

# The sale of heritage on eBay: Market trends and cultural value

Big Data & Society  
July–December: 1–17  
© The Author(s) 2020  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2053951720968865  
journals.sagepub.com/home/bds



Mark Altaweel<sup>1</sup>  and Tasoula Georgiou Hadjitofi<sup>2</sup>

## Abstract

The marketisation of heritage has been a major topic of interest among heritage specialists studying how the online marketplace shapes sales. Missing from that debate is a large-scale analysis seeking to understand market trends on popular selling platforms such as eBay. Sites such as eBay can inform what heritage items are of interest to the wider public, and thus what is potentially of greater cultural value, while also demonstrating monetary value trends. To better understand the sale of heritage on eBay's international site, this work applies named entity recognition using conditional random fields, a method within natural language processing, and word dictionaries that inform on market trends. The methods demonstrate how Western markets, particularly the US and UK, have dominated sales for different cultures. Roman, Egyptian, Viking (Norse/Dane) and Near East objects are sold the most. Surprisingly, Cyprus and Egypt, two countries with relatively strict prohibition against the sale of heritage items, make the top 10 selling countries on eBay. Objects such as jewellery, statues and figurines, and religious items sell in relatively greater numbers, while masks and vessels (e.g. vases) sell at generally higher prices. Metal, stone and terracotta are commonly sold materials. More rare materials, such as those made of ivory, papyrus or wood, have relatively higher prices. Few sellers dominate the market, where in some months 40% of sales are controlled by the top 10 sellers. The tool used for the study is freely provided, demonstrating benefits in an automated approach to understanding sale trends.

## Keywords

eBay, heritage, antiquities, natural language processing, name entity recognition, conditional random field, machine learning

This article is a part of special theme on Heritage in a World of Big Data. To see a full list of all articles in this special theme, please click here: <https://journals.sagepub.com/page/bds/collections/heritageinworldbigdata>

## Introduction

The online market for heritage, represented by antiquities or portable cultural objects such as jewellery or clothing items, herewith simply antiquities, is highly dynamic and one that can respond to increased consumer demands and as cultural material becomes available. Increasingly, online platforms, including social media sites, are used to sell antiquities (Brodie, 2015; Huffer and Graham, 2017). There are effectively two antiquities markets: one illegal and the other legal, although even legal markets can be considered grey markets due to the fact that stolen or illegally removed items are sometimes sold in legal markets. The illegal

market is complex and generally obscure, with some recent work showing it is possible to study this market using social media sites such as Facebook (ATHAR, 2019). Generally, data about this market is difficult to obtain, including understanding its shifting

<sup>1</sup>Institute of Archaeology, University College London, London, UK

<sup>2</sup>Walk of Truth, The Hague, Netherlands

### Corresponding author:

Mark Altaweel, University College London, 31–34 Gordon Square, London WC1E 6BT, United Kingdom.

Email: [m.altaweel@ucl.ac.uk](mailto:m.altaweel@ucl.ac.uk)

market dynamics and opportunities (Brodie, 2012). The legal (or even grey) antiquities market provides easier data to obtain that demonstrates how wider consumer interests, including cultural value, intersect with supply that drive purchasing behaviour. Cultural value is defined here as ideals, interests and beliefs that can shape or affect social norms and behaviour including purchase behaviour (Frese, 2015). In the antiquities market, cultural value can affect what types of objects are of interest and purchased. For instance, education and media could shape interests for given cultures by purchasers, such as Roman, Egyptian and Near East cultures often taught as part of ancient history topics in Western schools (Cooper, 2017; Roupp, 2010). Additionally, researchers have argued that the legal market in antiquities may also drive interests in the illegal market, as consumers turn to both to obtain objects of interest (Brodie et al., 2006). Overall, it might be possible to use the legal antiquities market to obtain insight into consumer interests and market opportunities that provide insight into social factors affecting the illegal market.

Methods to understand the antiquities market have used both online and offline data (Brodie, 2012; Brodie and Renfrew, 2005; Fay, 2013; Mackenzie and Yates, 2017); where online data has been collected, efforts have rarely utilised relatively larger datasets over a sustained period of measurement. The online antiquities market, at least sites that publicly advertise sales, makes it possible to measure what type of objects is sold and where. Object types may include information on cultural affiliation (e.g. Roman, Scythian), general use types (e.g. tools, weapons, statues) and material composition (e.g. wood, stone). This information enables a clearer idea as to what cultures are in demand as well as knowledge on object characteristics that help to inform on sale trends. Collecting sales data from given locations also enables a method to measure where the market is most active and demonstrates cultural and economic value objects have for consumers. One potential is to use natural language processing (NLP) so that unstructured and structured text data that provides such information on objects can be better harvested and understood. Named entity recognition (NER), using a conditional random field (CRF) approach, is one method within NLP that enables labelling of unstructured texts where they can then be quantified and term patterns could be understood in order to provide broader insight for sales and descriptive data (Lafferty et al., 2001; Lee et al., 2006). Such approaches help classify text into either predefined or learned categorical names that then facilitates larger or more generalised text patterns to emerge.

This work applies an NLP approach using NER and term searches on data collected from eBay's US site.

The intent of this work is to understand sale patterns on eBay, a very common site used for the legal antiquities sales, so that the extent to which given cultures, object types and materials sold are better understood using more systematic methodology. Additionally, the results are framed and understood within a cultural value perspective to attempt to explain sale patterns. The work attempts to assess how cultural value drives sales. The sales data analysed is provided as part of this work and is collected over the period covering 21 October 2018 to 3 January 2020. Before presenting this data, the ethics statement and disclosure of this work are made. The data was collected directly from eBay using a Python developed scraping tool created by the author, as the API within eBay proved to be insufficient for large-scale data capture. The use of such data from eBay is legal and permissible so long as it is not used for commercial purposes; this work also anonymises sellers' names as numerical sequences so that sales data cannot be linked to any individual.

To begin the presentation, background works and summary on the antiquities market and factors that could shape it are presented. Part of this presentation includes discussion on the use of computational and NLP methods in understanding heritage along with a general background on NER. The method applied in this work is then introduced, building on earlier work from Altaweel (2019). The results are then presented with the subsequent discussion and conclusion focused on insights obtained on the antiquities market and how that may reflect wider cultural interests and values that shape the heritage market. Limitations and future work are explored, including the relevance of such work to Big Data analysis, as part of the discussion.

## Background

### *Antiquities, heritage and online sources*

The study of heritage has largely avoided automated data gathering from Internet sources to assess the antiquities market. However, some notable studies have been conducted. Recently, Huffer and Graham (2017) studied social media using hashtag searches in relation to the sale of human remains. The results showed a relatively large portion of human remains are sold by a relatively small circle of sellers. Machine learning methods have also been applied, in particular recently by Greenland et al. (2019), who used looted areas to estimate the value of antiquities sold on the illegal heritage market. In fact, this might be the only study to use machine learning-based approaches to study the illegal antiquities market from source sites where objects are looted. The ATHAR (2019) project recently released a report detailing how Facebook has

been extensively used by social networks to sell Syrian antiquities. Altaweel (2019) is the most similar to this work, and the methods applied in that work are applied here, although there are updates to those methods. This work also expands on this earlier effort by investigating over a longer period of time and assessing sellers in different countries, rather than just sales, for part of the investigated period. Unlike the earlier work, the results here are linked to cultural value, and an attempt is made to provide underlying cultural reasons that may explain market trends observed. Other heritage-related work using NLP and machine learning includes Bonacchi et al. (2018), although the focus is not related to the sale of antiquities, but, rather, investigates how identity politics and ideas about the Iron Age, Roman and post-Roman past have shaped the Brexit debate on Facebook.

Although there have been few systematic studies looking at the antiquities market applying NLP or machine learning, other studies do exist that demonstrate the relevance of the online antiquities market. Researchers have acknowledged the increasing role and importance of the Internet in selling antiquities, including the illegal and legal market (Fay, 2011, 2013; Mackenzie and Yates, 2017). The online market is described as a potential grey area of antiquities sales, where objects are often of dubious provenance and are difficult to prove that they derived from legal means. The market of antiquities has been demonstrated to be international, with Western states often dominating sales (Barker, 2018; Bowman, 2008). eBay has emerged as a leading site in the sale of antiquities because it offers an international reach and it makes it easy to sell objects anywhere (Barker, 2018; Fay, 2011). What eBay provides is an auction platform allowing users to bid on objects placed by individual sellers; eBay also operates sites in different countries as well as its US flagship site. For antiquities, eBay mostly offers lower-end antiquities, although these are also the most common form of antiquities, and such a site could potentially serve as a measure of broader interests and sales in antiquities because it provides a large number of sales, demonstrating also wider participation by independent individuals and some auction sellers (Kotha and Basu, 2011). eBay has been shown to be a place often selling fakes or copies of objects (Stanish, 2009). Even if a large number of objects are fakes, eBay, given its scale, still serves as a useful proxy that gauges interests in the antiquities market in that it shows where many sales are occurring. We do have to be wary in assuming all objects sold on eBay are truly authentic, but online sites such as eBay demonstrate what interests, including cultures, types of objects and materials composition of objects sold, are evident in the antiquities market based on legal sale trends.

