



## Image Retrieval in Digital Libraries

### A Pragmatic Approach

#### Jean-Philippe Moreux

Preservation dpt, Digitization service, Bibliothèque nationale de France, Paris, France.

[jean-philippe.moreux@bnf.fr](mailto:jean-philippe.moreux@bnf.fr)

#### Guillaume Chiron

L3i Lab, University of La Rochelle, France

[guillaume.chiron@univ-lr.fr](mailto:guillaume.chiron@univ-lr.fr)



Copyright © 2017 by JP Moreux G. Chiron. This work is made available under the terms of the Creative Commons Attribution 4.0 Unported License: <https://creativecommons.org/licenses/by/4.0/>

---

### Abstract:

While historically digital heritage libraries were first powered in image mode, they quickly took advantage of OCR technology to index printed collections and consequently improve the perimeter and performance of the information retrieval service offered to users. But the access to iconographic resources has not progressed in the same way, and the latter remain in the shadows: manual incomplete and heterogeneous indexation, data silos by iconographic genre. Today, however, it would be possible to make better use of these resources, especially by exploiting the enormous volumes of OCR produced during the last two decades, and thus valorize these engravings, drawings, photographs, maps, etc. for their own value but also as an attractive entry point into the collections, supporting the discovery and serenity from document to document and collection to collection. This article presents an ETL (extract-transform-load) approach to this need, that aims to: Identify and extract iconography wherever it may be found, in image collections but also in print (dailies, magazines, monographies); Transform, harmonize and enrich the descriptive metadata (in particular with automatic classification tools); Load it all into a web portal dedicated to iconographic research. The approach is pragmatically dual, since it involves leveraging existing digital resources and (virtually) on-the-shelf technologies.

**Keywords:** digital libraries; image retrieval; image classification; data mining; metadata; OCR; heritage documents

---