Report on Preparing Europeana for IIIF involvement
Task Force

Appendix D: Implementation Examples
D.1: BnF use case on IIIF

Emmanuelle Bermès May 2017

Content

Appendix D.1: BnF use case on IIIF

1. Context

20 years ago, the national library of France (BnF) has created Gallica, a collective digital library. Today, Gallica displays online, free of charge, more than 4 million digitized documents from the BnF collections and from over 300 other partner institutions in France. The BnF has always been keen on disseminating its data, either bibliographic or of other kind, using standard formats and protocols: MARC/Z39.50, DC/OAI-PMH, and since 2011 RDF Linked data via the data.bnf.fr service.

Since 2010, the BnF has expressed interest in the development of IIIF and participated in the initial working groups that elaborated the standard (2011-2013). The first implementation of the image API within Gallica was achieved in 2014. The complete Gallica IIIF service (Image API + Manifests for the whole collection) was fully operational in 2016.

2. Objectives

Several objectives have guided the implementation of IIIF at BnF. They have also evolved over time.

2.1. Interoperability and a focus on niche collections: Biblissima

IIIF was first imagined by the scholarly community studying medieval manuscripts. It was seen as a promising way to develop an interoperable way to work remotely on digital collections and to gather virtual libraries of scattered documents.

Gallica being inclusive of a wide variety of digitised material (books, serials, newspapers, prints, photographs, objects, sounds, videos and of course, manuscripts both ancient and modern), its interface cannot be as specialised as would be needed to study specific types of documents such as medieval manuscripts using adequate tools. IIIF was thus foreseen as a mean to make Gallica's content available in more specialised interfaces, dedicated to specific usage by a niche of academic users.

The Biblissima project, in which the BnF was involved since its beginning, demonstrated what can be achieved by creating a series of prototypes. One of them demonstrates the reconstitution of a manuscript kept at the Châteauroux public library which illuminations were removed and are now held in the Prints and Photographs department of the BnF (see online prototype):

