

Luana Bulla, Maria Chiara Frangipane, Maria Letizia Mancinelli, Ludovica Marinucci, Misael Mongiovì, Margherita Porena, Valentina Presutti, and Chiara Veninata

Developing and Aligning a Detailed Controlled Vocabulary for Artwork

ISTC - National Research Council, Rome and Catania, Italy

LILEC - University of Bologna, Italy

ICCD - Ministry of Culture, Italy

OUTLINE

AIM OF THE WORK

MATERIALS

METHODOLOGY

- Matching and alignment algorithm
- Linking with Getty's AAT

RESULTS

FUTURE WORKS

Motivation

Controlled vocabularies enable overcoming the semantic heterogeneity commonly encountered in artwork catalogs and provide a common terminology for describing and cataloging artwork objects.

Most catalogs data rely on free text type descriptions, which makes classification arbitrary and integration between different sources difficult.

Aim

We aim to extend in order to improve the coverage of controlled vocabularies in the Italian cultural heritage domains.

We also aim to implement the limited number of Italian terms in Getty's Art & Architecture Thesaurus (AAT) through an automatic alignment process followed by manual validation.

MATERIALS

Getty: Conservation
Research
Foundation Museum



The Art & Architecture Thesaurus (AAT)¹
More than 400,400 terms of which only 8% have
associated Italian terms.

ArCo KG:
Architetture della
Conoscenza



Ontologies and Knowledge Graphs for the
description of italian
cultural heritage²
~ 2.5M records -
169 million triples about 820 thousand cultural entities

ICCD: Istituto
Centrale per il
Catalogo e la
Documentazione



Vocabulary of Artworks³
4.225 Italian terms,
organized into 7 general categories

Vocabulary of Archeological finds (VAF)⁴
1.784 Italian terms,
organized into 9 general categories
manual alignment with AAT

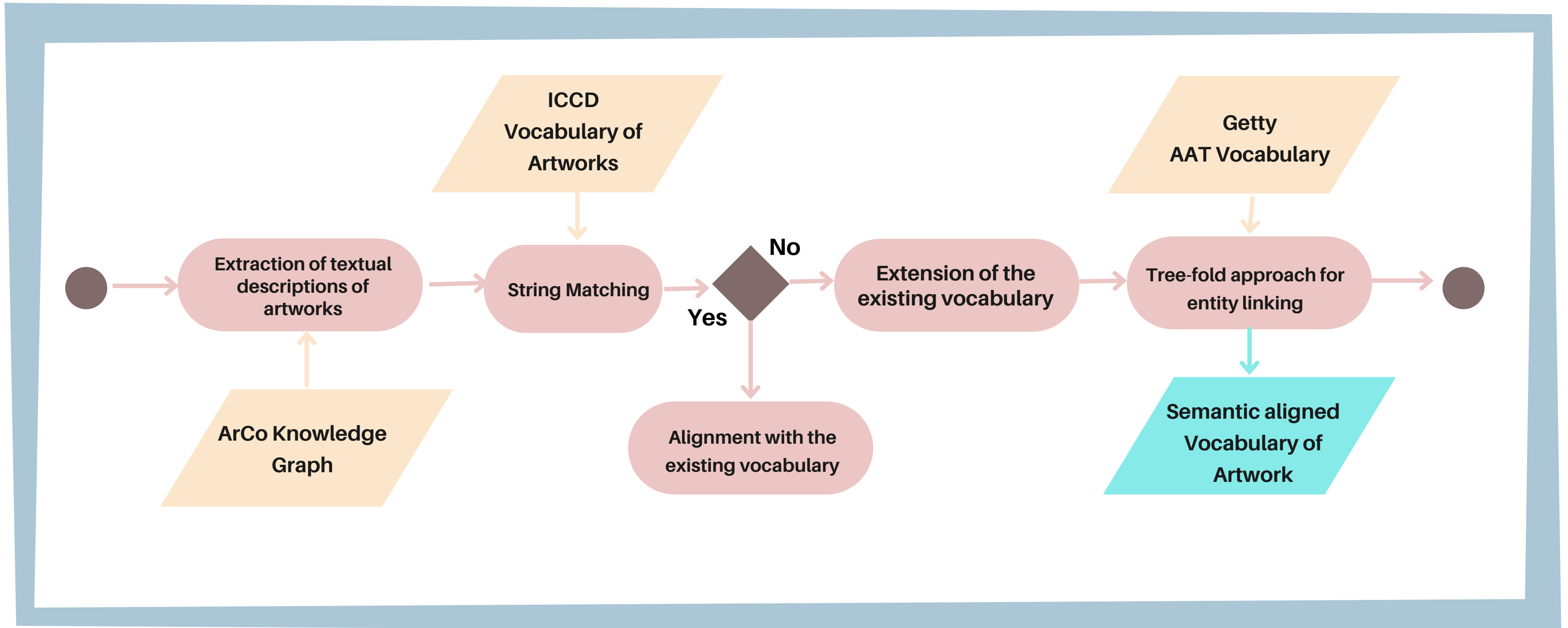
1. <http://www.getty.edu/research/tools/vocabularies/aat/aat%20faq.html#numbe%20r>

