MEGA-Jordan (Drzewiecki and Arinat 2017), DocArtis (www.docartis.com) and a national database (www.inp.tn/cnsa) for Tunisia, and a heritage gazetteer (www.slsgazetteer.org) for Libya. Many such datasets focus on documenting what has been identified as looting (summary in Kersel and Hill 2020), the remote identification of which is well-established in Afghanistan (Franklin and Hammer 2018), Egypt (2015), Iraq (Hritz 2008, Hanson 2012), Jordan (Contreras and Brodie 2010) and Syria (Casana and Panahipour 2014; Cunliffe 2014; Tapete et al. 2016). Meanwhile, equally pressing and detrimental factors, including demographic pressures and climate
change tend to be less visible in existing big archaeological datasets, though attempts are underway to address some of those issues (Rayne et al. 2020; Westley et al. 2021).

Increasing conversations about the use and analysis of, and infrastructure development for, big data in archaeology urges us to think the impact of large datasets on data collection and narrative production. The epistemological advancements associated with geospatial big data inevitably come with empirical, conceptual and ethical limitations, with associated challenges including sample bias toward larger, better discernible features, as well as questions surrounding the intellectual ownership of remotely produced or amalgamated geographical and archaeological data.

3. Ontological challenges

An important benefit of big data is the unprecedented analytical capacity afforded by the use of standardised data classification, such as the CIDOC-CRM standard ontological model for cultural heritage (Crofts et al. 2008), which facilitates comparative studies and understanding of broader patterns. Big data also allow a more effective organisation, communication and retrieval of information. Depending on the ontologies used in each database, they can enable different combinations of information, including data previously embedded in catalogues or archives with more limited capacities for access and analysis. In the context of archaeology, big datasets participating in linked open data (Candela et al. 2018) allow researchers to produce archaeological assemblages that are not limited by political, geographical or contextual boundaries.

To take advantage of this, hosting projects or institutions require experts with diverse specialisations, and a constant upgrade of technological systems and infrastructure. Moreover, investment in time and expertise on issues surrounding data interoperability are necessary to facilitate information extraction (Vlachidis and Tudhope 2015), data discovery and accessibility (e.g., ARIADNE – Meghini et al. 2017) and conceptual referencing among different regions (Binding and Tudhope 2016; Henninger 2017).

Reasonable critique has arisen on the use of rigid categorisations (e.g., Gupta et al. 2020), particularly when observations are characterised by qualitative aspects. For example, during data collection via satellite imagery analysis, researchers will use their distinct knowledge, training and field experience to identify and interpret visible anomalies potentially without
being able to conduct exhaustive research on the ontological dimensions of applied terminology. Critical discussions are also directed at the proliferation of data classifications that do not engage with indigenous ontologies and alternative understandings of cultural heritage. We explore this issue in a section below.

Overall, while the use of big data combined with consistent terminology allow for transregional analyses (McCoy 2020) and enables data integration of diverse archaeological assemblages, terminological standardisation can mask local and regional variations. Those variations, however, are better targeted through a finer-grained examination. This can be facilitated in terms of time and financial investment, when informed by patterns highlighted though the analysis of satellite-derived big data. On the south-central coast of Cyprus, for example, large-scale documentation of erosion produced a vulnerability model that highlighted archaeological landscapes experiencing aggravated land loss (Andreou 2018). A finer-grained examination directed by the model output and in-situ visits, have documented eroding and newly-exposed remains (Andreou et al. 2019). Once we move beyond the misguided expectation that big data consist of complete datasets, it is possible to identify ways to maximise their potential even in challenging contexts.

4. Epistemological challenges

Though often viewed as a new epistemological paradigm (Kitchin 2014), the implications of big data in archaeology have raised concerns (Huggett 2020), particularly the widespread misconception that large amounts of data equal higher precision or improved interpretations. Critique has focused on surmises on the validity, deductive capacity and credibility of analyses resulting from big data (e.g. Lohr 2015; McCoy 2017). In turn, the increase in the number of identified archaeological sites has created the need for robust quality control strategies and enhancement of metadata standards detailing the methodologies used for site identification (Opitz and Herrman 2018; McCoy 2017, 2020). When it comes to the remote collection of large-scale archaeological datasets, three types of approaches are used widely, each with its epistemological challenges.

The first approach is expert-led site identification, as used by the Oriental Institute and the EAMENA project. Casana (2014) used the term “brute force” to refer to teams of specialised researchers that scan images and tag features. This approach is considered to provide as informed and accurate as possible remote archaeological observations (Casana 2014), the
effectiveness of which for site identification and minimised inter-observer variability, has been demonstrated via ground-verification (Casana 2020). MarEA, has adopted this approach and supports a central research team leading on initial imagery-based site identification combined with bibliographic research and geographical observations which can be followed-up by in-country surveys as effective means of quality control.

A second approach focuses on crowdsourcing data generated by untrained eyes. Perhaps the most known and controversial case is GlobalXplorer®, an online platform that trains non-specialists to identify sites and even looting from satellite imagery (Parcak 2019). Beyond the ethical shortcomings of such approaches (discussed in the following section), the production of a large database by non-experts and without a clearly stated quality control strategy, has raised concern on the unregulated production of thousands of false positives, which can only be minimised via field-based observations (McCoy 2020; Casana 2020). Without concrete quality control measures, crowdsourced big archaeological datasets suffer from important epistemological limitations.