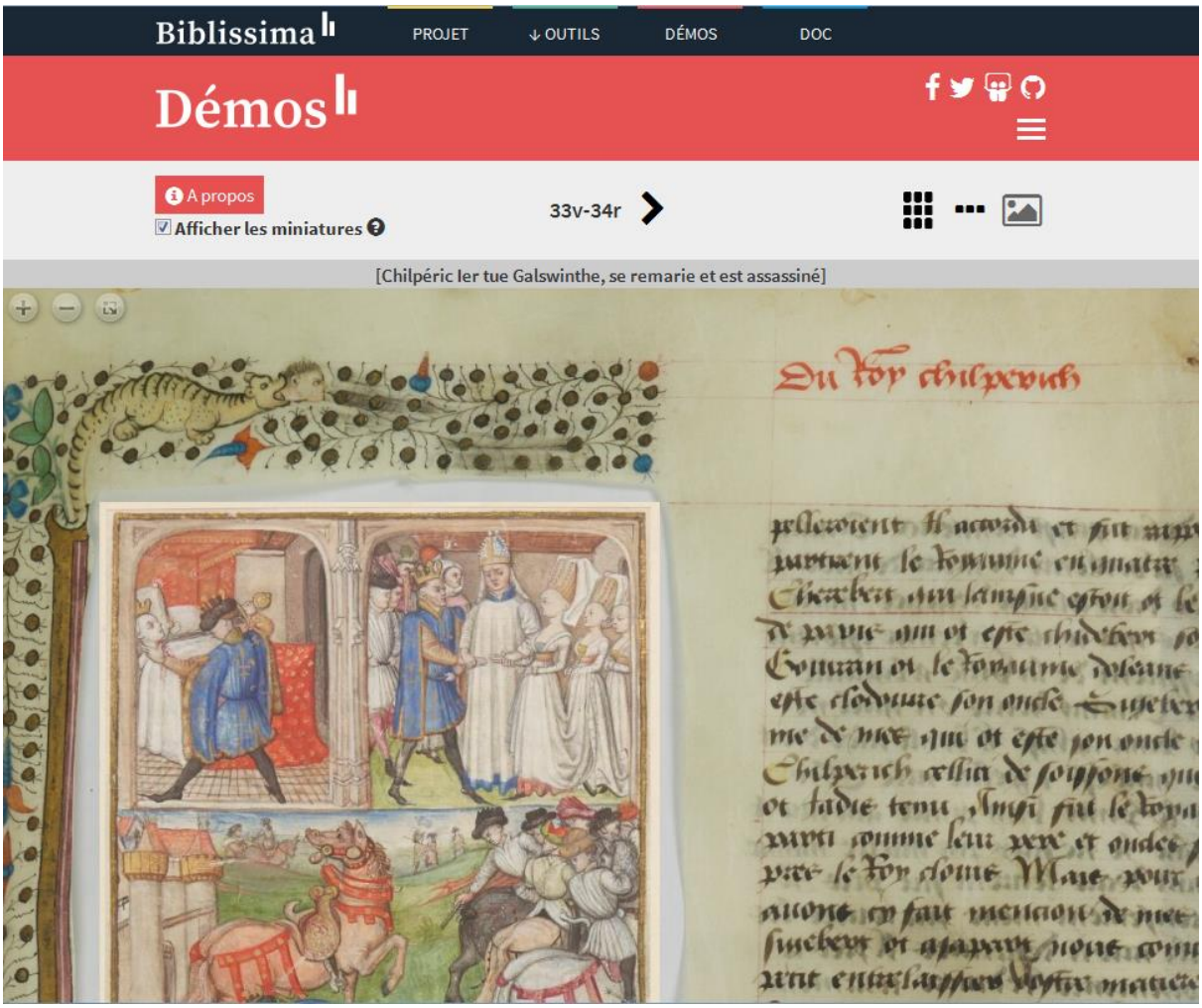

Illustration 1: online demo of the Châteauroux 5 MSS with its illuminations reinserted using IIIF.

For the purpose of Biblissima, a first instance of a IIIF Image Server for Gallica was developed in 2014, displaying only images from material that was relevant to Biblissima (mainly manuscripts).

2.2. Making the system modular: a IIIF-based zoom feature

The first Biblissima experiments with IIIF led the BnF to consider using the image standard inside the Gallica system in order to improve the high resolution zoom feature. This decision seemed natural in the context of the BnF's global strategy to make its architecture more modular and use internal APIs, as much as possible relying on standards (for instance, the search API for Gallica is based on SRU).

In order to achieve this goal, the main challenge was to set up a system that would convert all the master images produced in TIFF into JPEG2000. This is done on the fly, but a cache system was created in order to store the JPEG2000 version once it has been generated and thus provide quicker access for the end user.

