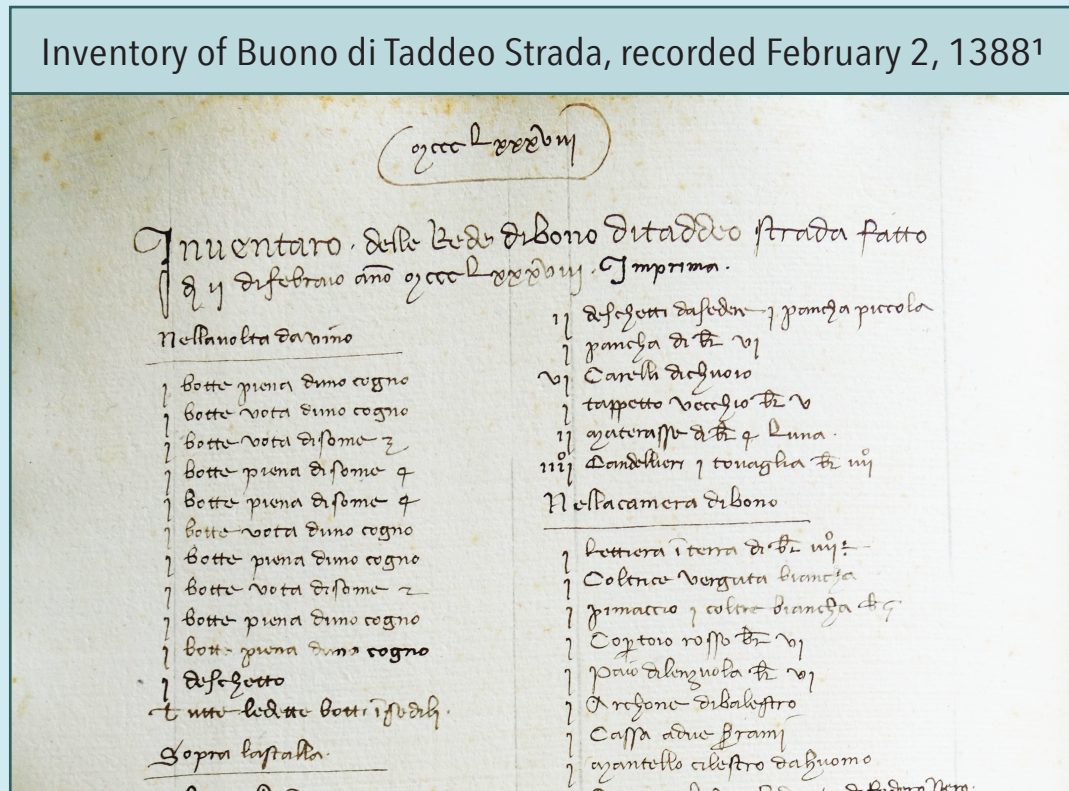


## DOCUMENTARY ARCHAEOLOGY

DALME is part of the larger **Initiative for Documentary Archaeology (IDA)**, whose goal is to lay the theoretical and methodological foundations for a novel approach to the study of past material culture. We call this approach a *Documentary Archaeology* because it enables researchers to **interrogate vast datasets of textual things alongside tangible material culture** by providing a framework that translates between the **modern domain ontologies** used by scholars to characterize museum objects and archaeological artefacts and the **historical folk taxonomies** used to describe objects by people in the past.

## INVENTORIES

Inventories are a type of record found across a period of centuries in Europe, making it a **stable platform for serial analyses** of data across time and space. They usually represent a **large proportion of a household's contents** and derive from a **spectrum of social ranks**, capturing a far broader range of material culture than sources oriented toward high-status objects. Inventories are particularly valuable because the **systematic biases that generate silences in the textual record are often quite different from those that affect the survival of tangible things**. Comparing evidence from each of the two domains makes the systematic biases inherent in each readily apparent and subject to statistical analysis. They also offer a crucial practical advantage: as legal documents, they tend to be **highly formalized and exhibit only a small range of variation** in terms of how the data is presented.