A third approach is automated feature detection informed by training data collected and cross-examined by experts. Depending on the quantity and quality of that data, inherent biases in the training sample will be reproduced in the automated detection process (Ringer and Loschky 2018; Hoffmann 2018). As such, appropriate training and supervision are required to control those biases, as well as to identify the impact of potential inter-observer variability. It is important to note, however that the size of the resulting data is one of the key limitations that prevents its systematic quality control (Casana 2020, S95; Savage et al. 2017; and more broadly Woodall et al. 2014). Reasonable questions have been raised on how quality control can be undertaken, particularly when archaeological publications do not offer exhaustive documentation of their digital data collection and classification strategies (Caraher 2016), to the extent that some have argued that digital and/or remote collection and analysis methodologies are often not understood in satisfactory depth by their users (Kvamme 2018, 75). In other words, large amounts of data combined with limited understanding of the collection methodologies and with an under-theorised approach to heritage, likely enable less critically examined narratives.

Overall, even though automated approaches are considered by some to be the way forward (Orengo et al. 2020), systematic, intensive and detailed expert-led analysis of satellite imagery, combined with bibliographic research and field-based observations by local
collaborators undeniably offers more nuanced and contextualised understandings of complex archaeological landscapes. Though resulting in relatively smaller volumes of data, brute force facilitates the collection of more accurate data and less dubious results. A combination of brute force and automated approaches presents important potential in the maximisation of the deductive capacity of big data, particularly when brute force-derived data are used as training samples for automated approaches.

5. Ethical Considerations

The establishment of large databases in the MENA region has fostered conversations on the ethics and politics of globalised archaeological databases. Beyond important discussions on the ethics of digital archaeology (Dennis 2020), concerns pertain to the remote collection, management, and distribution of big data (VanValkenburg and Dufton 2020). When projects derive data from formerly colonised countries and/or locations devastated by war, or administered with limited resources, postcolonial critique has turned the discussion toward the imbalanced relations between those who collect and analyse data (experts) and those who live in the geographic areas from where the data originate. Particularly, the globalisation of data relevant to the Global South through storing, management and curation outside of their geographic origin, is reminiscent of critiqued colonial practices (e.g., Azoulay 2019; Hicks 2020). With the MENA region at the forefront of conversations on postcolonial critique and archaeological ethics (Meskell 2020b), this section will discuss the issues most pertinent to the MarEA project. These include the use of remote sensing technologies to collect data and the use of data classification systems, both well-established methodologies in archaeological research and both with a long colonial legacy in the MENA region.

Many have discussed the panoptic view afforded by aerial imagery through its problematic historical links to colonialism and military surveillance (Hamilakis 2009; Pollock 2016; Pollock and Bernbeck 2018; Meskell 2020b), while others contrasted remotely-sensed data with the valuable, yet often marginalised proximate knowledge of local communities (McAnany and Rowe 2015; Mickel 2020). Fisher et al. (2021) also remind us that remote sensing offers overseas researchers access to geographical data, the collection of which traditionally required in-situ visits and permits from local authorities. Those permits are tied to local antiquities laws that often developed as a response to the exportation of antiquities to Western countries (Goode 2007). In other words, freely-available satellite imagery has
enabled researchers to bypass in-country oversight to collect data and document human practices (e.g., landscape alterations) without the informed consent of people living in those areas (Pollock and Bernbeck 2018; Meskell 2020b, 220-221). We consider that a mutual understanding of the benefits of remote sensing in monitoring archaeological sites can be fostered through the collaborative collection or production of information with local heritage practitioners that have a more contextual understanding of associated challenges.

It is also important to consider who the beneficiaries of big datasets are given that their archaeological contents are rarely entirely open access due to concerns that they may enable illicit excavations (Fernandez-Diaz et al. 2018; Parcak 2007). This issue becomes more acute, because questions of data ownership and legality, especially when data is derived from by freely-available imagery (Myers 2010a, 2010b; Scassa 2017), are an understudied topic in archaeology. In this context, the theoretical dimensions of big data have received less attention compared to the methodological and technological capacity of big databases.

Similarly, while acknowledging the advantages of digital documentation, some have voiced concern in the use of methods that are accessible mainly to those trained or living outside of the MENA region (Meskell 2018, 59-89; Rico 2017a; Stobiecka 2020). Rico (2017b, 742) highlighted potential power imbalances sustained between heritage experts and local communities, including local archaeology practitioners. Others have critiqued the under-theorised approaches of digital archaeology (Dallas 2015; Huggett 2015; Huvila and Huggett 2018; Dennis 2020), as well as the lack of ethical and historical reflection in digital heritage practices (Meskell 2020a; Stobiecka 2020). Although recent scholarship has demonstrated effective and mutually beneficial data collection and analysis (Kersel and Hill 2020; Gupta et al. 2020), one cannot ignore the absence of local archaeologists and institutions in charge of the data collation and the stewardship of their country’s heritage. It is unsurprising then that digital technocracy, combined with unclear strategies for local participation, especially in former colonies, has been received with scepticism by local scholarly communities (Abu-Khafajah and Miqdadi 2019).