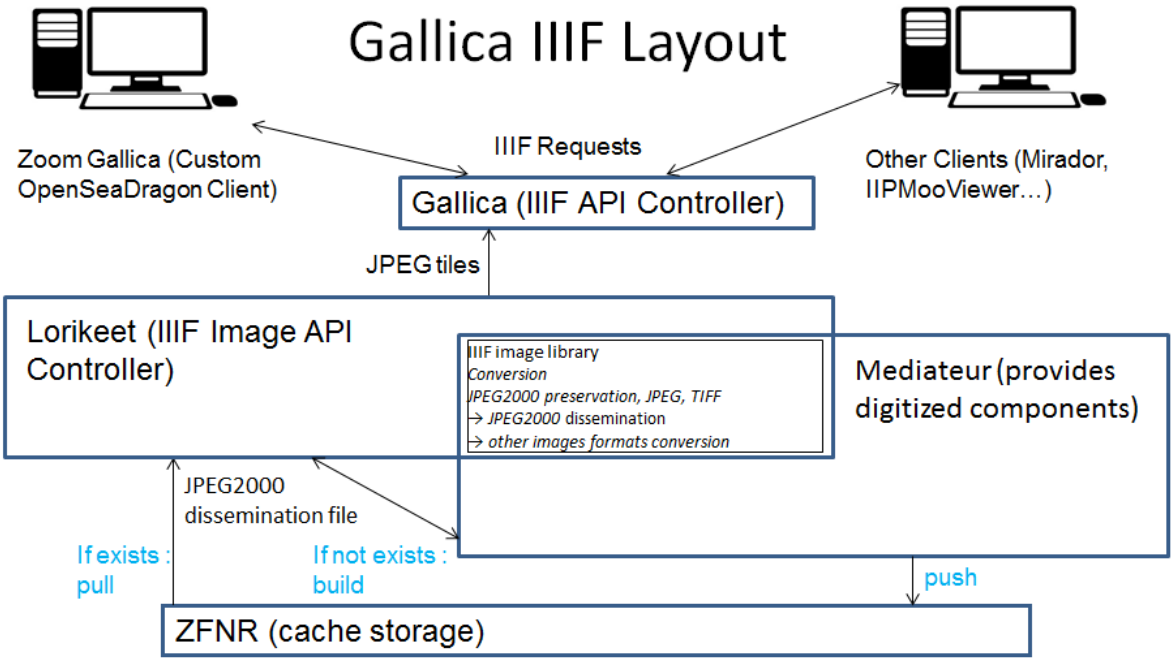

Illustration 2: architecture of Gallica's IIIF zoom service.

As shown on the diagram above, a web app called Lorikeet was developed to interpret the queries and deliver images in return. If the image is already available as JPEG2000 in the diffusion storage zone (ZFNR), it is sent directly to the Lorikeet app. If not, the server (called Mediateur) operates a conversion from TIFF to JPEG2000 and stores it in the storage zone.

In october 2015, this new architecture was ready. A new version of Gallica was published, where all the images were available through the IIIF Image API. The new zoom interface, based on a custom OpenSeaDragon client, was available to end users.

2.3. Increasing visibility: IIIF and Europeana

As soon as the IIIF image API was available for all of Gallica, it was seen as an opportunity to increase tremendously the visibility of Gallica's material across the web and to improve user experience.

The Europeana Portal was the first external website to implement IIIF as a way to access Gallica's content. Before, only the metadata was aggregated in Europeana and if the user wanted to access the content, he/she had to leave the Europeana portal and go to Gallica. This could be difficult to handle for the user, because he had to get used to a new design and different features each time he accessed content via a different provider. Moreover, he had to go back and forth from the portal to the provider in order to conduct his research.

IIIF makes it possible to integrate access to the content on the portal, including features such as turning pages or high resolution zoom, without having to transfer the actual digitised images from the BnF to Europeana and without decreasing the audience of Gallica.

