

An Archaeology of Textual Things

Gabriel Pizzorno, Digital History, Harvard University

Daniel Smail, Department of History, Harvard University

In the past few decades, the efforts of museums and other cultural institutions to digitize and make their collections accessible online have made massive amounts of information about artifacts available to the general public and researchers alike. From Harvard alone, one can obtain detailed information about hundreds of thousands of objects, ranging from those in museums, such as the Art Museums (250,000 objects) or the Peabody (700,000 records), to those of individual projects, such as the Sardis Expedition, featured in the Spring 2017 issue of *In Situ*.

This availability of information about *tangible things*, however, has not been matched by a similar increase in the access to information about *textual things*, that is, objects described in textual sources. Our project, the Documentary Archaeology of Late Medieval Europe (DALME), focuses on the latter, and aims to develop a publicly accessible and fully searchable online database of material culture that will enable researchers to seamlessly integrate object descriptions in contemporary documents, such as inventories, as the textual counterparts of objects found in museum collections and artefacts retrieved from in archaeological excavations.

In its current phase, our project focuses on a corpus of late-medieval household inventories and records of debt collection from France, Italy, and other regions of Europe (Figure 10). This type of record is found in both Europe and the Americas across a period of centuries, making it a stable platform for serial analyses of data that seek to make comparisons across time and space. Inventories usually represent a large proportion of a household's contents, thus

capturing a far broader spectrum of material culture than sources biased toward high-status objects.

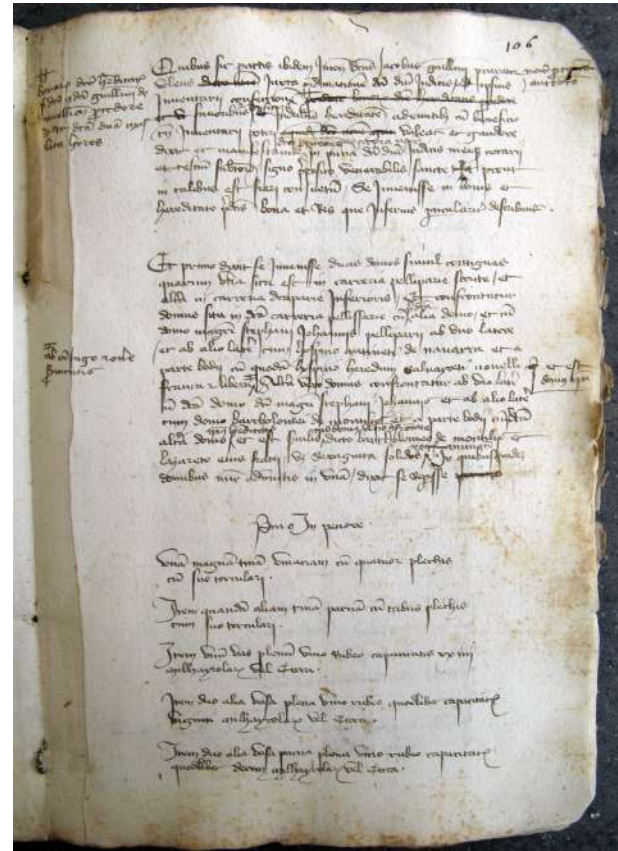


Figure 10: A folio from the inventory of the estate of Guilhem de Cavallhon, a relatively poor resident of Marseille who died in 1405. The inventory was requested by Guilhem's widow and heir, Massileta, in order to decide whether to accept the inheritance (if its debts were too great she might have chosen not to accept it), and conducted by her brother, Jacme Guilhem. (Archives Départementales des Bouches-du-Rhône 145, folio 106r, 12 November 1405)

The project relies on a sophisticated digital architecture that helps us to manage the project's data and documentation. At its most basic, the project's workflow digitization (i.e. generating machine readable versions of the documents), lexicalization (i.e. creating the

auxiliary data necessary to translate them), and parsing (i.e. reading and classifying their contents to generate database records of the objects described therein).