In the same context, research has raised scepticism on the disembodied documentation and visualisation of sites through the digital gaze (Hamilakis 2014, 104-108). This issue becomes more evident with the reproduction of data classification and management practices that do not engage with indigenous, transcultural and diachronic understandings and ontologies of cultural heritage, and presume a universal heritage value (Hamilakis 2007, 2012; Harrison
The concepts of “preservation” or “conservation” (Hamilakis 2007; Meskell 2018; Rico 2015, 2020) and “endangered” or “at risk” (Rico 2020; Vidal and Dias 2015) are at the heart of many conversations on how digital archaeological practices tend to reflect Western understandings of what is worth being documented and preserved. Similarly, the conceptual processes of the classification of heritage as “endangered” or activities as “(il)legal” have been criticised for suggesting that external viewers (in many cases non-local and without in-depth understanding of local practices and policies) have knowledge of the degree to which local communities are engaging in activities that are not permitted by law in their own countries (Vanvalkenburg and Dufton 2020, S4).

As archaeology continues to expand its areas of intersection, it is important to reflect on inherent biases in heritage studies and examine how our contributions to the archaeology of the MENA region perpetuate or seek to alleviate, historical tensions. While we acknowledge that limited research has been published on local sentiments toward endangered heritage (e.g., Cunliffe and Curini 2018), we also observe in the recent years that digital documentation projects in the MENA region acknowledge problematic legacies in heritage studies (Fisher et al. 2021) and increasingly seek to ensure heritage rights, with discussions on capacity building, community engagement and participatory research (Khatchadourian 2020; LaBianca et al. 2020; Henderson et al. 2021).

In MarEA, we acknowledge that our data is disseminated in a way that enables their incorporation in the predominant discourse of archaeology, using vocabularies and classifications that are Western and largely Anglophone. We consider the use of these ontologies a crucial step to integrate our data with existing interdisciplinary conversations in the fields of heritage studies, maritime archaeology and coastal/marine science. We also acknowledge that the burden or managing heritage on the ground is on the shoulders of local archaeological communities and other stakeholders that often lack access to or input into those databases. Beyond the problematic scholarly legacies in the archaeology in the MENA, we also acknowledge the environmental impact of colonial and other interventions in the region, which contribute to ecological crisis, actively affecting local communities and their heritage. These include, among others, pollution resulting from the Suez Canal (El-Magd et al. 2020), sedimentary imbalances resulting from damming (e.g., Hzami et al. 2021), militarisation (e.g., Ragin and Riccò 2019), the politics of resource exploitation particularly
underwater (Stocker 2012), as well as the ecological footprint of fossil fuel infrastructure and exportation (e.g., Samargandi 2021).

With these challenges in mind, MarEA actively incorporates stakeholders (heritage specialists) based in the region, who contribute with their archaeological expertise and nuanced understanding of local perceptions of heritage and ontologies. Our local collaborators, primarily archaeologists based in the region, also have the opportunity via MarEA to train in maritime archaeological theory and methods, enhancing their skillsets for more effective monitoring of MCH. In this respect, MarEA aims to engage with recommendations in the context of community archaeology (LaBianca et al. 2020), including prioritising capacity building, investing in establishing relations of mutual understanding, affirming the role of local heritage stewards, and embracing multidisciplinary and multiagency cooperation.

6. Maritime archaeology in the MENA: data sources and challenges

Due to its geographical and environmental context, MCH is exceptionally vulnerable to natural processes including inundation and erosion (Erlandson 2012), as well as anthropogenic processes related to colonial and other interventions in the region, demographic expansion (Flatman 2009) and the development of coastal areas for tourism or industry. These challenges have in the last decade encouraged an expansion of methodological approaches focusing on the impact of climate change on MCH (Van de Noort 2013) and rapid documentation processes (e.g., McCarthy et al. 2019). In the MENA region, research has focussed on sustainable recording methodologies (El-Asmar et al. 2012; Andreou et al. 2017; Andreou 2018; Pourkerman et al. 2018), mitigation strategies (e.g., Bitan et al. 2020), the production of regional baseline data (Galili and Rosen 2010; Westley et al. 2021) and the consideration of inclusive spatial and environmental planning (e.g., Breen et al. 2021). Compelling work has also been conducted on capacity building for the documentation, study and preservation of MCH (Khalil 2008; Semaan 2018; Blue and Breen 2019; Demesticha et al. 2019), sometimes emphasising endangered MCH (Galili et al. 2018; Recinos and Blue 2019).

While recent scientific research has demonstrated the value of global models and classifications to assess, for instance, littoralisation (Neumann et al. 2015), sea-level fluctuations (Muis et al. 2016), flooding and erosion (Vousdoukas et al. 2018; 2020), research
on the MENA’s vulnerable MCH has largely been regionally or micro-regionally focussed. The limited and fragmentary basis of this research presents, therefore, significant challenges in developing MCH site documentation or monitoring strategies, which are crucial for the subsequent integration of the MCH resource into spatial planning policies and legislation, as well as documentation of the resource prior to loss.