The goal of this work is to identify market trends and insights from the antiquities market using a legal site. Although it is hard to measure the outputs of this work relative to how the wider market has been, most works would agree that the Western market, particularly Europe and North America, and East Asian countries are often seen as among the most active markets for sales (Brodie, 2006). This is in large part driven by wealth in these societies that foster capabilities and interests in obtaining antiquities. Cultures affiliated with Western society are often of greater interests in many regions where antiquities are sold (Kersel, 2006). Examples include ancient Rome and the ancient Near East, with Rome seen as having established roots of modern Western societies and the ancient Near East being the home for major Western religions (Walton, 2006; Weisband and Thomas, 2015). Long-held fascination with some cultures, in particular ancient Egypt, has also driven demand (Fritze, 2016). However, today, objects can derive from many countries of origin given how easy online sites make sales. Many items often sold are portable, with objects sold generally being jewellery, coins, statues and other relatively small items, although larger, less portable items are sometimes looted or sold legally (Atwood, 2004). In effect, cultural values and how they are historically shaped have helped to drive the antiquities market, while general wealth in societies is also another key factor in shaping sales. These observations are general but they can allow this work to determine to what extent cultural values have shaped the antiquities market based on eBay's sales and what has been observed elsewhere, while also producing a more systematic assessment of a key online market.

### *Name entity recognition*

NLP techniques have long been used to help classify text, including categorising information in order to determine larger text patterns in topics or identifying common themes in text. This includes using machine learning and statistical methods, with deep learning techniques also becoming more common (Kamath et al., 2019; Powers and Turk, 1989). Within NLP approaches, NER is commonly applied as a means to identify and categorise terms from specific to usually more general terms that have informative value in given domains (Konkol et al., 2015; Nadeau and Sekine, 2007). NER application has included a variety of fields and domains that range from medicine to the social sciences. Approaches include supervised, unsupervised and semi-supervised methods (Nadeau et al., 2006; Ritter et al., 2011). Among common approaches in NER is CRF, which is a statistical-based approach that takes context of terms and their relationships, that is how surrounding terms for the term of interest relate,

and treats them as an undirected graph that enables term categorisation based on term occurrence probabilities (Lafferty et al., 2001). The conditional distribution of term relationships becomes the model that determines the likelihood that a given term falls under a given categorisation. The method is generally applied as a form of supervised learning that requires terms for training and those terms should be categorised based on the desired designations. In a CRF, word order and commonality of co-occurrences condition probabilities in the undirected term graphs using a Markov approach for probabilities. This has meant that CRF requires a large number of texts to make them acceptably accurate (Bundschuh et al., 2008; Sutton, 2012). To address this, some researchers have applied semi-supervised techniques as well as weighing techniques that can balance category selection for terms (Zafarian et al., 2015). Other methods also include term dictionary searches that match and designate given terms to a category using a domain-specific search (Altaweel, 2019; McCallum and Li, 2003; Wang et al., 2017; e.g., *axe* designated as a ‘weapon’, which is a more general categorisation). Overall, NER and CRF have not been extensively applied in studying the antiquities market.

## Method

This work applies the eBayScrapper tool found in GitHub (2020); the version applied for this work is found in an online repository for download (see Data Accessibility, supplementary material). This repository includes relevant data discussed in this article, including the sales data used for analysis and code applied. The methods deployed generally follow Altaweel (2019); however, there are some important modifications to this original work that are indicated and presented below.

The tool, as applied here, searches the US eBay site ([https://www.ebay.com/sch/37903/i.html?\\_sop=13&\\_sadis=15&LH\\_Complete=1&LH\\_Sold=1&\\_stpos=90278-4805&\\_from=R40&\\_nkw=%27+Antiquities+%27&LH\\_All=1](https://www.ebay.com/sch/37903/i.html?_sop=13&_sadis=15&LH_Complete=1&LH_Sold=1&_stpos=90278-4805&_from=R40&_nkw=%27+Antiquities+%27&LH_All=1)) dedicated to antiquities; this site is often used for international sales. The sales data could also be searched using eBay’s API; however, there is a limitation in the number of searches one can carry out on a daily basis. Therefore, collecting data was found to be easier by simply scraping data from the public sales data site using a scraping module launched every month. Figure 1 provides the flowchart describing the workflow applied; the description below discusses the method. Overall, eBayScapper is composed of two distinct tools. The eBayScrapper tool applies Python text scraping, NER analysis, dictionary searches and some analytical capabilities, including

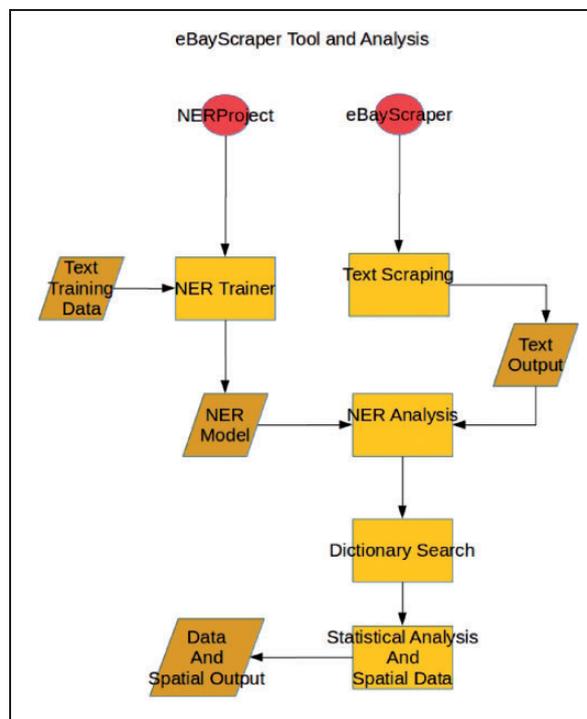


Figure 1. Workflow of the applied methodology.

outputting the results as CSV and shapefiles for statistical and spatial analysis. However, most of the statistical analysis presented is conducted in R (3.6) in this work. The other tool (NERProject) contains Stanford NLP libraries and uses training data to create the NER model applying CRF (described below). The eBayScrapper tool was written in Python 2.7+, but it has now been converted to work in Python 3 (3.8). The NERProject uses Java 8. Below, a summary of the applied steps in Figure 1 are described. The repository link (supplementary material) provided with this work gives more detailed information and description of key modules and methods in the code, with code level documentation also given. This can also be used to guide interested users of this tool.

### Text scraping

Using the eBay URL link provided, sub-sites, that is sites within the main antiquities sales site, are classified to include objects from various cultures (e.g. Celtic, Roman, Near Easter, etc.). Although objects are generally categorised based on their associated cultures, the cultures assigned to objects are often mislabelled or labelled in multiple categories. Duplicate scraped data is removed at the end of the text scraping process. Overall, because information is often mischaracterised, it is necessary to classify objects using the descriptive information provided rather than where objects are

classified in the sales data given by eBay. Scraped data recovered from eBay sales information includes: the date when an object is sold, the US dollar value the item was sold for, description of the object, location of the seller and seller username (collected from 8 July 2019). The collection of seller-related data is new in this work relative to the earlier work; the data given here is anonymised (Altaweel, 2019). All of the data, with the exception of the object descriptions, is structured. Each sub-site’s structured and unstructured scraped text is then outputted into a CSV file that can then be analysed in the next steps. The data scraped, including anonymised sellers, and analysed is provided as part of the repository link (supplementary material) provided. Data from all the sub-sites is aggregated into one file for subsequent analysis. The object descriptions are used to classify text based on culture, object type and material composition in the NER and dictionary search methods.

### Name entity recognition analysis

The NER method deployed utilises CRF using its deployment in the Stanford NLP toolkit (Finkel et al., 2005; Manning et al., 2014; Stanford NLP, 2020). The tool is wrapped within the NERProject tool within eBayScraper, where the NERProject uses training data to create an NER model that can then be deployed for the NER analysis. The training data provided indicates terms that are categorised using

designations created by this effort (Table 1). Application of CRFs utilises sentence structure, including word order, to inform on the probability that given terms could be classified using the prescribed categorical terms (Lafferty et al., 2001). The CRF method applied using Stanford’s NLP tool can be given in this summarised form:

$$p(y \vee x, \lambda) = \frac{1}{Z}(x) \exp\left(\sum_j \lambda_j F_j(y, x)\right) \quad (1)$$

with  $y$  referencing a label sequence, that is identified terms or entities, with a sequence used for the observations defined as  $x$ ;  $F$  represents a feature function that derives from an empirical distribution based on training, with a normalised factor  $Z$ , and  $\lambda$  representing the estimated parameter using the trained data (Finkel et al., 2005). Frequently appearing terms in association with a categorical term help define outputs. For instance, the term *iron* could fall under the category “*metal*”, such as in *iron axe* indicating a type of ‘metal’ object, while the term *Iron Age* would not be in ‘metal’ due to that term association. Training data determines if a term is likely to fall in one categorisation or another, while also differentiating potentially ambiguous contexts or usage of terms. Table 1 indicates the categories used for training and in the model; it also shows example terms in relevant categories. These categories are determined through

**Table 1.** Categories and example terms used in the NER CRF approach.

Cultures	Example term	Object types	Example term	Materials	Example term
Africa	Moroccan	Clothing	Button	Bone	Femur
America	Columbian	Coin	Drachma	Glass	Glass
Bronze age	Bronze Age	Decoration	Plaque	Ivory	Ivory
Buddhist	Buddhist	Household	Purse	Leather	Leather
Cambodian	Dong Son	Jewellery	Necklace	Metal	Iron
Celtic	Celt	Mask	Face	Papyrus	Papyrus
Central Asia	Indo-Scythian	Religious	Ritual	Terracotta	Ceramic
China	Chinese	Statue	Statute	Stone	Flint
Egyptian	Pharaonic	Text	Document	Wood	Wooden
European	Anglo-Saxon	Tool	Stylus		
Greek	Hellenic	Painting	Painted		
India	Mughal	Portrait	Portrait		
Iron age	Geometric Period	Vessel	Olpe		
Islamic	Al Andalus	Weapon	Sword		
Japan	Japan				
Medieval	9th century				
Near East	Mesopotamia				
Pre-Historic	Neolithic				
Roman	Rome				
Russian	Rus				
Viking	Nordic				

The three information types categorised are cultures, material types and material composition for objects.

observations of cultures, object types and materials that appeared in descriptive data as well as creating general terms that describe antiquity objects. The effort balances general descriptive terms with more specific and distinct terms. The 44 categories used are reviewed by the author over several iterations using domain experience and observation from samples in the empirical data. If descriptions do not allow for categorisation then the result is ‘unknown’, that is returning as undetermined. The spellings used in the category designations given in the results are UK spellings, although both US and UK spelling standards are utilised in training and categorising terms. In a deviation from Altaweel (2019), this work assesses deliberately misspelled words and corrects misspellings given by the seller in the NER, considering the outputs of both as possible for categorisation. In other words, misspellings are incorporated as part of the analysis, while words that can be corrected to proper spelling are also corrected before they are then analysed in NER. Over 5000 lines of training texts are used with this data deriving from sale samples randomly selected from the overall sale corpus, with the CRF model created over 99 iterations. For the analysed text, multiple categorisations are possible, as terms can either fall in one or multiple categories, while lines of text analysed could also describe multiple categories and could be relevant for different cultures, object types and materials.