As we parse the documents to detect objects and their attributes, we semantically link these to concepts in a controlled vocabulary. For this we use, and extend where necessary, the Getty Thesaurus of Art and Architecture. By way of example, the words in various languages for the object known in English as a “chest” are each linked to a headword in their respective lexicons; those headwords, in turn, are linked to a higher-order entity bearing the ID “300038992” in the Getty thesaurus (for convenience we call this a “chest”). In this way, a simple search for “chest” will allow the user to find objects known in their particular languages as “cassia,” “capcia,” “caxia,” “archibanc,” and so on, as well as to access broader conceptual categories in the hierarchy, such as “case furniture,” “furniture,” “furnishings,” and so on. The same is true for headwords that describe

attributes. The words “blodeus” and “sanguineus”, for example, are both used to describe a deep, blood red color. The corresponding headwords are linked via the Getty ID “300310722”, “blood red” in English. This way, one can search the database for “blood red” with certainty that the results will include all objects described as such, regardless of the specific terms used. Similarly, by using the Getty taxonomy to which the concept belongs, one could ask the computer to return increasingly broader categories such as all “red colors”, all “chromatic colors”, or all objects with “physical attributes”. It is this taxonomy which lies at the very heart of the DALME database and provides enormous user flexibility in devising queries and analyses.

The simple idea behind our project is that textual sources, however imperfect, provide a useful complement to archaeological evidence. Inventories in particular are an especially valuable source for archaeological work because the systematic biases that generate silences in the

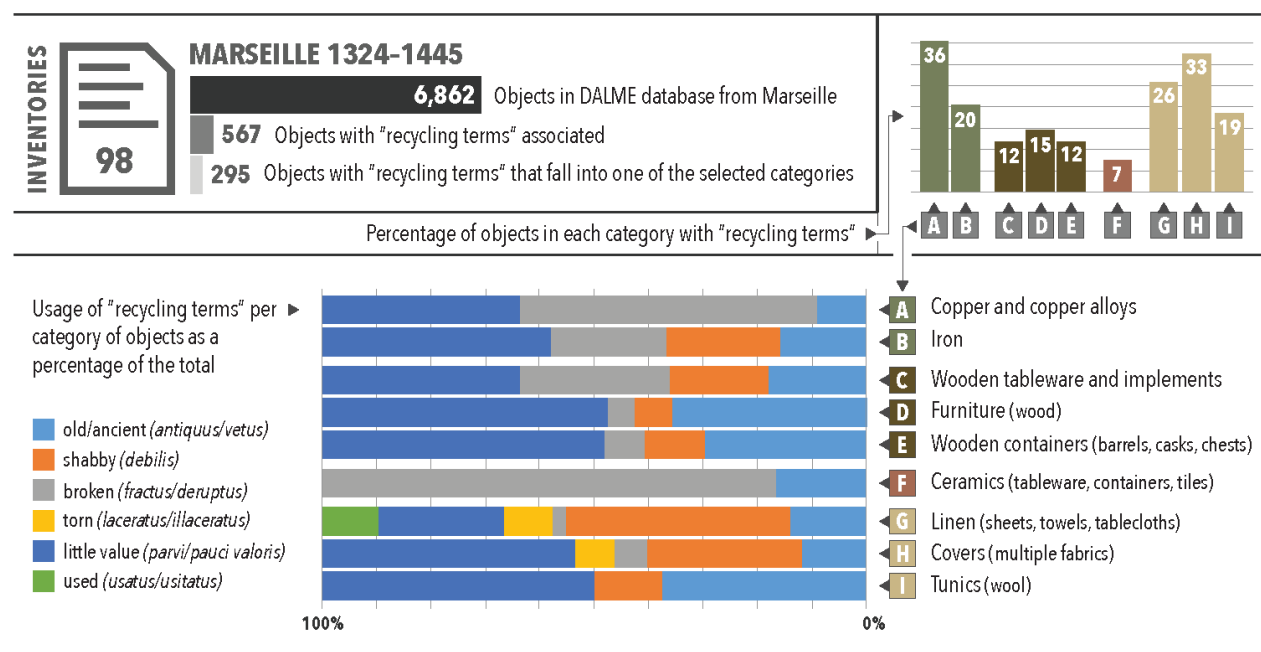


Figure 11: Explicit references to the recycling of objects are rare in our sources, but the frequent use of certain terms and expressions can be used to track implicit attestations of it. This graphic shows the usage of some of those terms for certain categories of material culture from inventories compiled in the city of Marseille between 1324 and 1445.