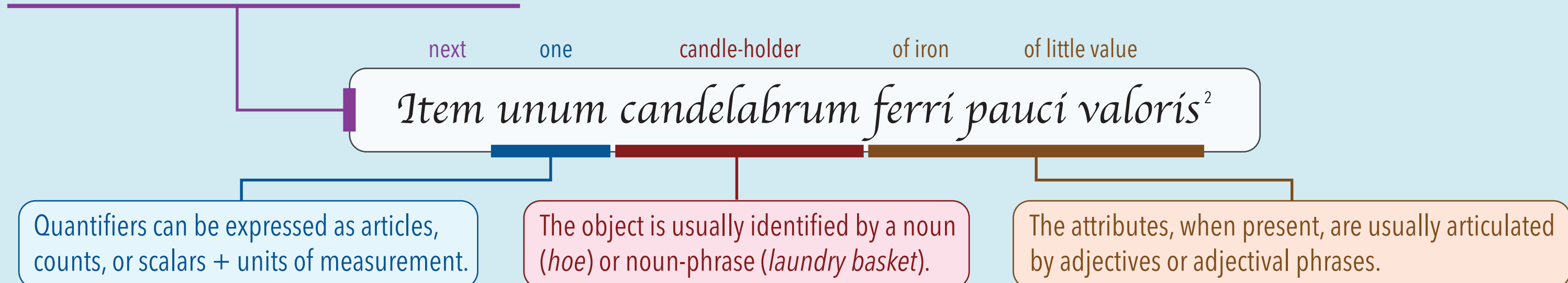
## DALME DATASETS

Initially, the project focused on a corpus of late-medieval household inventories and records of debt collection from **France and Italy**. We are now in the process of augmenting our existing dataset by transcribing and incorporating additional inventories from **Toulouse, Lleida, Sicily, Florence, Bologna, Tuscany, Montpellier, Valencia** and several other cities and regions.



## OBJECT DESCRIPTIONS

Object descriptions within inventories can usually be separated into discrete records consisting of strings of words with well-defined boundaries (eg **order-determining phrases, such as *first* or *next***). A record typically consists of a quantifier, the name of an object, and a list of its attributes:

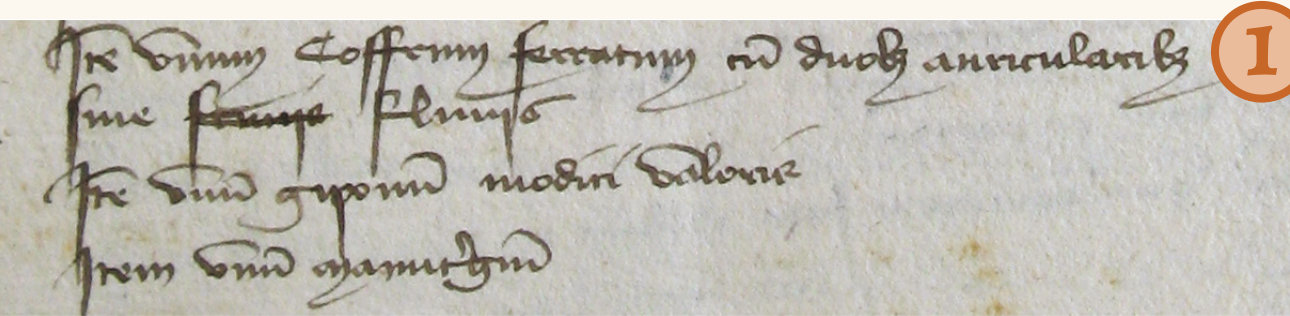


## OUR WORKFLOW

The project relies on a sophisticated digital architecture that helps us manage our data and documentation. Our basic workflow can be divided into three steps: **digitization**, **lexicalization**, and **semantic parsing**.

### DIGITIZATION

Digitization refers to the process of generating machine-readable versions of the inventories by using **Optical Character Recognition** software (in the case of published editions) or manually transcribing them (in the case of archival materials). In both cases we use **TEI mark-up** to tag key elements of the text to facilitate importing the contents into the database as well as preserving the format and editorial marks of the document.

**A**  **1**

**B**

**C**

**A** Item unum coffrum ferratum cum duobus auricularibus

**B** sive -frumis- flunis

**C** Item unum giponum modici valoris

Item unum manutergium

**A** <list type="inventory">

**B** <item>Item unum coffrum ferratum cum duobus auricularibus</b>

**C** sive<del rend="overstrike">frumis</del> flunis</item></b>

<item>Item unum giponum modici valoris</item></b>

<item>Item unum manutergium</item></b>

</list>

### LEXICALIZATION

Lexicalization consists of creating the auxiliary data necessary to translate the contents of the documents. This involves **tokenizing** the contents of the documents, **lemmatizing** the words, and associating them with entries in the corresponding **lexicon** module.

CORPUS	WORD LIST	
auricularibus	auricularibus	auricular, noun, oreiller (Du Cange)
coffrum	coffrum	coffus, noun, coffret (Razo)
cum	cum	cum, prep., with (Lewis & Short)
duobus	duobus	duo, adj. pl., two (Lewis & Short)
ferratum	ferratum	ferratus, noun, fere (Razo)
flunis	flunis	fluna, noun, taie d'oreiller (Razo)
frumis	frumis	
giponum	giponum	giponus, noun, jupon/jupe (Razo)
item	item	item, adv., besides, also (Lewis & Short)
manutergium	manutergium	manutergium, noun, torchon (Razo)
modici	modici	modicus, adj., moderate (Lewis & Short)
sive	sive	sive, conj., or if (Lewis & Short)
unum	unum	unus, prop., one (Lewis & Short)
valoris	valoris	valor, noun, value (Du Cange)

### SEMANTIC PARSING

Semantic parsing consists of **reading and classifying** the contents of the inventories to generate **database records** of the objects described therein.