### *Term dictionaries*

One deficiency of the CRF method is that it requires a large number of texts to reach a very high degree of accuracy in any model and subsequent analysis. To address this, the approach taken here is to use word dictionaries in regular expression searches. Searches are not case sensitive but they match comparable terms searched. The search applies spell checking on descriptions, while alternative spelling is also searched in analysed text, applying what is discussed above. Changes from Altaweel (2019) include new terms added to search categories in Table 1. Lemmatised terms, that is term inflections, are additionally searched. Generally, terms that are relatively less ambiguous are used in term searches, where this is based on domain knowledge. Similar to the NER method, terms are designated for given categories specified (see Table 1), although in this case terms are not trained to create the NER model. The exact terms used in dictionary searches are presented in the repository link (supplementary material) provided. They are listed under the culture, object and material categories as defined in the NER approach. The dictionary approach is used after the NER approach is used, effectively giving it a second

layer of analysis that can categorise terms if categorisation is missed in the NER. Python’s Natural Language Toolkit is utilised within eBayScraper for processing text, including tokenisation and lemmantisation.

### *Statistical analysis and spatial data*

Data is aggregated and organised by sale date, cultures, type and material composition for objects, along with where objects are sold and seller data. Data categorised based on the analysis is outputted in a .csv and is also outputted to a shapefile for spatial analysis. The seller’s username is given only for dates after 8 July 2019, but anonymised using random numbers assigned to usernames. The aggregated results also return top monetary sellers for countries in the shapefile output. The incorporation of seller information and results is an added feature to the previous eBayScraper version (Altaweel, 2019). Additional statistical analysis is also done in the R statistical language and is provided as part of the results and discussion below.

### *Precision and recall*

To measure the effectiveness of CRF in NER and term dictionaries applied here, a precision and recall test is applied. This simply uses tests as described elsewhere (Goutte and Gaussier, 2005; Powers, 2011). This results in an F-score, or  $F_1$  score, that is a harmonic mean value measuring both precision and recall. Precision evaluates how accurate the categorisation given for the text is; recall is how well relevant documents are retrieved successfully. Precision measures accuracy of the approach; recall is the sensitivity of the approach in obtaining desired information. Overall, the closer the  $F_1$  score is to 1 the better the approach is in being precise and sensitive to relevant data based on the applied design and categorisation created here; values over 0.9 are considered sufficient in demonstrating the methods given here accomplished accurate categorisation and sensitivity in capturing relevant text. If both precision and recall score over 0.9, then this result is considered desirable. In testing precision and recall, this work randomly selects 500 documents and conducts the  $F_1$  score for the cultures, types and materials for analysed objects. This gives a total of 1500 tests conducted that create the  $F_1$  score. The random samples selection for precision and recall testing was conducted in the randomSelector.py module in the test folder in eBayScraper.

**Table 2.** The precision (P) and recall (R) results that include the  $F_1$  score.

P	R	$F_1$
0.97	0.92	0.94

## Results

### Summary results

Overall, there are 108,559 individual sold items analysed, with a total value of \$5,088,174 and over 1.5 million descriptive terms assessed. The most expensive item sold is for \$15,200, while over 91% of items sold are under \$100. Table 2 provides the precision and recall test results, with an  $F_1$  score showing that the approach generally works adequately as defined by this study, with a high rate of accuracy and acceptable sensitivity in displaying relevant categorised data. Table 3 provides summary data that indicates top selling cultures, object types and material composition of objects as determined by the NER and word dictionary search method. Seller (anonymised numbers) data is also provided.

### Artefact results

Overall, Table 3 makes it evident that Roman, jewellery and metal objects were sold the most for the period assessed. Roman, Egyptian, Viking (i.e., Norse- or Dane-related) and Near East artefacts represent the most common cultures respectively, with unknown excluded in this case. Egyptian objects are slightly less than half the total Roman object sales; Viking are one-third the sales of Roman objects; Near East sales are about 30% of Roman objects. Overall, the top four cultures sold, excluding the unknown category, represent over 65% of sales. For object types, jewellery are the most common object sold while unknown objects, that is undetermined types, are the second most common object type. Statues, which include figurines, are about half the sales of jewellery and religious objects are slightly behind statues. For materials, metals are over 50% of total sales; unknown materials represent about 26% of objects sold. Stone and terracotta represent about 19% and 14% of objects sold, respectively.

Figure 2 represents the distribution of sales for different object types and materials for the top four most common cultures sold. Generally, relatively rare objects have higher mean and median sale prices. For Roman objects, texts have the highest mean (\$112.88) and median (\$50) sale prices, with one standard deviation for these objects being \$660.56. Clothing items have the lowest mean sales (\$24.58) and weapons

have the lowest median (\$14.01) sales. Jewellery, the most common object sold, has \$42.81 and \$20 for mean and median sales, respectively, but it also has a fairly wide sales spread (\$89.55 standard deviation). Statues also had relatively high sale prices with \$82.57 and \$47.81 for mean and median sale prices. Papyrus materials command the highest mean (\$66) and median (\$80) sale price and are rare, while common metal objects are \$50.44 and \$23.26, respectively, for the same measures. Leather, on the other hand, had the lowest mean and median sale values (\$11.76 and \$9.78). For Egyptian objects, less common objects also obtained high prices. Masks have a mean sale price of \$109.33 and median price of \$53.09; clothing has a mean sale price of \$73.75 and median of \$65. The most expensive item sold for this culture is for \$12,900, which is a text object. For materials, wood (\$99.96 mean; \$69.06 median) and papyrus (\$82.78 mean; \$55.5 median) commanded higher prices. Metal Egyptian materials generally received higher prices than their Roman counterparts (\$67.78 mean; \$43.14 median), with a wide standard deviation (\$288.62). Only glass (\$19.37 median) had a median or mean value under \$20 for Egyptian object types and materials. For Viking objects, vessels had high mean (\$103.65) and median (\$78.71) values. Statues (68.72 mean; 47.23 median) and tools (\$75.87 mean; \$25.84 median) also gained relatively higher sales. Tools and weapons had wide standard deviations (\$194.48 and \$112, respectively). Discounting unknown materials, terracotta (\$68.06 mean; \$60.35 median) received the highest prices. Leather (\$25.54 mean; \$8.71 median) objects are relatively cheap, similar to other cultures. Near East objects stand out for their relatively high prices. Masks (\$404 mean; \$404.92 median), texts (\$213.69 mean; \$51.5 median) and clothing (\$136.88 mean; \$27.1 median) are some types of objects that have relatively high prices compared to other cultures. Metal (\$76.40 mean; \$22.50 median) and terracotta (\$68.18 mean; \$43.93 median) are materials that received the highest prices for the culture. One inscribed silk object (\$15,200) received the highest price in the entire study period.

### Seller results

One novel result in this work is seller data is tracked for at least part of the dataset (8 July 2019 to 3 January 2020). The top sellers, and the total sold by that seller, are listed in Table 3. Looking deeper into these statistics, it is evident that relatively few sellers in the period when data is available sold a large percentage of objects (Figure 3(a)). Three sellers alone sold over 31% of sales during July 2019 to January 2020. Looking at the top four sellers and assessing object categories, these sellers

**Table 3.** Summary eBay results showing cultures (grey highlight), object types and material composition (grey highlight) categories for total objects sold (USD) using the NER/dictionary search.