In short, MENA’s MCH is challenging to monitor, relatively marginalised in cultural heritage policies and possesses a peripheral role in existing archaeological databases. Even though maritime archaeological sites, particularly coastal ones, are incorporated in databases mentioned earlier and others such as Pleiades (pleiades.stoa.org), Fragile Crescent (Lawrence et al. 2012), Leicester Trans-Sahara Project (Mattingly et al. 2017), those databases rarely represent the full array of material found in nearshore and subtidal contexts. Moreover, databases dedicated specifically to maritime archaeology in the MENA are limited. Available examples include chronologically-focused shipwreck databases and maps (The Oxford Roman Economy Project), databases by non-specialists emphasising on ports and harbours (www.ancientportsantiques.com), and a few national maritime archaeology databases managed by local institutions. Among the aforementioned databases, those that are open access, tend to emphasise specific local or regional scales of archaeological investigation, making more challenging our understanding of human experience beyond those contexts.

The marginalisation of fully subtidal MCH in existing big datasets, likely reflects the lack of accessible high-resolution seabed mapping data (Wölfl et al. 2019), particularly compared to (often) open access global coverage satellite imagery and elevation models at multiple spatial and temporal resolutions. It also reflects the higher expenses involved in maritime archaeology (Samuels 2009), which can be challenging in contexts with limited infrastructure and funding (Long 2000). This disparity also reflects the lack of specialists, due to the more complex technical requirements and training required for underwater work. More broadly, the collection of data by non-local archaeologists with the resources and technical expertise to direct most maritime projects in the MENA region, often focusing on the materiality of colonialism (e.g., shipwrecks, European settlements) perpetuates the impression that maritime archaeology is a Westernist practice (Blue and Breen 2019, 326).

It would appear therefore that important aspects are lacking from existing MCH management infrastructure that would enhance existing capacity building in the MENA region. One of them is baseline data including (1) the location of ancient maritime activities/landscapes; (2)
accessible high resolution seabed mapping data; (3) knowledge on associated threats beyond the site level that would allow further research on landscape change. Given the challenges discussed in the previous paragraphs, a large, synthetic database with consistent terminology is an obvious advantage in the evaluation of the maritime archaeology of the MENA region. In this paper, we argue that the collaborative production of an aggregated, synthetic and terminologically consistent database of known information on the MCH of the MENA region, would facilitate the identification of patterns affecting archaeology beyond the micro-regional scale of observation, and mobilise action (site monitoring, development of mitigation strategies etc.).

7. Introduction to the maritime endangered archaeology project (MarEA)

The MarEA project aims to rapidly and comprehensively document and assess maritime cultural landscapes in the MENA region. Key criteria for documentation are (1) the pre-modern chronology of features, or their association with pre-modern intangible heritage (e.g., boat building); (2) evidence for alterations by natural and/or anthropogenic factors and (3) evidence or reasonable concern that the features are likely to experience alterations (threats) in the future.

The primary methods of documentation include bibliographic research and satellite imagery analysis. The latter falls under Casana’s (2014) expert-led method and engages researchers with different regional and methodological specialisations, who examine satellite imagery, largely derived from very high-resolution imagery (0.4—1m) available via Google Earth and Bing. Quality control is achieved through the combination of multiple evidential avenues, including bibliographic and archival research, geophysical and bathymetric data (Andreou et al. 2020). Information is enhanced via collaboration with specialists and partnerships with local archaeologists with proximate knowledge to the cultural heritage of the region.

7a. The database

Information collected by MarEA is incorporated in the EAMENA database (database.eamena.org), which is an implementation of the Arches Project Cultural Heritage Inventory and Management Software (www.archesproject.org). Arches is an open-source heritage inventory app developed by the Getty Conservation Institute and the World
Monuments Fund. Arches also uses CIDOC CRM, which is the ISO standard used in research institutions and museums to describe heritage data (Zerbini 2018). The use of CIDOC CRM enables long-term sustainability of the EAMENA/MarEA data, introduces our observations within a common frame of understanding and facilitates data interoperability with other projects (Henninger 2017, 666-668).

Originally devised to offer a paradigm for national heritage data management (Zerbini 2018), the EAMENA database has come to encompass a wide array of information including at a basic level, the coordinates and extent of features or sites accompanied by a tentative interpretation, condition assessment and exposure or vulnerability to future threats. EAMENA emphasises on documenting the data collection processes, including value classifications (e.g., certainty qualifiers) on the sources and observations used to produce archaeological interpretations (site location, periodisation, disturbances, threats etc.) (Fisher et al. 2021, 6) (FIGURE 2a-c). The formulation of terminology in EAMENA took into consideration a variety of geographical zones, chronological periods and interdisciplinary approaches deriving from the project’s diverse research community (Zerbini 2018). The presence of expert-led identification combined with ground-verification in the EAMENA database, have led to its description as more “complete” and engaging (McCoy 2020, S23), but also more suitable for modelling future impacts.

