

## “Discoverability” in the IIF digital ecosystem

Paola Manoni<sup>(a)</sup>

a) Biblioteca Apostolica Vaticana, <https://orcid.org/0000-0001-7802-2718>

**Contact:** Paola Manoni, [manoni@vatlib.it](mailto:manoni@vatlib.it)

**Received:** 25 August 2021; **Accepted:** 21 September 2021; **First Published:** 15 January 2022

### ABSTRACT

The IIF APIs have been used since 2012 by a community of research, national and state libraries, museums, companies and image repositories committed to providing access to image resources. The IIF technical groups have developed compelling tools for the display of more than a billion IIF-compatible images.

We can figure out that with hundreds of institutions participating worldwide, the possibilities, for instance, for IIF-based scholarship are growing so one question could be about the discovery of those images relevant to one's research interests in order to discover them for their consultation or, even more, for their reuse.

While IIF specifications discussion has focused on the machine-to-machine mechanisms of making IIF resources harvestable, we have yet to implement an end-to-end solution that demonstrates how discovery might be accomplished at scale and across a range of differing standards for metadata arising from libraries, archives, and museums.

### KEYWORDS

IIF; LAM; Discovery; Digital ecosystem.

## 1. Discoverability

The International Image Interoperability Framework<sup>1</sup> as an interoperability protocol for image resources held in libraries, archives, museums, has produced over a billion IIIF-compliant images. This paper will focus on how this vast production is actually changing not only the use of digital objects online in the context of tools at the convenience of digital humanities, for instance with the well known abilities of IIIF viewers, such as Mirador<sup>2</sup>, but also the concept of discoverability of the knowledge objects now available via IIIF.

Discoverability is the quality of being able to be discovered or found and in relation to online content, it is the quality of being easy to find via a search engine, within an application, or on a website.

If we focus on the discoverability in the IIIF context we can refer to two main aspects:

- Which are the requirements that make a web platform a discoverable digital library service in the light of IIIF;
- How is it possible to discover IIIF-compliant content through current web platforms.

A first look about the context of the two issues, is concerning the non-trivial meaning of LAM data in the universe of a single domain, for an evaluation of their impact on the discoverability of IIIF objects. We thus consider the abstraction of LAM data produced within the digital ecosystem starting from the traditional statements related to the classes of:

- *Structured data* – In the LAM domain they include bibliographies, catalogs, indexing and abstracting databases, authority files. Structured data is generally stored in databases where all key / value pairs have clear identifiers and relationships and follow an explicit data model
- *Semi-structured data* - they are the unstructured sections within metadata descriptions as well as any unstructured portions of structured datasets.
- *Unstructured data* – they are the typical “everything else” pertaining to documents and other information-bearing objects in all kinds of formats.” (Zeng 2019).

We consider the typical elements of the IIIF Presentation API, keeping in mind that IIIF Presentation API provides:

- A model for describing digital representations of objects: just the metadata chosen in a completely arbitrary way in order to offer a remote viewing experience.
- A format for software - viewing tools, annotation clients, web sites - to consume and render the objects and any other associated content in the form of annotations.

This does not mean that descriptive metadata has no place in a digital object provided by the Presentation API. In fact, it is important that the object is linked to its description and to all the information relating to it. The presentation API provides this human readable information, so that viewers can interpret the important contextual information to end-users.

---

<sup>1</sup> Cfr. *International Image Interoperability Protocol* < <https://iiif.io/>>. Accessed April 15, 2021.

<sup>2</sup> Cfr. <https://projectmirador.org/>. Accessed April 15, 2021. Mirador is a fully IIIF-compatible tool capable of interpreting IIIF APIs. Mirador is an open source image, Javascript and HTML5 viewer that delivers high resolution images in a workspace that enables image annotation and comparison of images from repositories dispersed around the world, starting from compatibility with Image API that specifies a web service returning an image in response to a standard HTTP or HTTPS request.

The information pertaining to Presentation API is the IIIF manifest of the digital object represented as a “thing” and enriched with the complex knowledge data related to it.

A so-called IIIF manifest contains a descriptive section of the digital object but the specifications do not define any rules relating to metadata. In other words, we can say that IIIF requirements are completely agnostic as to which descriptive metadata to apply as well as to which image formats. The galaxy of data pertaining to the “thing” represented in the manifest is completely scalable and referable to the different meanings of LAM data (structured, semi-structured, and unstructured) we mentioned.

Moreover, as for the Presentation API, the meaning of any accompanying descriptive metadata for display in a viewer is not taken into any account. The purpose of this API is the representation of the content of the work – for example the pages of the book, the painting – or the link where users can get information about the meaning of the content of the work.

The objective of the IIIF Presentation API is to provide the information necessary to allow a rich, online viewing environment for primarily image-based objects to be presented to a human user, likely in conjunction with the IIIF Image API. In other words, the IIIF Presentation API gives us a specification for “presenting” a digital object and the data describing it in order to view, annotate it, or compare it with other objects. A IIIF client can also display any accompanying metadata included as pairs of labels and values within the manifest. But it needs no definition or scheme for what that metadata means. It is *outside of the scope* of the Presentation API (Crane 2017).

The user can view important semantic metadata, but the scope of the Presentation API is just to leverage that text. In the Presentation API, the semantic meaning is *elsewhere* because it is not belonging to its the specifications. An API client should simply render them.