Type	Total	Mean	Median	SD	Min	Max	Seller	Total by seller
Africa	13,942.3	23.59	14.95	72.27	0.99	1,692	73,957,878	782.89
America	206,646.44	83.73	29.00	214.97	0.01	3,850	66,204,376	12,499
Bronze age	60,466.33	60.11	33.16	80.75	0.95	861	23,911,541	4,851
Buddhist	146.7	36.68	34.95	24.05	7.95	68.85	75,344,177	68.85
Cambodian	2,322.77	59.56	26.05	81.31	5.5	375	10,672,785	375
Celtic	121,732.51	39.00	20.74	74.00	0.74	1,800	97,422,287	12,021.16
Central Asia	50,939.4	72.56	17.36	248.20	0.75	2,738.88	55,608,283	2,101.07
China	30,841.86	34.77	16.40	105.60	0.99	2,611	34,234,785	2,270.28
Egyptian	748,420.31	46.33	29.14	157.89	0.01	12,900	97,422,287	43,928.71
European	68,718.51	46.46	24.99	115.25	0.07	3,300	97,422,287	21,627.96
Greek	420,108.1	50.91	24.99	169.48	0.01	12,900	34,234,785	30,500.25
India	59,212.76	50.10	39.30	56.75	0.99	1,000	26,778,633	2,898.22
Iron age	12,771.35	69.03	30.00	113.28	0.95	600	23,911,541	5,070
Islamic	301,948.33	141.69	28.00	520.10	0.01	15,200	77,395,849	19,707
Japan	6,832.2	89.90	35.21	218.24	5.5	1,650	94,962,355	2,675
Medieval	245,707.25	29.39	15.51	68.90	0.01	2,332.17	34,234,785	7,511.57
Mongol	2,075.03	86.46	25.50	113.24	0.1	345.87	66,496,171	894.5
Near east	481,840.92	48.65	19.06	211.55	0.01	15,200	26,778,633	31,952.84
Pre-historic	179,245.68	35.93	16.14	83.68	0.01	2,738	90,343,768	11,840.78
Roman	1,560,321.05	45.74	21.51	114.56	0.01	12,900	97,422,287	91,833.83
Russian	89,757.35	51.70	25.00	116.64	0.99	2,650	66,496,171	13,761.14
Unknown	598,379.05	51.05	19.00	154.65	0.01	7,000	34,234,785	17,654.5
Viking	546,653.48	42.12	20.00	91.85	0.01	3,051.55	97,422,287	29,674.8
Coin	111,451.77	31.60	15.00	78.16	0.01	2,325	25,238,208	9,164.08
Decoration	124,536.45	45.02	23.83	76.82	0.01	1,451.58	34,234,785	16,419.04
Household	129,600.51	37.81	18.50	85.64	0.01	1,775	97,422,287	5,515.81
Jewellery	1,511,003.47	33.75	17.50	73.54	0.01	3,300	97,422,287	58,279.32
Mask	37,706.11	113.92	49.98	203.91	0.01	2,332.17	68,239,848	1,959
Religious	710,416.99	39.82	23.30	74.72	0.01	4,750	97,422,287	31,662.13
Statue	750,678.86	60.41	39.25	100.15	0.01	3,800	97,422,287	69,038.18
Text	301,045.52	63.70	32.49	320.08	0.01	15,200	97,422,287	18,784.78
Tool	562,755.48	40.40	20.00	117.81	0.01	10,000	97,422,287	27,901.5
Unknown	767,625.08	70.99	24.99	208.57	0.01	7,000	97,422,287	15,871.88
Vessel	615,332.01	71.14	38.05	186.09	0.01	5,249	97,422,287	24,706.49
Weapon	556,968.05	53.03	19.99	131.17	0.01	2,738.88	34,234,785	25,318.27
Bone	23,163.05	42.97	26.88	55.69	0.01	500	34,234,785	2,426.7
Glass	223,997.6	33.94	20.65	46.56	0.01	1,000	97,422,287	11,868.24
Ivory	139.14	46.38	54.00	17.88	21.69	63.45	34,234,785	85.14
Leather	3,134.94	27.50	0.99	69.50	0.01	461	30,249,188	350.67
Metal	2,546,570.29	48.35	21.19	135.92	0.01	10,000	97,422,287	14,4241.99
Papyrus	9,952.31	88.86	56.50	124.86	0.99	1,000	56,654,242	1,995.59
Stone	708,300.08	42.83	18.50	102.72	0.01	3,850	97,422,287	31,938.78
Terracotta	463,468.97	49.64	32.49	68.96	0.01	1,800	97,422,287	31,903.07
Unknown	1,341,573.27	47.83	20.50	174.80	0.01	15,200	26,778,633	42,369.63
Wood	46,621.3	86.98	49.98	177.48	0.01	2,500	26,778,633	4,732.28

Data includes standard deviation (SD) and other summary statistics. Seller data is for the top seller in each category. Seller total sales data (USD) is from 8 July 2019 to 3 January 2020.

have disproportional influence on items such as statues (48% of sales) and glass ( $\approx 59\%$ ; Figure 3(b) and (c)). Overall, the top 10 sellers sold about 40% of objects from July 2019 to January 2020.

The relative dominance that top sellers command is assessed here by looking at the percentage of sales by

the top seller for each given object type and material in the cultures that sold the most (Figure 4). Generally, objects that sold less are dominated by the top seller, although even common objects show evidence that one seller in many cases had a large sales percentage. In the case of Roman objects, about 44% of masks and

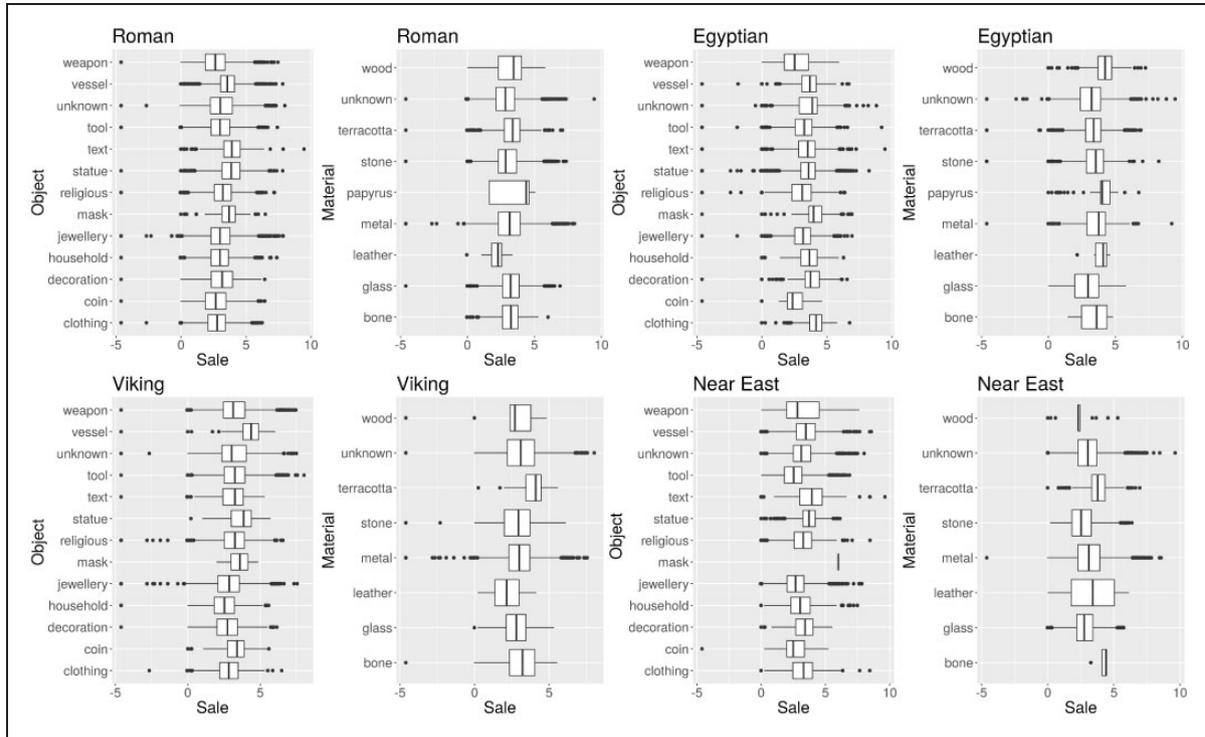


Figure 2. Distribution of sales (natural log) for different object types and materials for the four most common cultures sold.

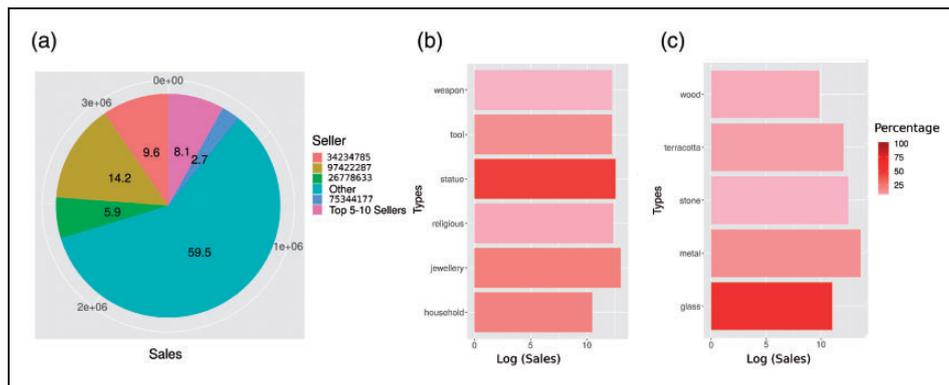
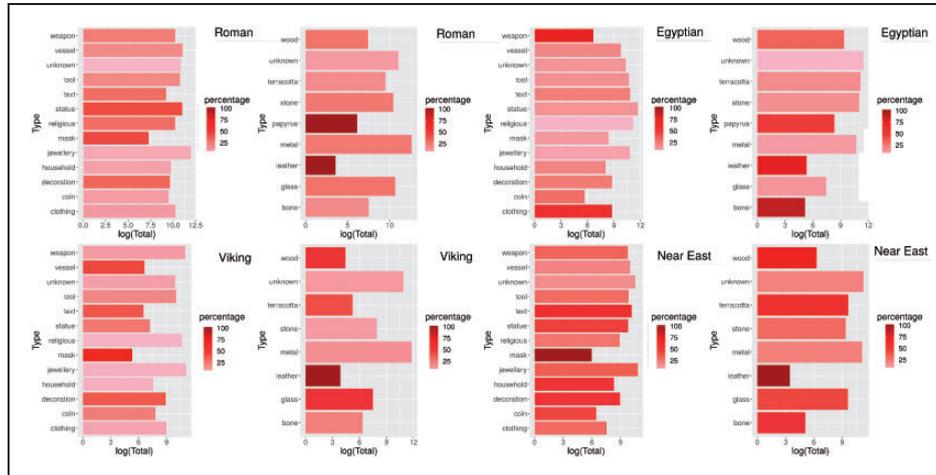


Figure 3. Seller data covering 8 July 2019 to 3 January 2020. Charts show percentage of sales in total (a) for the top 10 sellers and for object (b) and material (c) types for the top four sellers.

statues sold are by one seller. Even for more commonly sold objects, such as jewellery (15%) and weapons (30%), one seller has a relatively large percentage of sales. All of Roman leather and papyrus are sold by one seller; about 27% of glass and wood sales are by one seller. The top seller sold nearly 25% of metals, the most commonly sold type of material. For Egyptian objects, clothing (56%) and weapons (70%) are mostly sold by the top seller. The largest seller of texts, which had over \$58,000 in sales for the assessed period, sold about 29% of all text objects. The top seller sold at least 11% of any given object type. One seller sold more than 67% of all leather and 74% of all

bone, which are relatively rare objects; more common materials such as metals (24%) and terracotta (29%) also had many sales by the top seller. Viking objects show a similar pattern as the other cultures; objects that are rare, such as texts (42%), vessels (46%) and masks (63%), show a dominant sale pattern by one seller. More common objects, such as jewellery, weapons, religious and tools, show ranges between 15% and 28% of sales by one seller. Common materials included metal objects, which have 20% of sales by one seller. Less commonly sold objects, including wood, glass and leather, have over 50% of sales by one seller. For the Near East, texts sold for the assessed period are over



**Figure 4.** Sales percentages (log USD) for the top four cultures (Roman, Egypt, Viking, Near East) by the top seller in each object and material category assessed between 8 July 2019 and 3 January 2020.