MarEA adopted the EAMENA database and in the process enhanced its existing terminology and documentation forms (referred to as heritage resource models) to include maritime-specific requirements. This included extending the thesauri to descriptions relevant to maritime environments and sites and incorporating new data documentation fields that contextualise maritime heritage and coastal change. One such development is the Geoarchaeological Resource Model (GRM), which enables the documentation of geological and geomorphological data (FIGURE 3). This is intended to provide information on environmental and landscape change which can aid in archaeological prospection, interpretation and heritage management. This is a necessity for maritime archaeology given changes in coastal configurations driven by Quaternary relative sea-level fluctuations, tectonics and localised sedimentary processes, all of which operate to some extent across the MENA shoreline (Anzeidi et al. 2011; Lambeck et al. 2011; Benjamin et al. 2017; Vacchi et al. 2018). The implications for the location and preservation of heritage sites are considerable. For instance, former terrestrial or coastal sites can be found submerged on the continental shelf, whilst former marine or coastal sites can be found buried under coastal
16

sediments or uplifted above present sea-level (Flemming 1965; Besancon et al. 2004; Marriner and Morhange 2008; Galili et al. 2020). Finding and interpreting these sites within their appropriate palaeo-landscape context requires geological and geomorphological evidence of past environmental and landscape change (Flemming et al. 2017). Within MENA, this evidence is presently spread across the academic literature, and with some elements contained within specialised databases (e.g., indicators of sea-level change: Vacchi et al. 2018; Pedoja et al. 2014), and not necessarily visualised geospatially without additional processing. The GRM provides easy access to the relevant data, as well as enabling a rapid overview of the geographic distribution and nature of the evidence, and its relevance to archaeological research and management. This mirrors attempts elsewhere to integrate and record geological/geological evidence for archaeological purposes (Flemming et al. 2017; EMODnet submerged landscapes database).

7b. Data access

The EAMENA database is accessible through a registration process. Access permissions are provided at various levels from public access to researchers/users and data producers/editors. These levels of access guarantee the protection of sensitive data, if required, accounting for the requirements of national heritage partners and agencies, yet still encompassing the benefits of open data (transparency, accountability and diversification of viewpoints) (Fisher et al. 2021, 6). MarEA regularly shares observations recorded in the database as blog posts on our official website (marea.soton.ac.uk) and in social media, making them available for feedback from specialists and interested parties.

The classification of data in the Arches database reflects the aim of the EAMENA and MarEA projects to readily enable retrieval of trans-regional and diachronic combinations of information. For instance, if a local stakeholder or a researcher is interested in the impact of coastal erosion on cultural heritage, they can search for “Erosion/Retreat” and “Inundation”. The search can be narrowed geographically or even expanded thematically, for instance to explore the correlation of erosion with other impacts such as “Building/Development” or “Seismic Activity”. Resources are often linked to bibliographic references, as well as values that classify the degree of certainty of threat (“Definite”, “High”, “Medium”, “Low”, “Negligible”). This offers the user a prompt and informed selection of data that would be the most suitable for a particular question.
To date, the EAMENA database contains 331894 individual records, 8398 of which have been documented by MarEA. This includes 167335 Heritage Place records of which 8329 are produced by MarEA (FIGURE 4). Heritage place records include sites, features and archaeological landscapes, with associated information on nearby building and road infrastructure development, port development, offshore development, flooding and erosion.

8. Local collaborations

Engagement and close collaboration with local partners and stakeholders reinforces heritage networks with the MENA region, creates opportunities for capacity building and enables quality control on remotely sensed data. MarEA aims to showcase the importance of remote sensing in heritage management and collect data via mutually beneficial approaches to MCH. Local collaborators are archaeologists identified via existing networks with governmental heritage agencies, universities and via their research output. As such, they represent a wide array of interests and capacity.

8a. Training workshops

MarEA was included in EAMENA’s Cultural Protection Fund-supported training sessions and to date has offered MCH-specific training to archaeologists from Lebanon, Jordan and Palestine. Participants developed maritime archaeological skillsets, used the EAMENA database and engaged in discussion surrounding the ontological dimensions of the used terminology. These workshops encouraged the incorporation of specific local concerns, such as the potential of satellite-derived data to assess the maritime archaeology of Gaza. They also offered key information regarding how this data would be beneficial to non-academic local communities. One workshop included practical training in Aqaba, Jordan, where local partners ground-verified remotely-sensed information, but also tested site documentation and assessment forms that MarEA produced to collect information relevant to coastal vulnerability (FIGURE 5).

8b. Co-production of data
Local archaeologists actively contribute to MarEA with their regional, chronological, and material expertise. They also enhance the database through the co-production of information that bares local perceptions and engagements with heritage. One such example is our collaborative data amalgamation with specialists on the MCH of Morocco and local archaeologists (Trakadas and Karra 2020). Data relevant to the MCH of Morocco are inventoried using the EAMENA database’s terminology, alongside in-situ assessments of any damage and the vulnerability of those sites.

8c. Fieldwork

In Libya, maritime archaeology is a young but growing discipline. Recent training initiatives (Hobson 2019; Leone et al. 2020; Nikolaus et al. 2019) focused on increasing terrestrial survey skill as well as GIS and database knowledge. However, maritime-related recording techniques, data collection and analysis have received little attention.