2. <http://wit.istc.cnr.it/arco/?lang=en>

3. <http://www.iccd.beniculturali.it/it/ricercanormative/139/thesaurus-perla-definizione-dei-beni-storici-artistici>

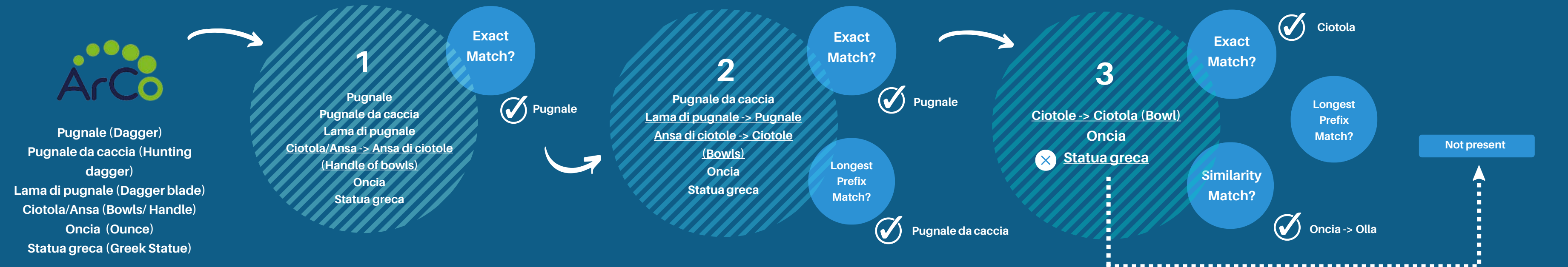
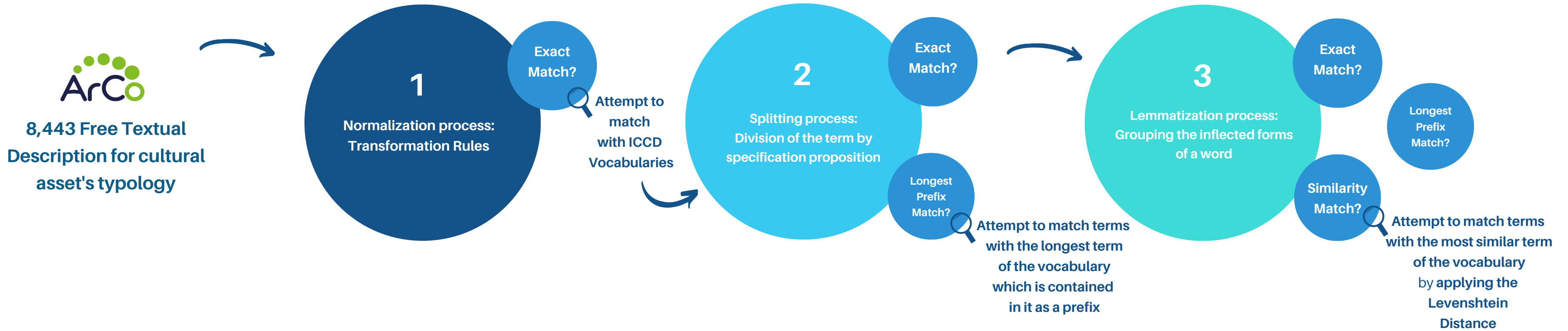
4. <http://www.iccd.beniculturali.it/it/ricercanormative/108/thesaurus-perla-definizione-dei-reperti-archeologici>