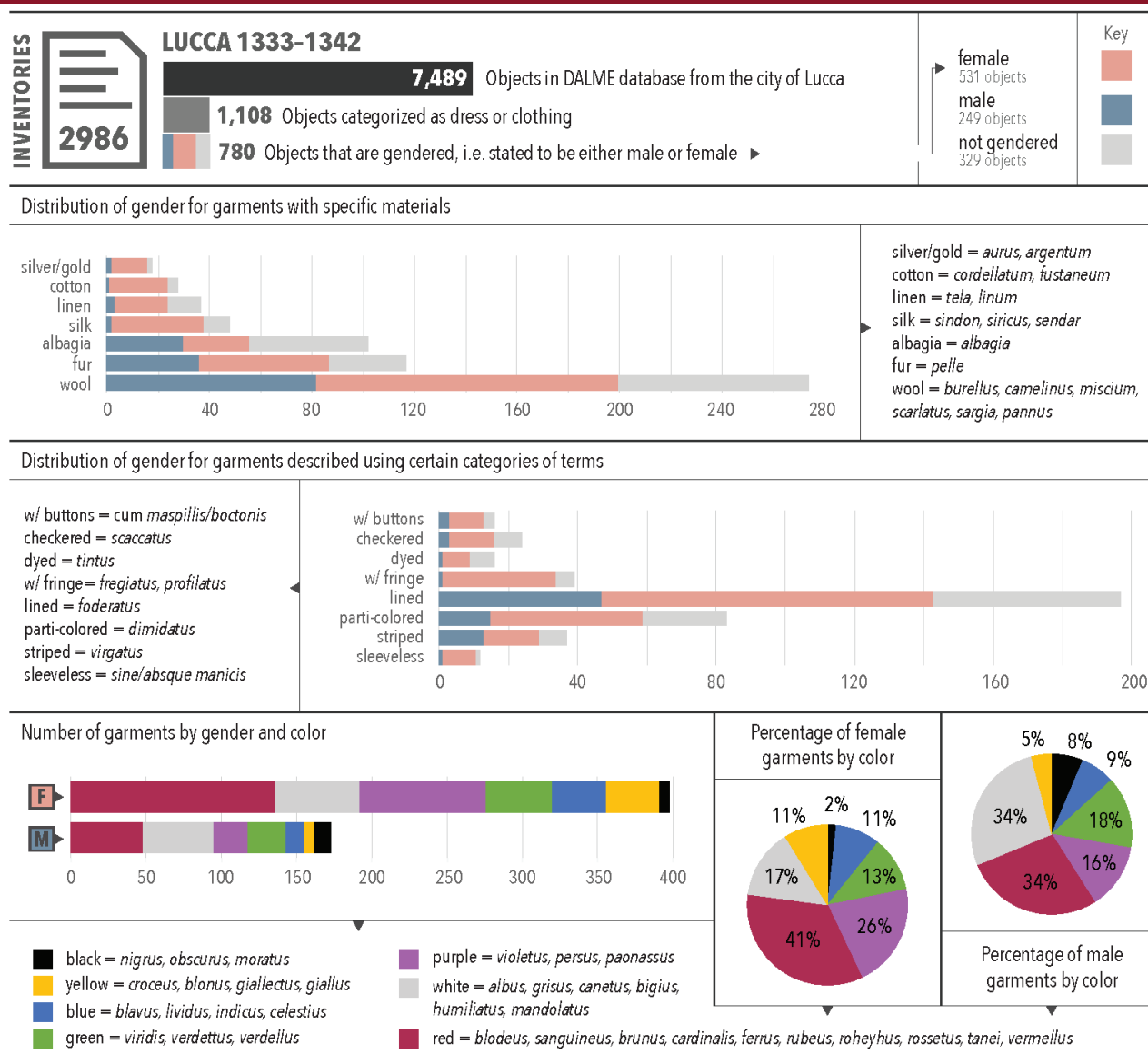


Figure 12: This graphic shows a preliminary analysis of items of clothing from the Italian city of Lucca between 1333 and 1342. Over 70% of the objects are gendered, that is, are explicitly described as being for men or women. A number of interesting patterns emerge from the dataset. If an object in the sample is gendered, it is twice more likely to be female than male. Of the materials in the sample, all but one appear to be gender-biased towards women, only albagia is predominantly used in male gendered items. Just as with materials, certain descriptive terms seem strongly linked to female objects, with fringed and dyed items being almost exclusively worn by women. Lastly, looking at the distribution of colors, red appears to be, by far, the most popular colour for women, while red and white seem equally popular with men. Preference for blue and green is similar for both genders.

textual record are often quite different from the systematic biases that responsible for differential preservation in the archaeological record. Indeed, the comparison of the evidence from each of the two domains will make the systematic biases inherent in each readily apparent.

In the medieval archaeological record, for example, evidence for the reuse of durable objects and materials is found everywhere. However, only some recycling processes leave behind traces that can be seen directly in the archaeological record. Historical sources are often equally fragmentary with respect to the information they offer about recycling, but the types of information provided complement

those found in the archaeological record. To take an example, the usage of certain expressions can be used to track implicit attestations of the phenomenon (Figure 11).