Since the Arab Spring (2011) and the ongoing political unrest in Libya, unregulated agricultural development and building activities have increased dramatically. Archaeological investigations along the northern and western stretch of Cyrenaica’s coastline have been limited to date, but previous surveys detail a variety of sites, including classical-period harbours and settlements, production and industrial sites, fortified farms, military features, and burial features (Little and Jones 1971; Hesein 2014; Emrage 2015; Buzaian 2019; Tusa and Buccellato 2019). While threats and damages to sites (particularly erosion) are mentioned sporadically in previous studies (Flemming 1965; Bennett 2018; Bennett et al. 2004), a systematic and comprehensive condition assessment of MCH sites along the Cyrenaica coastline was lacking.

In 2020, MarEA established a collaborative pilot survey with the Department of Antiquities in Cyrenaica, Libya (DoA), to record threats and damages to archaeological sites along the coast of Cyrenaica between Apollonia (Susah) and Teucheira (Tocra), and to create a comprehensive condition assessment of coastal and submerged archaeological sites. The outcome of this survey will form the basis for the joint development of comprehensive protection and mitigation strategies for particularly vulnerable sites (FIGURE 6).

During this collaboration, MarEA conducted a desk-based assessment using open-access satellite imagery and published literature to identify and map potential sites in advance of
field survey. The desk-based results were shared and discussed with the Libyan partners and a joint decision was made on the final selection of target sites and suitable terminology for this survey. The ground-verification and survey of these sites (coastal and underwater) was then carried out by members of the DoA using dedicated MarEA survey forms in Arabic that were tailored for threat and damage assessments (FIGURES 7-8).

This expert-led evaluation by the local survey team has provided accurate information about the validity of sites identified from satellite imagery. Preliminary results show that most sites recorded during the desk-based assessment were indeed pre-modern. However, a small number were, instead, natural rock formations or more recent constructions. Data quality was further improved by the identification of small features that are not identifiable on satellite imagery, such as small industrial installations and underwater features. The data collection and identification of threats facilitated through this collaboration allowed for enhanced data quality control and targeted, more-contextualised management strategies. It also contributed to the research capacity and preparedness of partner organisation to implement local/regional responses to threat. Research capacity and intellectual reciprocity were accompanied by the creation of a five-part MCH training programme in Arabic that covered subjects including underwater recording techniques and survey, GIS applications, drone surveys and photogrammetry.

9. Example of applications: Rapid assessment of the impact of Cyclone Shaheen in Oman

A big database that summarises cultural heritage locales and their vulnerability, can be instrumental in the impact assessment of unpredictable or inevitable catastrophic events, such as tropical cyclones that often strike the southern Arabian Peninsula. Oman is widely identified as exceptionally vulnerable to tropical cyclones, often affecting areas with a high density of maritime archaeological sites, such as the Muscat and the Dhofar Governorates (FIGURE 9). Despite substantial research on impact assessment (Al Ruheili and Radke 2020), modelling (Hereher et al. 2020) and mitigation strategies (Mansour et al. 2021), cultural heritage is not yet incorporated in national policies for the protection of the environment and infrastructure.

To demonstrate the potential of MarEA’s dataset, we use as a case study Cyclone Shaheen - a category-1 storm which made landfall in Oman on 3rd October 2021. During the weeks following Cyclone Shaheen, we amalgamated all available archaeological information
pertaining to the coastline between Muscat and Sohar, which experienced the worst consequences of this cyclone (FIGURE 9). This included 184 sites, of which 41 with definite and 38 with high archaeological certainty. Documented features include the remains of coastal and near-coastal Islamic forts, traditional boats, historic buildings, and cities mentioned in historical archives and charts.

We subsequently assessed the broader impact of the cyclone through a rapid inspection of open-source Sentinel-2 (S2) satellite images covering the coastal strip. This helped us identify the area around the lower reaches of the Wadi al Hawasinah as suffering particularly severe impacts (FIGURE 10). This was evidenced on the shoreline by breaches in previously-closed wadi mouths and inland by newly-cut or expanded networks of channels.

We then assessed the extent of potential impact in terms of flooding and shoreline change in the vicinity of these archaeological sites through a more detailed analysis. This comprised extraction of surface water from pre- and post-cyclone S2 images (Table 1) using the Sentinel Water Index (SWI) and automated (Otsu) thresholding (Jiang et al. 2021), followed by differencing of the extracted pre- and post-event water layers (see also Tapete and Cigna 2020 for similar approaches applied to archaeological sites). These analyses indicated breach and flooding of the previously-closed wadi mouth and hinted at intermittent shoreline retreat (FIGURE 10a-c). That this is not the result of higher tide is supported by modelled tides from the FES2014 global ocean tide atlas (Lyard et al. 2021; measured values from the Muscat tide gauge were not available at time of writing), which show that both pre- and post-event images were acquired at high water. Onshore, areas of standing water can be identified, with a particularly extensive zone located 1.6-0.4km northwest of the wadi. This was also verified using similar techniques applied to higher resolution Planetscope images (Table 1), the key differences identified with the use of McFeeters (1996) Normalized Difference Water Index (NDWI) and manual thresholding. These show broadly similar patterns, albeit with slightly reduced surface water, but stronger evidence of shoreline recession and inland penetration along newly-flooded inlets (FIGURE 10d-f). The severity of this impact is further illustrated by previous studies which show that the wadi mouth has been usually blocked since construction of a dam in 1995 and the coastline to its northwest is generally regarded as stable (Al-Hatrushi 2013).