From the three sample entries:

<b>A</b>	item	unum	coffrum	ferratum	cum	duobus	auricularibus	sive	flunis
<b>B</b>	item	unum	giponum	modici	valoris				
<b>C</b>	item	unum	manutergium						

Five object records are generated in the database:

<b>A</b>	1	COFFER	style :: iron-bound
	1	PILLOW / PILLOW CASE	
<b>B</b>	1	PILLOW / PILLOW CASE	
	1	JUPE	condition :: of little value
<b>C</b>	1	HAND TOWEL	

## HOW WE USE LINKED OPEN DATA

Our system uses a **faceted approach** that can assign different **scopes (temporal, spatial, formal)** to terms, in order to account for regional and temporal variations in spelling and meaning while **retaining semantic links** to the objects they describe. The latter is accomplished by defining said object as an abstract **concept** (reified only by a numeric ID in the database) to which as many terms as pertinent can be attached. As we parse the documents to detect objects and their attributes, we **semantically link** these to concepts in an **abstracted taxonomy of material culture**. For the latter we **use, and extend** where necessary, the Getty Research Institute's **Art & Architecture Thesaurus**.

2

Very different words can be used to describe attributes of an object that are in fact identical. The **two terms below are both used to describe a deep, blood red color**.

sanguineus  
blodeus

4

Likewise with objects: medieval Provençal used at least **fourteen different Latinized words to refer to the object we would call basket** in English.

banasta/banastonus  
bassellus  
benha  
belsa  
canestella/canestellus  
canistrum  
cofinus  
colla  
garbella/garbellinus  
gavanhola  
gorbinus  
panal  
sportinus  
sysra

vermillion  
incarnadine  
rose  
mauve  
blood red  
scarlet  
ruby  
cadmium  
Mars  
hematite  
crimson  
carmine  
alizarine  
lake  
magenta  
garnet  
geranium  
cochineal  
kermes  
spectrum  
para  
scarlet v.  
scarlet l.  
permanent  
dragon's blood  
cerise  
Tyrian purple

variable red  
vivid red  
strong red  
deep red  
moderate red  
dark red  
grayish red  
blackish red  
reddish gray  
reddish black  
purplish red  
red

color types  
neutrals  
chromatic colors  
achromatic colors  
brown  
yellow  
olive  
yellow-green  
green  
pink  
blue  
purple  
red  
orange

hues or tints  
color phenomena

Conditions and Effects  
Design Elements  
Attributes and Properties  
Color

**Linked Open data**  
GETTY VOCABULARIES

Semantic View (JSON, RDF, N3/Turtle, N-Triples)

ID: 300310722

blood red (variable red colors, red colors, ... Color (hierarchy name))

Note: Variable color term referring to any moderate to vivid red.

Terms:

- blood red (preferred, C,U,English-P,D,U,U)
- blood-red (C,U,English,U,F,U,U)
- 血紅色 (C,U,Chinese (traditional)-P,D,U,U)
- xiè hóng sè (C,U,Chinese (transliterated Hanyu Pinyin)-P,U,F,U,U)
- xie hong se (C,U,Chinese (transliterated Pinyin without tones)-P,U,F,U,U)
- hsieh hung se (C,U,Chinese (transliterated Wade-Giles)-P,U,F,U,U)
- bloodrood (C,U,Dutch-P,D,U,U)

Facet/Hierarchy Code: D.D.L

Hierarchical Position:

- Physical Attributes Facet
- Color (hierarchy name) (G)
- colors (hues or tints) (G)
- chromatic colors (G)
- red colors (G)
- variable red colors (G)
- blood red (G)

3

The terms **blodeus** and **sanguineus** are linked via the Getty ATT 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 ATT 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.

This abstracted taxonomy lies at the very heart of the DALME database and provides enormous user flexibility in devising queries and analyses.

Associated Concepts  
Physical Attributes  
Styles and Periods  
Objects  
Activities  
Materials  
Agents  
Brand Names

Genres  
Groupings and Systems  
Components  
Built Environment  
Furnishings and Equipment  
Visual/Verbal Comm

Tools and Equipment  
Vehicles  
Furnishings  
Sound Devices  
Containers  
Weapons and Ammo  
Recreational Artifacts  
Costume  
Measuring Devices

bronzes  
by location  
by function/context  
by form  
components

**FUTURE DIRECTIONS:** We believe that a number of other data points in our database can be managed in similar fashion to the objects and their attributes. We have recently begun to implement a similar solution for **geographic and personal names**.

## SOME RESULTS

**RECYCLING:** Textual sources, however imperfect, provide a useful complement to archaeological evidence. 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 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. Explicit references to recycling 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.