Inventories can also offer valuable information in the case of high-value materials that are more likely to be recycled, such as precious metals. For example, silver belt fittings in the textual sources from Marseille outnumber fittings made of copper alloy by a ratio of about 20:1. Among the extant archaeological finds from Marseille, however, there appear to be no silver belt fittings whatsoever, although we do find the occasional belt composed of copper alloy fittings. Likewise, documentary sources, inasmuch as they often record organic and inorganic materials in equal measure, can provide information that would otherwise be inaccessible to archaeology. An example is provided by the distribution of fabric colors and design details in clothing (Figure 12). Through textual sources, in other words, it is possible to

derive a more accurate picture of the assemblage of typical medieval households. Similarly, the abundance of documentary records can enable us to look for patterns at scales that would be difficult to approach with archaeological evidence alone (e.g. Figure 13).

In conclusion, our goal is to create a framework that allows us to look systematically at material culture in almost any historical context, regardless of whether the information about said objects comes from museums, archaeological excavations, or documents in archives. We view this framework as a guide that would enable us to map historically contingent descriptions of objects into attributes that can be managed alongside those generated by studying material objects themselves. The idea is not so much to provide a translation matrix (that's not possible for everything) but rather a systematic approach for managing assumptions about these objects and make them usable in rigorous research applications.

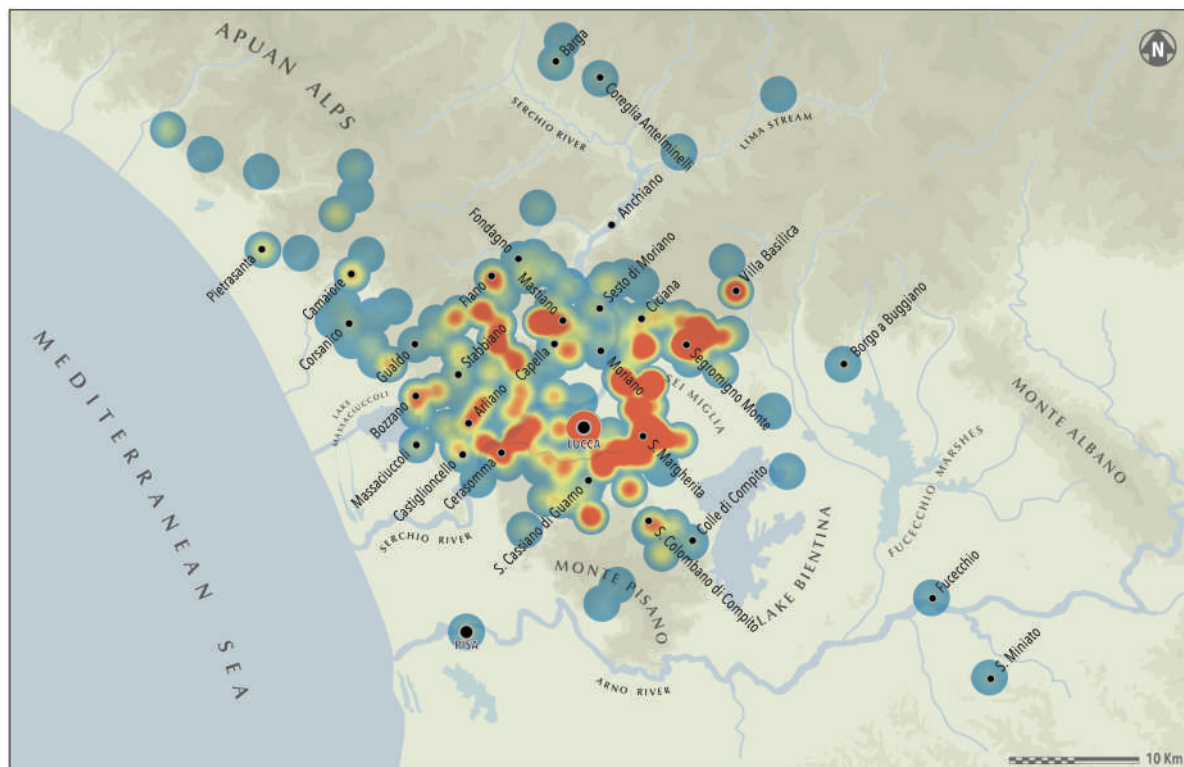


Figure 13: This heat map shows the density of *preda* (an object seized as part of debt collection) throughout the region surrounding the Italian city of Lucca between 1333-1342. The red shades indicate areas where predation was particularly intense.