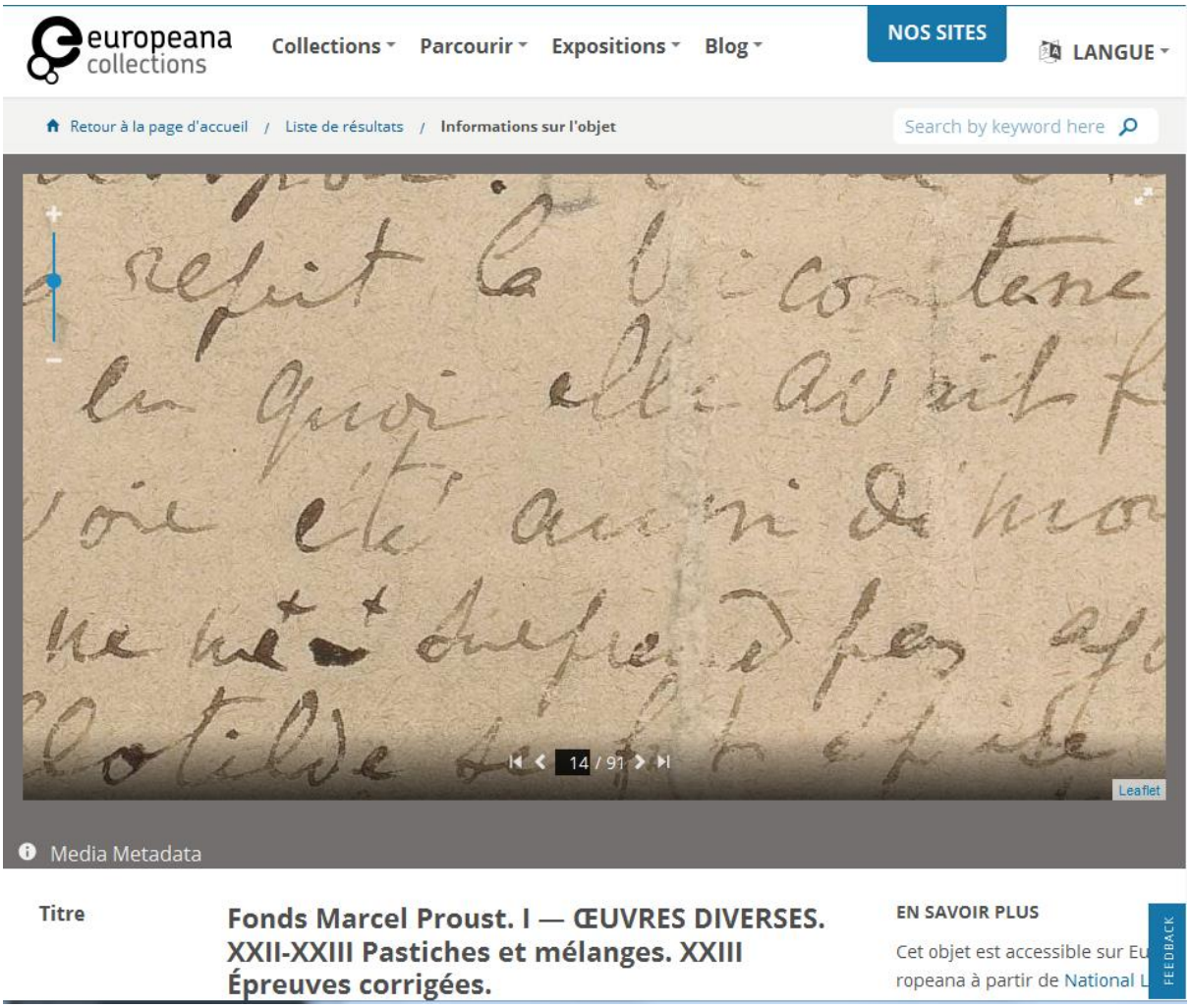

Illustration 3: zooming on a Proust Manuscript from the BnF within the Europeana Portal.

In order to make this new aggregation feature possible, the BnF had to implement not only the IIIF Image API but also the IIIF Presentation API, which provides metadata and information on the structure of document such as order of pages, etc. known as "Manifest". The Presentation API was available on every Gallica document in April 2016.

3. Perspectives and impact

IIIF as it has been developed until now at the BnF has opened new perspectives.

3.1. Shared collections

The Europeana implementation mentioned above has demonstrated the feasibility and interest of a new aggregation model: while digital libraries used to build shared portals based only on metadata, they can now add advanced display functions for the content itself.

The Polonsky project is a 3 year project funded by the Polonsky foundation to digitize 800 early medieval manuscripts (VIIIe – XIIIe c.) from the BnF and the British Library (BL) collections, 400 from each institution. The whole collection will be presented online in a IIIF-based viewer developed by the BnF.

For this project, the BnF has decided to adapt its shared infrastructure "Gallica Marque Blanche" or "Gallica Vanilla": a fully customizable digital library service designed to build a digital library with the partner's branding (in this case the project's branding) based on Gallica's infrastructure.

Within the Polonsky project, the BnF will develop a Gallica Vanilla website which will display digitised manuscripts from both institutions.