Methodology



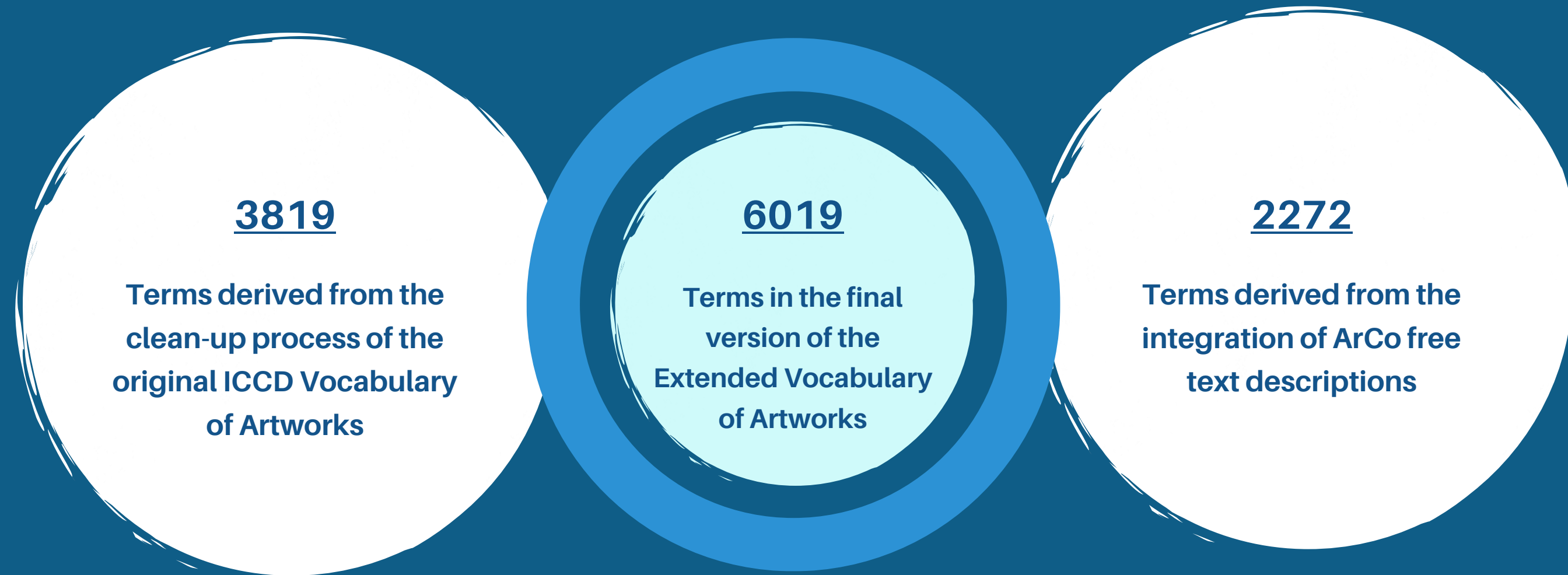
METHODOLOGY

Matching and alignment algorithm



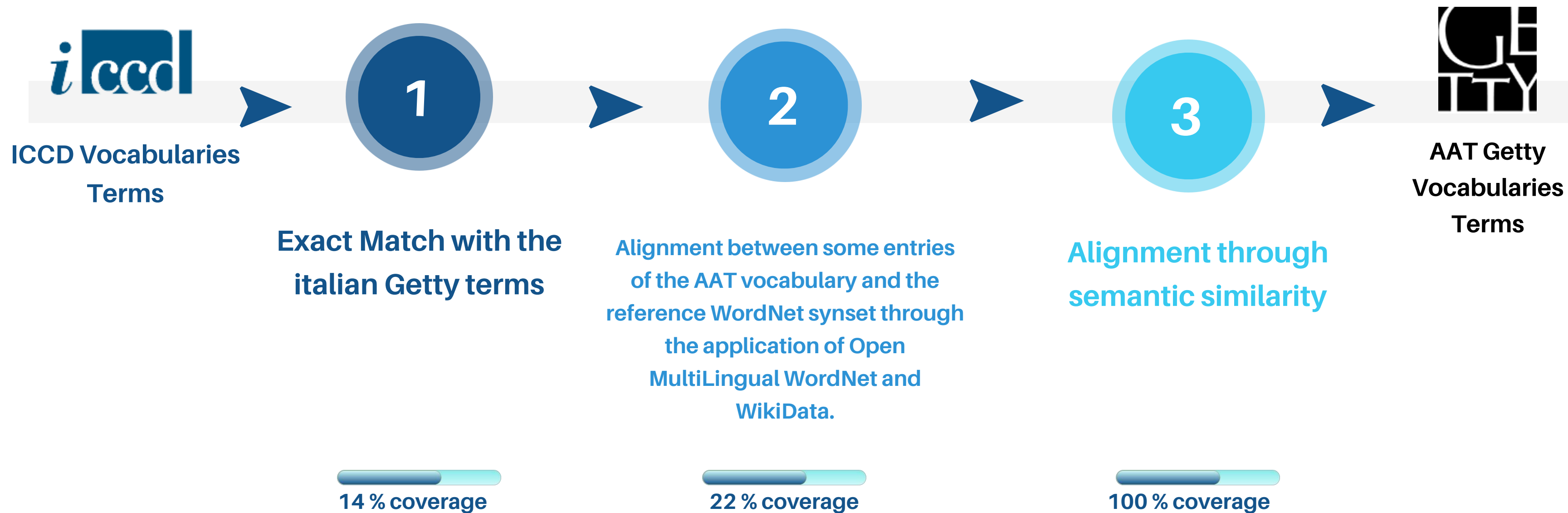
OUTCOME

Matching and alignment algorithm



METHODOLOGY

Linking with Getty's AAT



RESULTS ON VOCABULARY OF ARTWORKS

*by manual validation

Matching and alignment algorithm

Step	Precision	Recall
Exact match and prefix	0,994	0,780
Lemmatization	0,993	0,807
Similarity	0,989	0,962
Complete	0,982	0,993

Error analysis:

- Overuse of the similarity metric (e.g., the description "blocco" (block) paired with the term "bricco" (jug)).
- Incorrect input descriptions (i.e. improper use of prepositions, such as "di" (of) instead of "per" (for))
- Imprecision of the similarity metric.

Step	Precision	Coverage
Getty@it	81%	8%
WordNet	66%	2%
SBert	12%	10%
Total	21%	21%

Performance of the Getty's AAT vocabulary alignment with the Vocabulary of Artworks.

Step	Precision	Coverage
Getty@it	84%	3%
WordNet	67%	3%
SBert	17%	15%
Total	23%	23%

Performance of the AAT vocabulary alignment on the Vocabulary of Artworks' extension

