

Clare Llewellyn | Andrew Shapland  
Aruna Bhaugeerutty | Tyler Bonnet

Reconciling Knowledge Graphs and Social Media from Newspaper Articles to Twitter

Kevin Page | Robert Sanderson | Kelly Davis  
Emmanuelle Delmas-Glass | Kayla Shipp

## Abstract

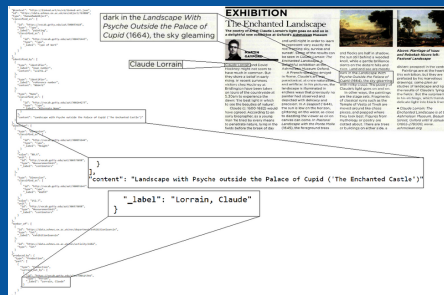
The co-exhibition of artworks has been a part of our cultural context for more than 250 years. Exhibitions, physical and digital, make art relevant and accessible to artists, scholars and the general public in a managed and directed environment. Such engagement influences emerging trends in style and the art market, but neither information on objects that were exhibited together nor their reception from the audience have been made available for art historical or digital humanities research. The Linked Art collaboration [1] has developed a modern Linked Open Usable Data (LOUD) [2] set of specifications to describe and publish art-related knowledge, including which objects were shown during which exhibitions. For this data to be truly connected, enabling scholars to seamlessly traverse institutional silos, descriptions of exhibition events and art objects must be 'reconciled': the descriptions of the same object should be connected by matching identifiers in different systems at different institutions. The Enriching Exhibition Scholarship (EES) project is developing and applying computational techniques to align socially-based textual and structured data, such as exhibition catalogues, newspapers and social media, making reconciliation easier and more effective. The alignment and enhancement records has required art history domain expertise, specific cultural heritage metadata knowledge and advanced computational research in text mining, analytics, machine learning and information retrieval.



## Matching Structured Data and Full Text

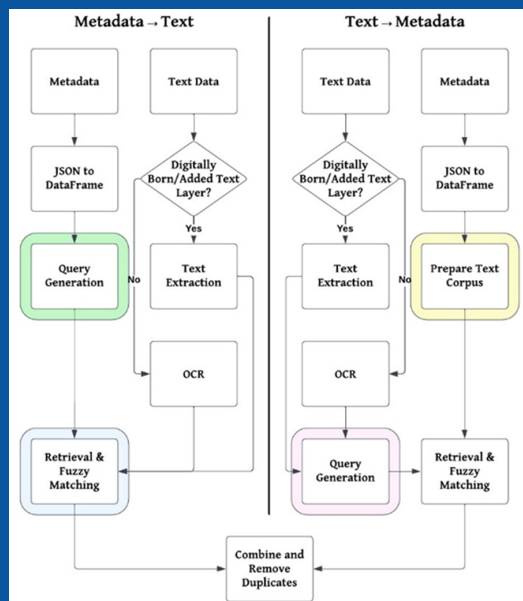
### Data

The project is working primarily with Linked Art metadata, newspaper articles, exhibition catalogues, and social media. The figure on the right shows an example of the type of matching we are doing. The figure illustrates a match between Linked Art metadata for an art object and a press cutting that is relevant to the object. This is considered a strong match as the object title, artist name, and an exhibition that featured the work are all present in the text. One of the aims of the project has been to develop a method of finding such matches automatically, reliably, and in a way that is highly scalable.



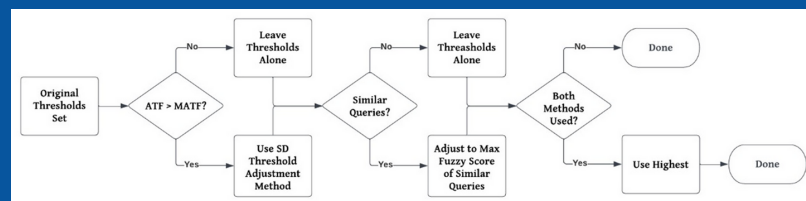
### NLP Pipeline

The natural language processing (NLP) pipeline developed for the project matches structured data and full text using information retrieval techniques in combination with fuzzy matching. It has two sides corresponding to the direction of matching: metadata to text where parts of metadata are extracted and used as queries to search the full text; and text to metadata where named entities are extracted from text and used as queries to search the metadata. The pipeline uses basic text extraction or OCR depending on whether the document has searchable text. In addition to providing an additional measurement of potential relevance, fuzzy matching allows the pipeline to handle OCR errors more effectively. The search function is set up using the information retrieval framework PyTerror [3] after running an experiment to determine which supported retrieval model performs best on a set of query relevance judgements made for the dataset. In addition to a retrieval score, a fuzzy matching score is calculated for each search result. For a search result to be considered a match, it must meet thresholds for both retrieval score and fuzzy matching score.