**Table 4.** Top 10 country-level results showing top cultures, objects and material types sold (USD) along with the top seller and the total sold by that seller.

Country	Total	Culture	Object	Material	Seller	Seller sales
United Kingdom	1,860,706.74	Roman	Jewellery	Metal	97,422,287	223,289
United States	784,652.43	Unknown	Jewellery	Metal	38,893,829	19,908
Cyprus	308,884.87	Roman	Jewellery	Metal	34,234,785	144,611
Thailand	244,156.61	Roman	Jewellery	Metal	53,002,313	2,773
Germany	223,628.49	Pre-historic	Weapon	Stone	87,180,606	14,624
Egypt	206,905.40	Egyptian	Statue	Metal	77,395,849	19,707
Ukraine	122,030.35	Viking	Weapon	Metal	66,496,171	30,369
Bulgaria	96,715.74	Roman	Jewellery	Metal	7,842,328	4,933
Russia	64,162.97	Viking	Jewellery	Metal	92,138,303	16,047
Austria	60,564.29	Roman	Jewellery	Metal	60,329,669	5,296

Note that seller data is from 8 July 2019 to 3 January 2020.

\$26,000, with one seller selling more than 60% of texts, in large part because the top sale of any sale was a text (inscribed silk) object. Even jewellery, a common object sold, had one seller selling more than 36% for all jewellery sales. For the Near East, all objects analysed showed that the top seller sold at least 15%, with Figure 4 making it evident that 10 objects types have over 30% of sales by one seller. For materials, somewhat common terracotta has 61% of sales by one seller. Over 20% of metal sales are by one seller and the ratio is higher (30%) for stone.

### Country-level outputs

Outputs are also displayed based on where sales occur. Table 4 lists summary results for the top 10 selling countries for culture, object type and material composition for objects. The UK and United States alone represent more than half of all sales. Cyprus, a

relatively small country with less than 2 million people, is third in total sales. Overall, Europe and North America are the two continents that have the most sales. For all countries where data is present, the top culture sold (a) and the total sales (b) are displayed (Figure 5). There is some spatial patterning in results, although not for all cultures. Viking objects are the top culture sold in eastern Europe; Near East objects are the top culture sold in some Middle East countries, but they are also the top cultures sold in Italy, Indonesia and Hong Kong. Egyptian objects are the top objects sold in Egypt but also elsewhere. Most top selling Roman object countries are in Europe, although outside of Europe it is also a top seller, while mostly Islamic countries tend to sell Islamic objects as their top culture. The top sellers in the UK (\$223,289; 34%), Cyprus (\$144,610; 91%), Ukraine (\$30,368; 48%) and Israel (\$27,301; 84%) have a relatively high portion of those countries' sales for dates where

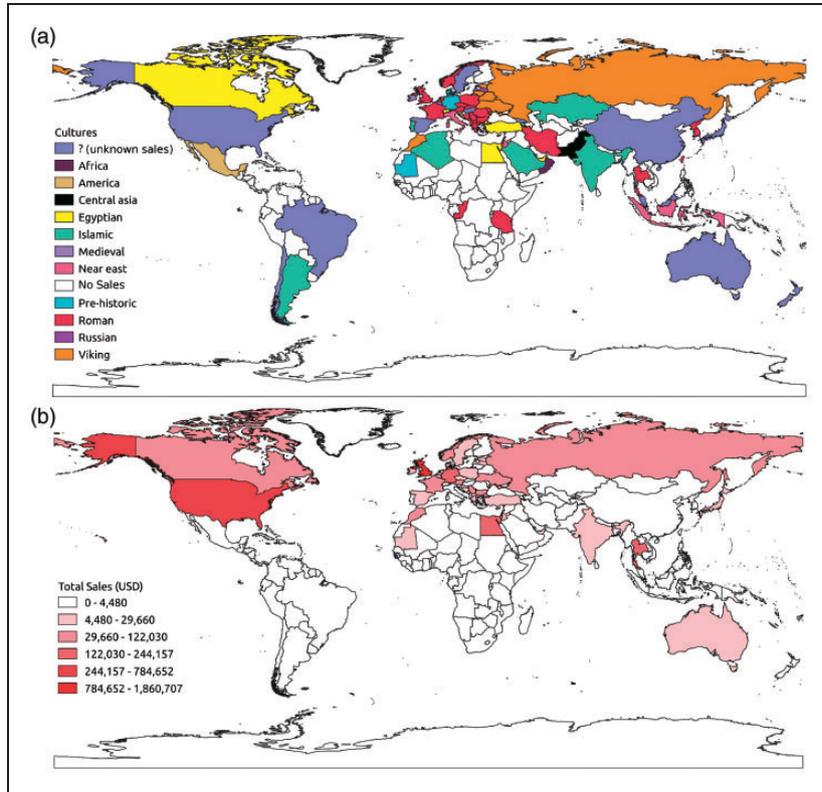


Figure 5. Maps showing where cultures were sold (a) and total sales for countries (USD) (b).

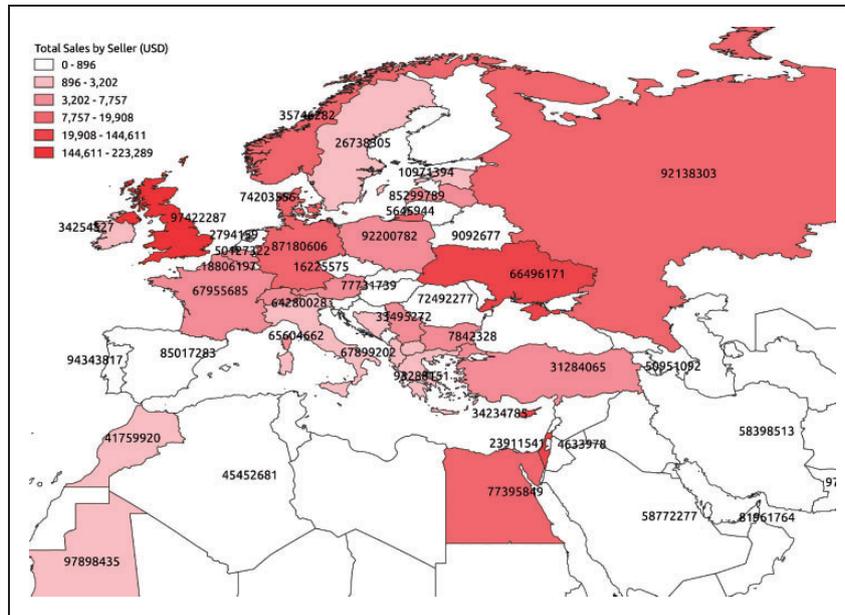


Figure 6. Top sellers indicated and their total sales (USD) in Europe from 8 July 2019 to 3 January 2020. Usernames are provided as anonymised numbers.

data is available. In Germany and Egypt, two other high selling countries, top sellers are not as dominant, with less than 19% of sales for the top sellers (Figure 6).

Looking at all sales for cultures, object types and material composition, it is evident that there is a strong positive correlation, using a multiple correlation analysis, between many types of objects across

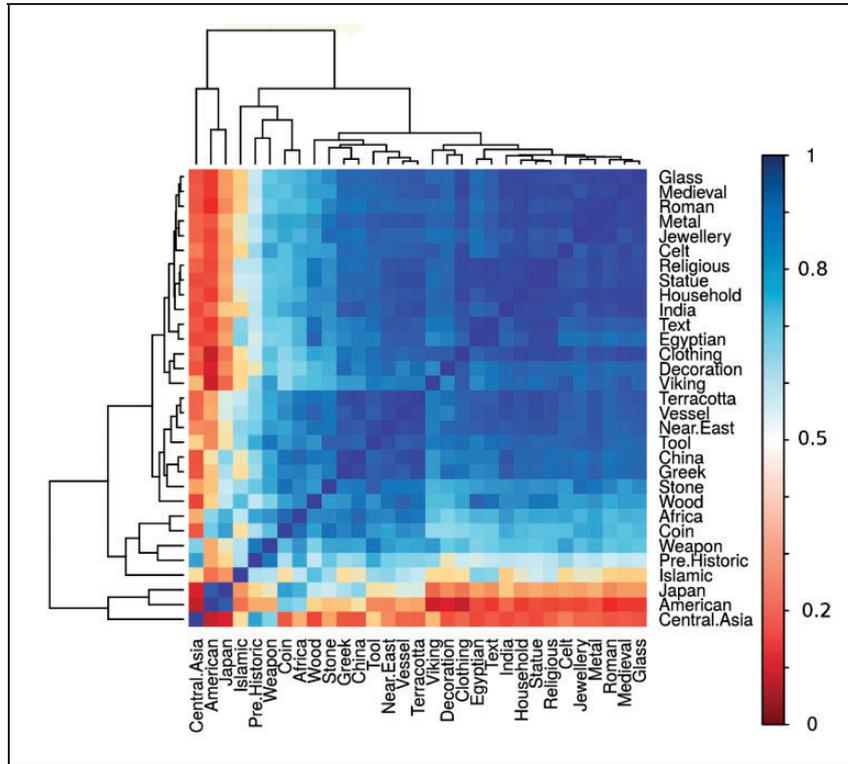


Figure 7. Multiple correlation heat map for culture, object type and material composition for objects sold in different countries.

