



HAL
open science

The French digital infrastructure for written heritage Biblissima+ and the challenges of musical heritage

Kévin Roger

► **To cite this version:**

Kévin Roger. The French digital infrastructure for written heritage Biblissima+ and the challenges of musical heritage. Medieval and Renaissance Music Conference 2023, Jul 2023, Munich (Allemagne), Germany. 2023, 10.5281/zenodo.8414821 . hal-04231647

HAL Id: hal-04231647

<https://hal.science/hal-04231647>

Submitted on 6 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The French digital infrastructure for written heritage Biblissima+ “The challenges of musical heritage”

Kévin Roger & David Fiala

Biblissima+

Biblissima+ is a web portal and a multi-site digital infrastructure devoted to the history of transmission of ancient texts. It is organized into 7 clusters bringing together researchers, curators and engineers around specific objects and/or technologies.

Cluster 2 - Study and searchability of data materials

Cluster 1 - Acquisition of interoperable source corpora

Cluster 3 - Artificial intelligence, pattern and handwriting recognition

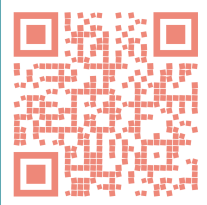
Cluster 4 - In-depth treatment of graphic systems

Cluster 5 - Editing sources in TEI

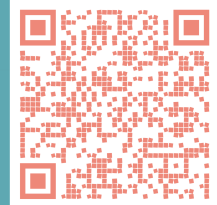
Cluster 7 - Text interoperability and analysis

Cluster 6 - The challenges of musical heritage

GitHub Repository



Zenodo Community



Cluster 6, led by D. Fiala and K. Roger (RicarcarLab/CESR), is dedicated to the digital scientific mediation of music books and scores. Its aim is to improve the quality and richness of musicological metadata and provide multimedia solutions for accessibility to digital data on early music.

Connecting music resources

While inventories of music sources abound on the Web, they are often not interoperable with the viewers of libraries and archives. Cluster 6 set itself the task of creating various technical gateways. A significant example is the DIAMMtoIIIF project, which aims to generate visualization files (IIIF manifests) connecting metadata from the Digital Image Archive of Medieval Music to the image files of the BnF (Gallica). To achieve this, Cluster 6 developed a Python program performing various tasks: in addition to querying the DIAMM API, it automatically generates the IIIF manifest. Once the IIIF manifest is opened in an online viewer like MIRADOR, the visualization of the source appears augmented with the various metadata of each work.

```
import json
import urllib.request
import re
from iiif_presentation_factory import ManifestFactory

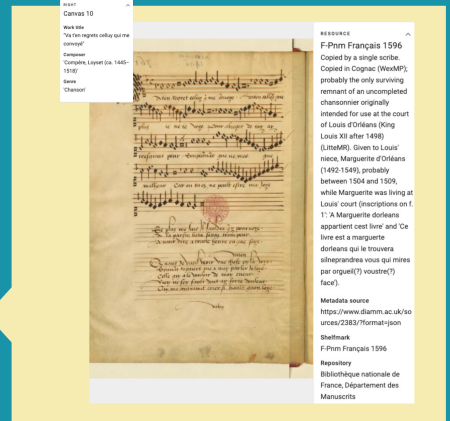
url = "https://www.diamm.ac.uk/sources/2387/ffm"
response = urllib.request.urlopen(url)
data = json.loads(response.read())

mf = ManifestFactory()
factory.set_base_uri("https://example.com/")
factory.set_base_image_uri("https://gallica.bnf.fr/iiif/image/info/version/2.0/")
factory.set_iiif_image_info_version("2.0")

mf = factory.manifest(data["display_name"])
mf.viewingDirection = "left-to-right"
mf.set_metadata({
    "Metadata source": data["url"],
    "id": data["display_name"],
    "Repository": data["archive"]["name"],
    "Date": data["date_statement"],
    "Type": data["source_type"],
    "Source type": data["source_type"],
    "Measurements": data["measurements"]
})
mf.description = data["notes"][0]["note"]
mf.viewingList = "pages"
d = str(data)

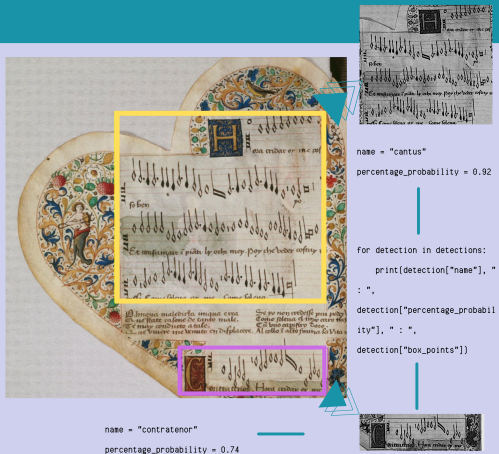
void_folios = 9 # voids the number of void folios
y = void_folios + 1

def get_title():
    titlelist = []
    titlepattern = re.compile("composition|volume|titlematch")
    titlematches = titlepattern.findall(d)
    for match in titlematches:
        titlelist.append(match)
```



AI model applied to early music

Cluster 6 has also started to develop artificial intelligence models for the automatic recognition of early musical sources (manuscripts and printed documents). The aim is, first, to recover detailed metadata generally ignored by conventional inventories (identification and precise location of musical parts on the page, notation used, etc.) but critical for the understanding of early music sources. In a second step, this information could be automatically integrated into IIIF manifests and processed for aligning various external resources such as music or text encodings, annotations or audio files. In conjunction with the BnF Data Lab, Cluster 6 intends to assemble a substantial dataset of musical sources from the 13th to the 16th centuries. Certain deep learning strategies have already been defined and executed in experimental versions of the model. Model training was developed in Python, using the ImageAI library as well as the YOLO recognition model.



Biblissima+ is funded by a French national grant, managed by the ANR under the Programme d'investissements d'avenir integrated into France 2030, bearing the reference ANR-21-ESRE-0005.