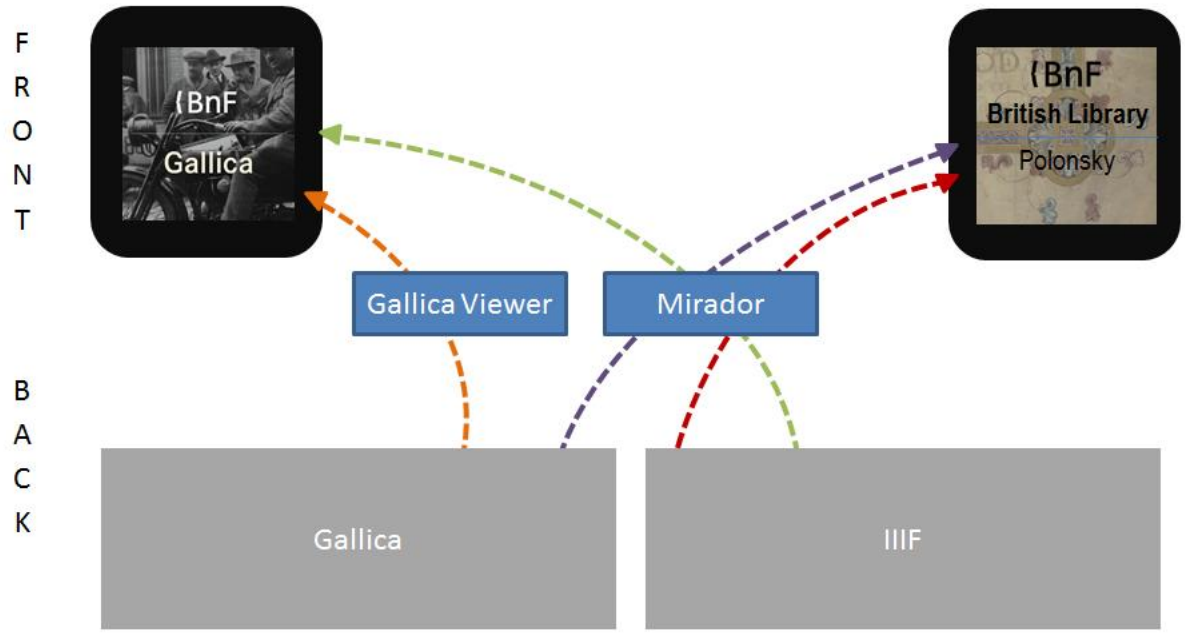

Illustration 4: how IIIF will improve Gallica's aggregation model.

The Gallica Vanilla infrastructure will evolve in order to make it possible to display images that are available remotely through a new viewer based on Mirador. This new feature will allow to display the BL's images in the Polonsky IIIF website without any transfer of digital files, but also to extend this feature to other partners of Gallica and Gallica Vanilla in the future.

Therefore, this improvement will later on benefit other Gallica contributors among the 300 partners that provide content to the digital library: currently, image files for Gallica's partners have to be uploaded to the BnF's servers before they can be displayed in Gallica or a Gallica Vanilla website. When the IIIF viewer is integrated, images from partners who have a IIIF server can be made available directly on Gallica without going through a cumbersome ingestion process.

3.2. Shared developments

Investigating the use of Mirador for the Polonsky project also opens perspectives of sharing the development effort with the wider community.

Mirador is a powerful IIIF-compliant viewer with a very complete set of image management features. However, features more focused on text (OCR display and search, etc.) would be required in order to cover all the tools currently offered by the Gallica viewer. As this work

in underway within the IIIF community, the BnF could contribute to this joint effort by developing on Mirador rather than continuing to build its own viewer, and this investment would also benefit the wider community.

Such shared developments on a common viewer are only possible because the underlying infrastructure relies on standard APIs, namely IIIF Image and Presentation APIs.

3.3. Unexpected use and reuse

Another impact of IIIF is the availability of the content to be reused by third parties that are not (yet) partners of the BnF. In 2016, the BnF organised its first hackathon: for this event the Gallica APIs, including IIIF, were for the first time documented and publicized (see the Hackathon 2016 Github). The availability of IIIF allowed for the development of new prototypes such as Gallica.lol, a meme generator for Gallica.

4. References

- Emmanuelle Bermes & Matthieu Bonicel, "IIIF and Biblissima". Presentation at IFLA 2016, Session 091: Worth a Thousand Words: *A Global Perspective on Image Description, Discovery, and Access*. See online.
- Biblissima Demo website and Beta website.
- Documentation of the BnF's IIIF API on the Hackathon BnF 2016 wiki. See on