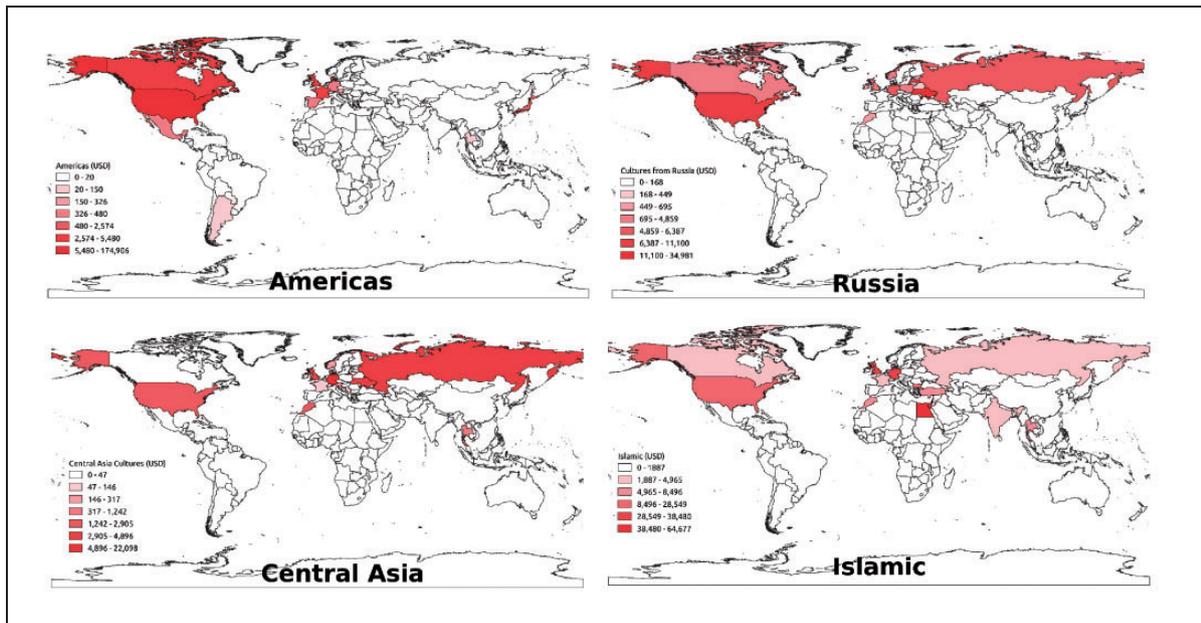


Figure 8. Some cultures and where they sold (USD).

countries (Figure 7). This suggests that these types of items tend to sell together in countries where these objects sell. Countries that tend to sell a high volume of objects generally sell a broad range of items that cover different cultures, object types and material

composition. Exceptions for items that do not have close positive correlation to sales of other items include Central Asian, American, Japanese, Islamic and Pre-historic objects. Object and material types tend to have strong correlations with other items, although

weapons and coins have relatively weaker correlations, indicating that many different object types are often found together in countries where they are sold. Wood also has a relatively weak correlation with other objects compared to other material types sold.

Despite evidence that a broad range of items have strong correlations in sales in given countries, one can also look at specific countries to see where given items may sell more. Given that many cultures are most commonly sold in the UK or US, there is evidence that

some countries do appear to specialise or at least focus more of their sales towards given cultures (Figure 8). For instance, cultures of the Americas are mostly sold in North America, with the US (\$174,906 sold) leading in sales, but Canada is the second highest seller. For Russian cultures, Ukraine (\$34,981) and Latvia (\$11,100) sold the most. Cultures from Central Asia are mostly sold in Germany (\$22,098), with the UK and Russia the second and third top sellers. For Islamic cultures, Egypt (\$64,676) and Germany

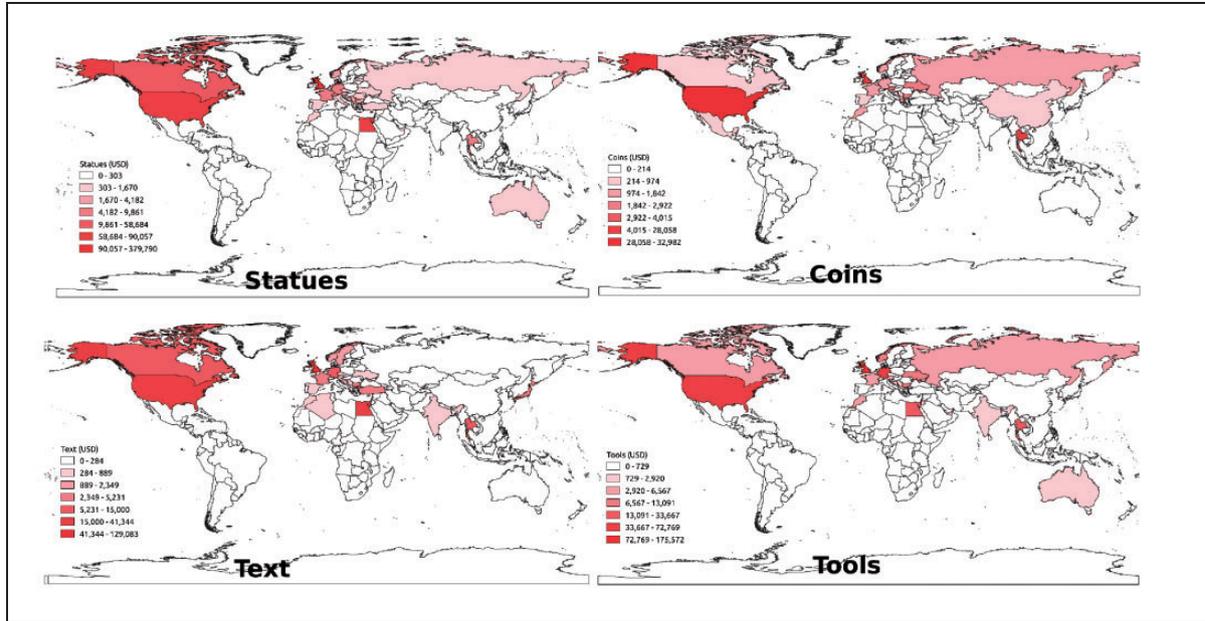


Figure 9. Some common and less common objects types sold (USD) in countries.

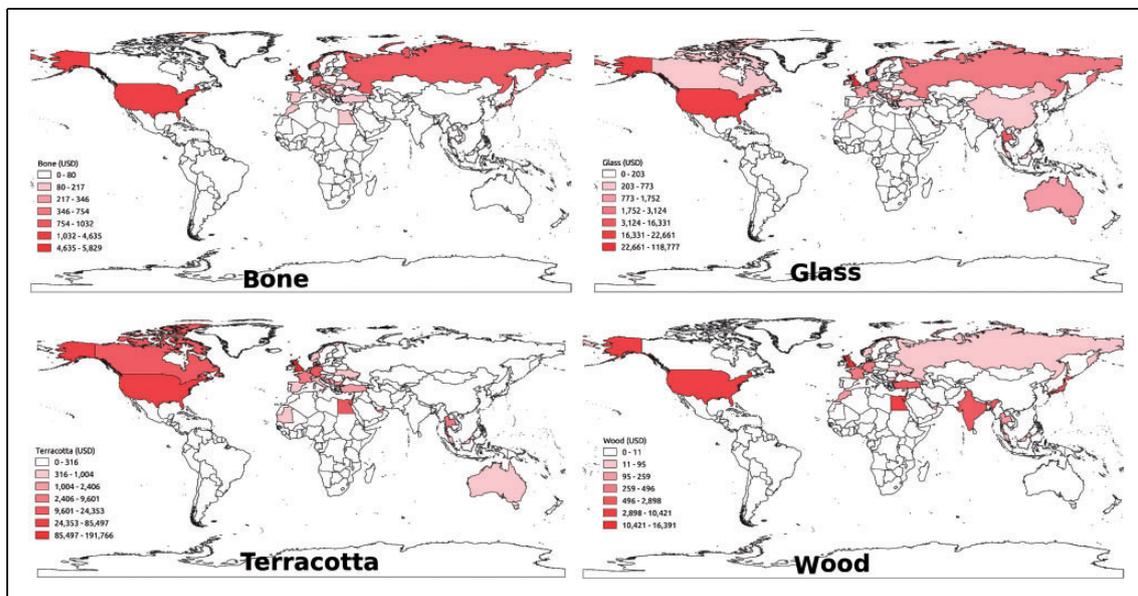


Figure 10. Some less common and more common material types sold (USD) in countries.