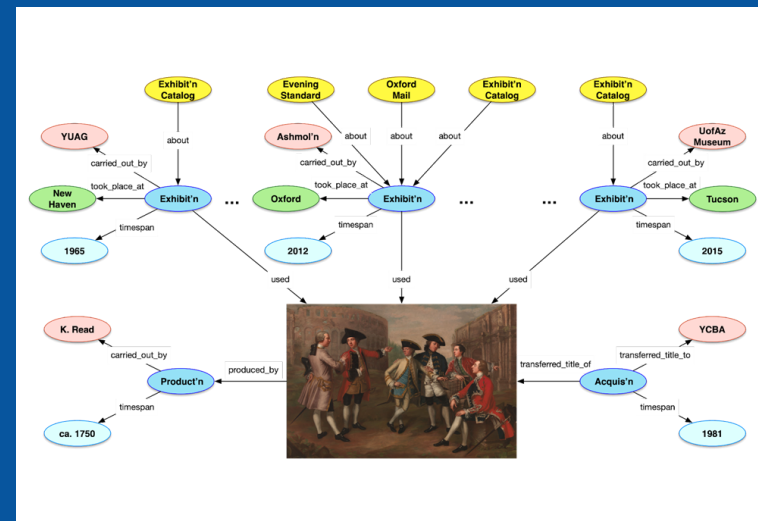
### Threshold Adjustment Algorithm

Retrieval is important for narrowing down the corpus to potential matches, but an additional process is needed to further separate the matches from the non-matches. This is done using a threshold adjustment algorithm (shown below), which is formed of two methods for adjusting retrieval and fuzzy matching score thresholds. First, for those query-text pairs that have an Average Term Frequency (ATF) greater than the Mean Average Term Frequency (MATF), thresholds are increased by the number of standard deviations difference between the ATF and MATF. Second, fuzzy matching score thresholds are raised to the maximum fuzzy matching score among those scores calculated for the given query and all other queries.

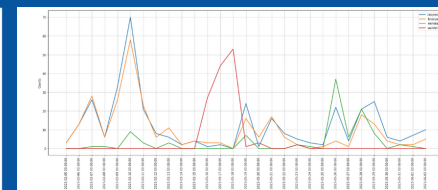
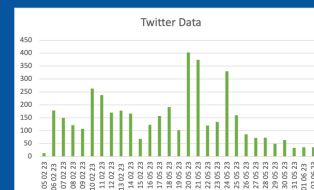


## Exhibitions

The Yale Center for British Art lent the Ashmolean Museum at Oxford 19 artworks for the exhibition The English Prize: The Capture of the Westmorland in 1712, including the painting British Gentlemen in Rome, by Katharine Read from circa 1750. That object was also lent to the Yale University Art Gallery for an exhibition in 1965 and most recently to the University of Arizona Museum of Art for their 2015 exhibition Rome: The Legacy of an Eternal City. Newspaper articles in The Evening Standard and The Oxford Mail describe the exhibition and reference other works that were present, including Batoni Pompeo's painting of Francis Basset, in the Museo del Prado. By mining these social texts (in yellow in the figure on the right), and connecting with the Linked Art structured data from across institutions, we can bring together both quantitative and qualitative information about objects presented and audience responses to them.



## Social Media Engagement With Exhibition



Visitors to the Labyrinth: Knossos, Myth & Reality Exhibition at the Ashmolean Museum were invited to engage via social media by using various hashtags such as #LostintheLabyrinth and #LabyrinthLabel. With the second hashtag, visitors were asked to write their own labels for objects on display. So far, 4194 Tweets have been gathered. There has also been engagement on Instagram.

### Write Your Own Label

Some visitors took us up on our invitation to write a label on Twitter (first right) and others decided to share their experience generally (second right).



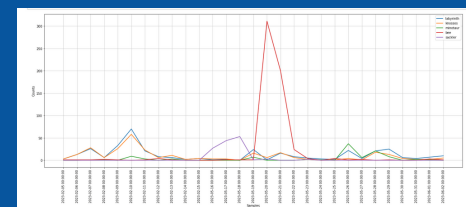
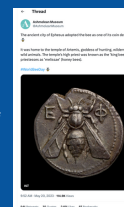
### Instagram

The project will also be looking into Instagram where we are also seeing engagement including visitors writing their own labels like the one below about the loom weight and weaving in ancient Greece.



### The Bee Tweet

One object that got special attention was a coin with a bee design from the ancient city of Ephesus. The second figure on the right shows a spike in engagement on World Bee Day.



## References

- [1] Page, Kevin; Delmas-Glass, Emmanuelle; Beaudet, David; Norling, Samantha; Rother, Lynn; Hänsli, Thomas. Linked Art: Networking Digital Collections and Scholarship. Digital Humanities 2020 (DH2020) Book of Abstracts, pp. 504-509.
- [2] Sanderson, Rob. Standards and Communities: Connected People, Consistent Data, Usable Applications (Keynote). 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL), pp. xxviii-xxix.
- [3] Macdonald, Craig, and Nicola Tonellotto. Declarative experimentation in information retrieval using PyTerror. In Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval, pp. 161-168. 2020.