Linking with Getty's AAT

RESULTS

Linking with Getty AAT considering the Vocabulary of Archaeological finds as a gold standard

Step	Precision w/o	Precision w	Recall
Getty@it	72%	63%	6%
Getty@it + WordNet	51%	48%	10%
Getty@it + WordNet + SBert	10%	4%	24%
Total	20%	10%	24%

- Precision w / o: considering as the gold standard only the aligned terms of the Vocabulary of Finds (67% of the vocabulary)
- Precision w: considering all terms of the Vocabulary of Finds as a gold standard.

FUTURE WORKS



IMPROVING MATCHING AND ALIGNMENT ALGORITHM

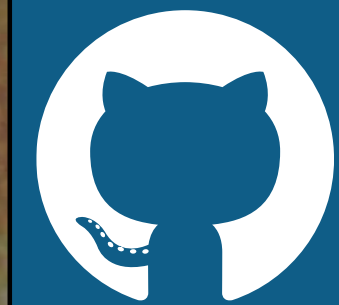
We intend to test different semantic similarity metrics for matching free text descriptions with vocabulary terms.



IMPROVING THE LINKING WITH GETTY'S AAT

We plan to enhance the alignment approaches between our controlled vocabularies and Getty's AAT Vocabulary by taking into consideration their hierarchical structure.

Thanks for your attention



<https://github.com/LuanaBulla/Controlled-Vocabularies-for-Cultural-Heritage>