(\$38,480) are the top sellers. For object types, Western Europe and North America dominate sales, with some notable patterns evident (Figure 9). Statues, which also include figurines, are mostly sold in the UK (\$379,790); Egypt (\$69,339), however, is the third highest seller after the US (\$90,057). For coins, the US (\$32,982) is the leading seller, with Thailand (\$12,866) being the third highest seller. For texts, the UK (\$129,083) is the highest seller, with Egypt (\$24,817) being the second highest. The UK (\$175,572) and US (\$72,769) also lead in selling tools respectively, with Thailand (\$33,667), Egypt (\$19,048) and Israel (\$8,127) being the 5th, 6th and 10th leading sellers. Most material types sold in the major selling countries such as the UK, US and others in the top 10 sellers, but some more rare materials are relatively sold more by other countries not among the more typical leading sellers (Figure 10). Relatively rare items include bone, which is mostly sold in the UK (\$5829) and Cyprus (\$4635). The UK (\$118,777) also led in glass sales, with the US (\$22,661), Cyprus (\$21,042), Thailand (\$16,331) and Israel (\$7,602) making the top five sellers in that category, respectively. For terracotta, a relatively more common material sold, the UK (\$191,766), US (\$85,498), Cyprus (\$44,149), Egypt (\$24,353) and Canada (\$18,060) lead sales. The top five sellers for wood materials are the UK (\$16,391), US (\$10,421), Egypt (\$8,610), India (\$2,898) and Israel (\$677).

## Discussion and conclusion

### Key results

Few works have applied machine learning and quantitative assessment on online antiquities sales. This is among the first works to do this, with results showing multiple key insights on eBay's antiquities trade. The results indicate that Roman, jewellery and metal objects are, by far, the most common antiquities sold for cultures, object types and materials, respectively. Many sales are by individual or corporate sellers, with the top 10 sellers selling around 40% of objects from July 2019 to January 2020. Sales mostly concentrate in Europe, with the UK dominating, or North America, with the US a key seller. Other major sellers are Thailand, the fourth highest seller, and Egypt, which is the sixth highest seller. Cyprus is a surprising third highest seller, a high position given the country's small size. For Cyprus, one seller sold 91% of that country's objects sold between July 2019 and January 2020. Other countries that had a high portion of their sales by one seller include the UK, Ukraine and Israel. A variety of cultures, object types and materials appear to sell together in countries, indicating countries that sell antiquities offer both common and more niche

objects. Although Western countries generally dominated these sales, countries that had generally lower sales are among the top sellers in some categories. This includes Canada, Latvia, India and Israel as leading countries in some culture, object and material types sold. In addition to Roman objects, Egyptian, Viking and Near East cultures sold the most. While jewellery is the top object type sold, statues, including figurines, and religious items are the second and third highest selling object types, respectively, when unknown or undetermined objects are excluded. Stone and terracotta are the second and third highest selling materials, excluding unknown or undetermined materials, respectively. Masks, papyrus and wood are objects that had relatively high selling prices, although these objects are relatively rare. Statues and glass objects appear to be dominated by four main sellers. For the top cultures, masks, leather, papyrus and bone items are often exclusively sold by one seller or sales are at least dominated by one seller.

The above results indicate that many objects sold are relatively cheap, as determined in previous work, with some objects likely being forgeries or copies (Fay, 2011). There is no way for this work to determine what objects sold are fakes or copies. Regardless of the number of forgeries or fakes, these results suggest what cultures are generally more valued by consumers. The results align with earlier work showing that Western countries dominate sales (Barker, 2018; Bowman, 2008) and that online sales do often have a disproportional representation by few sellers (Huffer and Graham, 2017). Cyprus was shown to be a leading seller; earlier studies have shown Cyprus to have been a major source country for illegal antiquities sold, although it is not clear how significant the country has been relative to other countries (Brodie and Renfrew, 2005). In relation to Thailand's relatively strong sales, the country has been known to sell objects particularly from neighbouring Cambodia, but the results do not indicate if objects sold on eBay specifically relate to this (Mackenzie and Davis, 2016). East Asia is seen as a growing region for antiquities sales; results suggest at least Thailand does fit this pattern potentially because of increased wealth in the region (Brodie, 2006). While East Asia could be experiencing increased demand in sales volume, demand is clearly focused towards Roman, Egyptian, Viking (Dane/Norse) and Near East objects. Western countries and countries that can provide these in demand cultures generally had higher sales. The results also demonstrate that these cultures that are valued by customers are also given higher cultural value, that is more interest and greater social focus in Western societies, through scholarship and public interest that has greatly focused on them. For these top four cultures, they are

commonly discussed in ancient history topics in public schools or even in public media in Western countries (Cooper, 2017; Roupp, 2010). Western societies often hold Roman ideals, such as the rule of law and government, and culture as a model (Hingley, 2001; Richard, 2011). The Near East and Egypt have long held fascination in the West and are seen as the birthplace of 'civilisation' and Western religions (Derricourt, 2015). The Vikings (Dane/Norse) are common in media and popular culture that discuss the past, including how they shaped Europe and the West during the early Medieval period (Harvard and Stadius, 2016).

Such results suggest cultural value does appear to align with monetary value or at least affect total sales in a direct manner. Cultural value helps to create greater interest and drive sales for given objects associated with given cultures. Another work (Brodie, 2014) has shown that leading countries selling antiquities, at least at the end of a long network of antiquities traders, often also correspond to those that have interest in top selling cultures such as those indicated by this work. The results demonstrated here correspond well with cultures that are associated with or are of great interest and influence to mostly North American and Western European audiences, who likely form most buyers online (Mackenzie and Yates, 2017). While cultural value is a construct developed within society, it shapes prices and demand witnessed on the art and antiquities market more broadly (Becker, 2008). The sales volume and dollar values for sales indicate the top selling cultures have greater market demand, with anecdotal, educational and media data all suggesting wider cultural interest in these sold cultures by buyers. A site such as eBay does make it possible to sell less common object types to specific buyers and interests, which is evident in the results. Less commonly sold items with relatively lower buyer interest included those associated with non-Western religions (e.g. Buddhist) or non-Western regions (e.g. Central Asia).

What is sold is also influenced by key sellers. Results, where available, show relatively few sellers seem to have the greatest influence in sales, reflecting the likely dominance of firms or larger sellers on eBay. Although there is a broad range of sellers, as Brodie (2015) has previously indicated, overall it is only a few sellers who characterise a large percentage of total sales even in online sites. Online platforms effectively make it easier for firms and enterprising individuals to reach much wider audiences than traditional auction houses. Furthermore, national laws seem to be of little consequence for sales, as both Cyprus and Egypt are among leading antiquities sellers even though they have strict national laws against antiquities sales (UNESCO, 2020). One can conclude eBay appears to be generally poor at ensuring objects sold comply with local laws

such as the total prohibition of selling antiquities, as in the case of Egypt, or checking to see if there is very clear permission for sales from local antiquities or heritage authorities, such as in the case of Cyprus. This supports and highlights the grey nature of this marketplace as discussed earlier (Fay, 2011, 2013).

### *Methodological implications*

The methodology deployed demonstrates utility in that the outputs achieved would need considerable time if one were to use non-automated methods. Ideally, dictionaries would not be needed and a strict NER using CRF would be sufficient. However, this requires training data that would take time to build. One goal to improve this work would be to gradually switch to a pure NER approach, but until this can produce sufficiently high F-scores the method deployed currently appears appropriate. In addition to gradually switching over to a pure NER approach, expanding the number of sites observed, such as major antiquities sellers and auction sites (e.g. Sothebys, 2020), could be another way to expand this work. That would help the approach to take a 'Big Data' perspective in capturing a broader variety of the antiquities market. However, current restrictions on some sites in obtaining data, including scraping blockers or limited APIs for data retrieval, make this endeavour challenging. Scaling efforts to have broader site coverage would not only enable a wider and more diverse understanding of the antiquities market and what is sold, but higher-end sales that eBay appears to mostly ignore could be better understood. Another potential approach could be to integrate image-based search, along with text analysis, which would enable both descriptive data and images to be combined and inform on the objects assessed and help classification (Cacko and Iwanowski, 2018). It was observed that descriptions by the dealers are sometimes incorrect in describing objects, suggesting a combination of image-based analysis and description could potentially improve NER classification. As stated, fakes or copies are also sold on eBay. One possibility is to conduct a future study to see if NLP methods, such as NER, could be used to determine descriptions used more commonly with fake or copy objects. Additionally, cross-checking items sold on eBay or other sites with national or international databases for protected heritage objects can help ensure antiquities are not sold illegally, particularly where objects have been known to be stolen. Overall, the approach presented facilitates large-scale data gathering and its understanding using machine learning, demonstrating its utility for understanding the antiquities market.

## Acknowledgements

We would like to thank the organisers of the 2019 *Digital Heritage in a World of Big Data Conference* held at the University of Stirling for inspiring this work.

## Data accessibility

The code and relevant scraped and analysed data, as well as data used for precision and recall testing, are all available in the following site: <https://discovery.ucl.ac.uk/id/eprint/10110929/>

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Mark Altaweel  <https://orcid.org/0000-0003-1807-3737>

## Supplemental material

Supplemental material for this article is available online.