These images are of insufficient resolution to conclusively determine impacts on the specific documented sites, and the 2-day delay for the post-cyclone images also means that the
immediate impact of flash flooding cannot be detected. Nevertheless, the results highlight that this area is vulnerable to cyclone-driven flash flooding and the overall extent of remaining standing water suggests that flood waters could have reached the sites but had drained away by the time of image acquisition. Shoreline retreat does not seem to have reached the sites in this instance, but rapid localised retreat of up to several metres is suggested by the analysis. This too highlights the continued vulnerability of the area in light of the rising 21st Century sea-level, intensified cyclone patterns and possible accelerated coastal retreat (Hereher et al. 2020; Vousdoukas et al. 2018; 2020). As such, this case study demonstrates how the MarEA database in conjunction with other ‘Big Data’ (in this case the ever-expanding archive of open-source satellite imagery) enables rapid assessment and can be used to identify sites which require more frequent monitoring. In this case, site inspections are suggested to assess flash flood impacts coupled with baseline survey against which future coastal changes can be monitored.

10. Conclusions

MCH constitutes some of the most significant datasets incorporated in narratives of human mobility and interregional interaction, but also contains the most vulnerable assemblages in terms of preservation. The MCH of the MENA region, which is relatively marginalised in cultural heritage policies, is challenging to monitor at large-scales and generally possesses a peripheral role in existing conversations on endangered archaeology. There is a critical need for rapid and cost-effective documentation of sites and associated threats in vulnerable, dynamic and rapidly changing maritime landscapes. Although small-scale research was able to highlight the detrimental impact of natural and anthropogenic processes on specific maritime archaeological sites, the broader geographical impact of those processes remains unclear.

Extensive historical and archaeological research on the MCH of the MENA region has produced large quantities of information on past human activities, diachronic land use and landscape alterations, which are disseminated in different formats, some more accessible than others. Moreover, the associated data have not been integrated. The aggregation of such vast datasets requires a systematic classification system that will reflect the historical and geographical particularities of the archaeologies of the MENA region via extensive consultation with a diverse array of experts. Though the primary purpose of the MarEA
dataset is to facilitate the identification of vulnerable sites and mobilise monitoring and the development of mitigation strategies, the dataset also offers a unique opportunity to scholars to develop new research syntheses and directions.

Viewing the lack of detailed and consistent documentation of MCH in the MENA region, a region central in global historical discourses, in this paper we aimed to demonstrate the importance of big data in the context of endangered MCH. Rather than viewing a big dataset as a means of producing answers to longstanding questions about pre-modern human societies, we instead engage with it as a tool for enabling new avenues for investigation. While emphasising the benefits of big data approaches to document endangered MCH, we are also keenly aware of the associated epistemological, ethical, and analytical challenges. Similarly, we share concerns about discourses focusing on the speed of data acquisition and processing (Caraher 2016, 2019), and invest in quality control through interdisciplinary research, in-situ visits when possible, and local partnerships.

With a critical gaze on data collection, management and dissemination processes, and close engagement with quality control and local communities, big data approaches on MCH offers significant new data to archaeological research. Such approaches can open regional interpretations that are more empirical, relying on a wider variety of archaeological features, and offering numerous archaeological assemblages through the combination of diverse datasets. At a local level, a large dataset would highlight broader patterns affecting MCH and mobilise action often impeded by the lack of accessible baseline data. With expert-led, locally informed, and regularly ground-verified data, we consider that MCH, people in charge of its stewardship, and local communities that have developed affective ties with it, would benefit immensely from big data approaches.

Bibliography


Emrage, A. S. 2015. “Roman Fortified Farms (Qsur) and Military Sites in the Region of the Wadi Al-Kuf, Cyrenaica (Eastern Libya)”. Ph.D. diss., University of Leicester, Leicester.


Hamilakis, Y. 2012. “...Not Being at Home is One’s Home”: Ontology, Temporality, Critique.” *Forum Kritische Archäologie* 1: 12-23.


of “At Risk” Coastal Archaeology: The Example of Ancient Siraf (Iran, Persian Gulf).” *Applied Geography* 101: 45-55.


Table 1. Summary metadata for satellite imagery used in the Cyclone Shaheen Assessment

Figure Captions

Figure 1: Map of the MENA region (produced on ArcGIS Pro using bathymetric data from the General Bathymetric Chart of the Oceans, www.gebco.net).

Figure 2: Example of data entry in the EAMENA database (Screenshot from database.eamena.org)

Figure 3: Ontological map of the Geoarchaeological Resource Model (produced on Microsoft Visio).

Figure 4: Map of MarEA site density in the EAMENA database (produced on ArcGIS Pro).

Figure 5: Coastal vulnerability assessment in Aqaba, Jordan (photo by William Deadman, November 2019).

Figure 6: Map showing sites documented during the first two phases of the Cyrenaica Coastal Survey (produced on ArcGIS).