## References

- Altaweel M (2019) The market for heritage: Evidence from eBay using natural language processing. *Social Science Computer Review*. Epub ahead of print 26 August 2019. DOI: 10.1177/0894439319871015; <https://doi.org/10.1177/0894439319871015>.
- Antiquities Trafficking and Heritage Anthropology Research (ATHAR) (2019) *Facebook's black market in antiquities*. Available at: <http://atharproject.org/report2019/> (accessed 1 December 2019).
- Atwood R (2004) *Stealing History: Tomb Raiders, Smugglers, and the Looting of the Ancient World*. New York: St. Martin's Griffin.
- Barker AW (2018) Looting, the antiquities trade, and competing valuations of the past. *Annual Review of Anthropology* 47: 455–474. DOI: 10.1146/annurev-anthro-102116-041320.
- Becker HS (2008) *Art Worlds*. Berkeley and Los Angeles: University of California Press.
- Bonacchi C, Altaweel M and Krzyzanska M (2018) The heritage of Brexit: Roles of the past in the construction of political identities through social media. *Journal of Social Archaeology* 18(2): 174–192.
- Bowman BA (2008) Transnational crimes against culture: Looting at archaeological sites and the “grey” market in antiquities. *Journal of Contemporary Criminal Justice* 24(3): 225–242.
- Brodie N (2006) An archaeological view of the trade in unprovenanced antiquities. In: Hoffman BT (ed.) *Art and Cultural Heritage: Law, Policy and Practice*. Cambridge: Cambridge University Press, pp.52–63.
- Brodie N (2012) *Uncovering the Antiquities Market*. Oxford: Oxford University Press.
- Brodie N (2014) The antiquities market: It's all in a price. *Heritage & Society* 7(1): 32–46.
- Brodie N (2015) The internet market in antiquities. In: Desmarais F (ed.) *Countering Illicit Traffic in Cultural Goods: The Global Challenge of Protecting the World's Heritage*. Paris: ICOM, pp.11–20.
- Brodie N, Kersel MM and Tubb KW (2006) Introduction. In: Brodie N, Kersel MM, Luke C, et al. (eds) *Archaeology, Cultural Heritage, and the Antiquities Trade*. Gainesville: University Press of Florida, pp.1–24.
- Brodie N and Renfrew C (2005) Looting and the world's archaeological heritage: The inadequate response. *Annual Review of Anthropology* 34(1): 343–361.
- Bundsbusch M, Dejori M, Stetter M, et al. (2008) Extraction of semantic biomedical relations from text using conditional random fields. *BMC Bioinformatics* 9(1): 207.
- Cacko A and Iwanowski M (2018) Image search enhanced by using external data sources and reasoning. In: Choraś M and Choraś RS (eds) *Image Processing and Communications Challenges 9*. New York: Springer International Publishing, pp.19–26.
- Cooper H (2017) *The Teaching of History in Primary Schools: Implementing the Revised National Curriculum*. London: Routledge.
- Derricourt RM (2015) *Antiquity Imagined: The Remarkable Legacy of Egypt and the Ancient Near East*. London: I.B. Tauris.
- Fay E (2011) Virtual artifacts: eBay, antiquities, and authenticity. *Journal of Contemporary Criminal Justice* 27(4): 449–464.
- Fay E (2013) *Trading in antiquities on eBay: The changing face of the illicit trade in antiquities*. Unpublished PhD Thesis, University of Keele, England.
- Finkel JR, Grenager T and Manning C (2005) Incorporating non-local information into information extraction systems by Gibbs sampling. In: *Proceedings of the 43rd annual meeting on association for computational linguistics - ACL '05*, Ann Arbor, Michigan, 2005, pp.363–370. Stroudsburg: Association for Computational Linguistics.
- Frese M (2015) Cultural practices, norms, and values. *Journal of Cross-Cultural Psychology* 46(10): 1327–1330.
- Fritze RH (2016) *Egyptomania: A History of Fascination, Obsession and Fantasy*. London: Reaktion Books.
- GitHub (2020) *eBayScraper*. <https://github.com/maltaweel/eBayScraper> (accessed 08 January 2020).
- Goutte C and Gaussier E (2005) A probabilistic interpretation of precision, recall and F-Score, with implication for evaluation. In: Losada DE and Fernández-Luna JM (eds) *Advances in Information Retrieval*. Berlin, Heidelberg: Springer, pp.345–359.
- Greenland F, Marrone JV, Topçuoğlu O, et al. (2019) A site-level market model of the antiquities trade. *International Journal of Cultural Property* 26(1): 21–47.
- Hardy SA (2016) *Illicit trafficking, provenance research and due diligence: The state of the art*. Report, United Nations

- Educational, Scientific and Cultural Organization, Paris, France.
- Harvard J and Stadius P (2016) *Communicating the North: Media Structures and Images in the Making of the Nordic Region*. London: Routledge. Available at: <https://www.taylorfrancis.com/books/e/9781315572925> (accessed 8 April 2020).
- Hingley R (ed.) (2001) *Images of Rome: Perceptions of Ancient Rome in Europe and the United States in the Modern Age*. Journal of Roman Archaeology Supplementary series. Portsmouth, RI: Journal of Roman Archaeology.
- Huffer D and Graham S (2017) The Insta-dead: The rhetoric of the human remains trade on Instagram. *Internet Archaeology* (45). <https://intarch.ac.uk/journal/issue45/5/index.html>
- Kamath U, Liu J and Whitaker J (2019) *Deep Learning for NLP and Speech Recognition*. Cham: Springer.
- Kersel MM (2006) From the ground to the buyer: A market analysis of the trade in illegal antiquities. In: Brodie N, Kersel MM, Luke C, et al. (eds) *Archaeology, Cultural Heritage, and the Antiquities Trade*. Gainesville: University Press of Florida, pp.188–205.
- Konkol M, Brychcín T and Konopík M (2015) Latent semantics in named entity recognition. *Expert Systems with Applications* 42(7): 3470–3479.
- Kotha S and Basu S (2011) Amazon and eBay: Online retailers as market makers. In: Hamilton GG, Petrovic M and Senauer B (eds) *The Market Makers: How Retailers Are Reshaping the Global Economy*. Oxford: Oxford University Press, pp.155–180.
- Lafferty J, McCallum A and Pereira F (2001) Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: *Proceedings of the 18th international conference on machine learning*, Boca Raton, FL, 16–19 December 2001, pp.282–289. San Francisco: Morgan Kaufmann.
- Lee C, Hwang Y-G, Oh H-J, et al. (2006) Fine-grained named entity recognition using conditional random fields for question answering. In: Ng HT, Leong M-K, Kan M-Y, et al. (eds) *Information Retrieval Technology*. Berlin, Heidelberg: Springer, pp.581–587.
- McCallum A and Li W (2003) Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In: *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003*, Edmonton, Canada, 2003, pp.188–191. Stroudsburg: Association for Computational Linguistics.
- Mackenzie S and Davis T (2016) Temple looting in Cambodia: Anatomy of a statue trafficking network. In: Charney N (ed.) *Art Crime*. London: Palgrave Macmillan, pp.179–199.
- Mackenzie S and Yates D (2017) What is grey about the “grey market” in antiquities? In: Beckert J and Dewey M (eds) *The Architecture of Illegal Markets: Towards an Economic Sociology of Illegality in the Economy*. Oxford: Oxford University Press, pp.70–86.
- Manning C, Surdeanu M, Bauer J, et al. (2014) The Stanford CoreNLP natural language processing toolkit. In: *Proceedings of 52nd annual meeting of the association for computational linguistics: System demonstrations*, Baltimore, MD, USA, 2014, pp.55–60. Stroudsburg: Association for Computational Linguistics.
- Nadeau D and Sekine S (2007) A survey of named entity recognition and classification. *Linguisticae Investigationes* 30(1): 3–26.
- Nadeau D, Turney PD and Matwin S (2006) Unsupervised named-entity recognition: Generating gazetteers and resolving ambiguity. In: Tawfik AY and Goodwin SD (eds) *Advances in Artificial Intelligence*. Berlin, Heidelberg: Springer, pp.266–277.
- Powers DMW (2011) Evaluation: From precision, recall and F-Measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies* 2(1): 37–63.
- Powers DMW and Turk CCR (1989) *Machine Learning of Natural Language*. London: Springer.
- Richard CJ (2011) *Why We're All Romans: The Roman Contribution to the Western World*. Lanham: Rowman & Littlefield.
- Ritter A, Clark S, Etzioni M, et al. (2011) Named entity recognition in tweets: An experimental study. In: *Proceedings of the conference on empirical methods in natural language processing (EMNLP '11)*, Stroudsburg, PA, USA, July 2011, pp.1524–1534. Stroudsburg: Association for Computational Linguistics.
- Roupp H (ed.) (2010) *Teaching World History in the Twenty-First Century: A Resource Book. Sources and Studies in World History*. Armonk: M.E. Sharpe.
- Sothebys (2020) Sothebys Auction Site. Available at: <https://www.sothebys.com/> (accessed 21 January 2020).
- Stanford NLP (2020) The Stanford natural language processing group. Available at: <https://nlp.stanford.edu/software/CRF-NER.html> (accessed 3 January 2020).
- Stanish C (2009) Forging ahead. *Archaeology* 62(3): <https://archive.archaeology.org/0905/etc/insider.html>.
- Sutton C (2012) An introduction to conditional random fields. *Foundations and Trends® in Machine Learning* 4(4): 267–373.
- UNESCO (2020) UNESCO Database of national cultural heritage laws. Available at: <https://en.unesco.org/cultnat/laws> (accessed 21 January 2020).
- Walton JH (2006) *Ancient near Eastern Thought and the Old Testament: Introducing the Conceptual World of the Hebrew Bible*. Grand Rapids: Baker Academic.
- Wang X, Jiang X, Liu M, et al. (2017) Bacterial named entity recognition based on dictionary and conditional random field. In: *2017 IEEE international conference on bioinformatics and biomedicine (BIBM)*, Kansas City, MO, November 2017, pp.439–444. Piscataway: IEEE.
- Weisband E and Thomas CIP (2015) *Political Culture and the Making of Modern Nation-States*. Boulder: Paradigm Publishers.
- Zafarian A, Rokni A, Khadivi S, et al. (2015) Semi-supervised learning for named entity recognition using weakly labeled training data. In: *2015 The international symposium on artificial intelligence and signal processing (AISP)*, Mashhad, Iran, March 2015, pp.129–135. Piscataway: IEEE.