Figure 7: Rock-cut tombs at Phycus, many of which are used as storage facilities and domestic animal shelters (photo by the CCS team, March 2021).

Figure 8: Remains of structures on the coast of Phycus (photo by the CCS team, March 2021).

Figure 9: Map of maritime archaeological sites in Oman documented in the EAMENA database in conjunction with major cyclonic tracks over the past 130 years (National Oceanic and Atmospheric Administration).

Figure 10: Satellite image showing the impact of Cyclone Shaheen at Wadi al Hawasinah. Documented forts are shown as yellow dots. A) Sentinel-2 pre-cyclone, true colour image (dashed boxes indicates location of detailed view using Planetscope imagery). B) Sentinel-2
<table>
<thead>
<tr>
<th>Satellite</th>
<th>Bands</th>
<th>Acquisition date; time (UTC)</th>
<th>Spatial resolution</th>
<th>Availability (Source)</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentinel-2 (Level 2A)</td>
<td>13 bands (Vis, NIR, SWIR)</td>
<td>27/09/2021; 06:31</td>
<td>10-60m (band dependent)</td>
<td>Public (Google Earth Engine)</td>
<td>Pre-cyclone Shaheen</td>
</tr>
<tr>
<td>Sentinel-2 (Level 2A)</td>
<td>13 bands (Vis, NIR, SWIR)</td>
<td>05/10/2021; 06:47</td>
<td>10-60m (band dependent)</td>
<td>Public (Google Earth Engine)</td>
<td>Post-cyclone Shaheen</td>
</tr>
<tr>
<td>Planetscope (Level 3A)</td>
<td>4 bands (Vis, NIR)</td>
<td>27/09/2021; 06:25</td>
<td>3.7m</td>
<td>Free academic license (Planet)</td>
<td>Pre-cyclone Shaheen</td>
</tr>
<tr>
<td>Planetscope (Level 3A)</td>
<td>4 bands (Vis, NIR)</td>
<td>05/10/2021; 05:59</td>
<td>3.7m</td>
<td>Free academic license (Planet)</td>
<td>Post-cyclone Shaheen</td>
</tr>
</tbody>
</table>

Table 1. Summary metadata for satellite imagery used in the Cyclone Shaheen Assessment
Figure 2

Figure 2a-c.tif
Figure 3: EAMENA-MarEA database
Geoarchaeology/Palaeolandsapes
Resource Model

- **Geoarchaeology (GA)**
  - **Assessment Summary**
    - Investigator Name
    - Assessment Activity (e.g., Diver Survey, Marine Geophysical Survey)
    - Assessment Activity Date
  - **Resource Summary**
    - Resource Name
    - Description
    - Process Indicator (e.g., Sea-level Rise, Tectonic Uplift, Channel Migration)
  - **Geography/Geometry**
    - Location Certainty
    - Country
    - Maritime Region (e.g., Territorial Waters)
    - Location Type (e.g., Data Extent, Interpolated Boundary)
  - **Environment Assessment**
    - Bedrock Geology
    - Surficial Geology
    - Depositional Process
    - Wave Climate
    - Tidal Energy
    - Depth/Elevation
    - Topography
    - Land Cover
  - **GA Assessment**
    - GA Certainty
    - Source of Evidence (e.g., Core, Borehole, Seafloor Bathymetry)
    - General Date (e.g., MIS, Quaternary Division)
    - GA Feature Sediment (e.g., Sand, Coarse Sediment)
    - GA Feature Landform (e.g., Terrace, Channel)
    - GA Interpretation (e.g., Marine Terrace, Glacial)
  - **Condition Assessment**
    - Overall Site Condition
    - Damage Extent
    - Disturbances (e.g., Natural, Tourism)
    - Disturbance Causes (e.g., Coastal Erosion, Demolition, Construction)
    - Effects (e.g., Alteration of Terrain, Water Damage)
    - Threats (e.g., Trawling, Dredging, Wind Action)
Georgia Andreou (PhD 2015, University of Edinburgh) is a researcher at the University of Southampton and an associate lecturer at the Institute of Archaeology at UCL. Her interests include the prehistory of the Mediterranean and the Near East.

Julia Nikolaus (PhD 2017, University of Leicester) is a researcher at Ulster University with a focus on North African archaeology and heritage protection.

Kieran Westley (PhD 2007, University of Southampton) is a researcher at Ulster University. His interests include climate change, remote sensing and submerged prehistory.

Crystal el Safadi (PhD 2018, University of Southampton) is a researcher at the University of Southampton with expertise in maritime archaeology and archaeological computing. Her main interests are in digital approaches to the human past, prehistoric maritime mobility, and mapping technologies.

Ash Smith (PhD 2008, University of Southampton) is a database engineer working at the University of Southampton.

Lucy Blue (DPhil 1996, University of Oxford) is a lecturer at the University of Southampton and co-director of the MarEA project. She is also the part-time maritime archaeological director of the Honor Frost Foundation.

Colin Breen (PhD, 2003, National University of Ireland, Galway) is a Reader in the School of Geography and Environmental Sciences, Ulster University, and co-director of the MarEA project. He has research interests in maritime societies, environment change and conflict.