In a nutshell, a manifest is what a IIIF viewer loads to display the object. A manifest could be used to represent the object within a web service as well as it could be used to add annotations to the represented object or even to be aggregated within a new manifest thus realizing its reuse.

The structure of the manifest also includes the concepts of sequence, which is of fundamental importance for aggregated resources (e.g. books, manuscripts and archive materials composed of page, leaf, folio or sheet) and canvas.

Each view of the object, for example each page is represented by a canvas. A Manifest contains one or more **Sequences of Canvases**. But a canvas is not the same as an image. “The canvas is an abstraction, a virtual container for content” (Crane 2017).

A Canvas is the digital surrogate for a physical page which should be rendered to the user. Each Canvas has a rectangular aspect ratio, and is positioned such that the top left hand corner of the Canvas corresponds to the top left hand corner of a rectangular bounding box around the page, and similarly for the bottom right hand corners. The identifier for the Canvas is not an identifier for the physical page, it identifies the digital representation of it.<sup>3</sup>

The canvas is a kind of conceptual extra layer in which an object is included.

The canvas keeps the content separate from the conceptual model of the page of the book, paint-

---

<sup>3</sup> Cfr. *Shared Canvas Data Model* <<http://iiif-io.us-east-1.elasticbeanstalk.com/model/shared-canvas/>>. Accessed April 15, 2021.

ing or archival unit. The content, we are referring to, could be blocks of text, videos, links to other resources, and it is exactly mapped on the canvas. By including a canvas in a manifest, you provide a space on which users and scholars can annotate the content.

All association of content with a canvas is done by **annotation**. The IIIF Presentation API is built on the W3C Web Annotation Data Model<sup>4</sup>.

Annotations associate content resources with Canvases. The same mechanism is used for the visible and/or audible resources as is used for transcriptions, commentary, tags and other content. This provides a single, unified method for aligning information, and provides a standards-based framework for distinguishing parts of resources and parts of Canvases.<sup>5</sup>

The canvas establishes a stage in which the simplest case – one image per canvas – is straightforward, but more complex cases, more complex and interesting associations of content, can be managed.

The latest specification of the IIIF, still in beta version, is the IIIF Content State API<sup>6</sup> which demonstrates another purpose for a representation by sharing a content to be represented on a canvas.

In its scope there are two examples:

- A user follows a link from a search result, which opens a IIIF viewer. The viewer focuses on the relevant part of the object, such as a particular line of text that contains the searched-for term.
- A user opens several IIIF Manifests to compare paintings, then wishes to share this set of views with a colleague.

These are examples of sharing a resource, or better, a *particular view* of a resource. Other examples include bookmarks, citations, playlists and deep linking into digital objects.

The objective of the IIIF Content State API is to provide a standardized format for sharing of a particular view of one or more IIIF Presentation API resources, such as a Collection, a Manifest, or a particular part of a Manifest.

Content State API is **how we can point at things in IIIF** and this demonstrates how the concept of digital resource and its reuse expand to include new ways of knowing the resources and new ways of citing them. In fact, it basically means dereferencing URIs of annotations whose motivation such as *content state* will be included in a manifest.

Content State is a way for humans to share bookmarks, and it's also a way for search results to point at the exact part of a digital object that they match (Crane 2021).

We can argue at this regard that the semantic enrichment process pertaining to the IIIF's vision of LAM objects and data reflects the broader general transformation from document-centric to entity-centric knowledge modeling due to the many relations for each canvas.

---

<sup>4</sup> Cfr. *Web Annotation Data Model*. Accessed April 15, 2021. <https://www.w3.org/TR/annotation-model/>. The Model does not prescribe a transport protocol for creating, managing and retrieving annotations. Instead, it describes a resource oriented structure and serialization of that structure that could be carried over many different protocols.

<sup>5</sup> Cfr. *IIIF Presentation API 3.0* Accessed April 15, 2021. <https://iiif.io/api/presentation/3.0/>.

<sup>6</sup> Cfr. *IIIF Content State API 0.3* Accessed April 15, 2021. <https://iiif.io/api/content-state/0.3/>.

Let us now go back to consider the discoverability of the IIIF in the light of Presentation API and the first question

- Which are the requirements that make a web platform a discoverable digital library service in the light of the IIIF;

We may focus on this by considering the use case of the Vatican Library as an example.

## 2. The use case of the Vatican Library

DVL (the DigiVatLib, <<https://digi.vatlib.it>>) is a digital library service. It provides free access to the Vatican Library's digitized collections: manuscripts, incunabula, archival materials and inventories as well as graphic materials, coins and medals, printed materials. It is fully based on the International Image Interoperability Framework technology, making digital materials easily accessible and usable.

- The viewer is able to zoom, browse and 'turn pages' of JPEG2000 images as well as allow scholars to compare digital objects from different IIIF repositories of other digital libraries.
- Descriptions and bibliographic references from the online catalogues are indexed and linked to digital materials.
- Each object is equipped with URIs for the discovery of IIIF manifests.
- The guided navigation ('faceted search') leverages metadata elements for narrowing or refining queries.

The Library has promoted a *new* perspective to the study of manuscripts by means of web communication and IIIF.

To meet this challenge the Library has implemented a project to enrich the digital delivery of these materials by annotating some exemplary manuscripts with scholarly analysis.

The use case of annotations in IIIF was a three-year Mellon-funded project, held between 2016 and 2019, in conjunction with Stanford University Libraries, which produced over 26,000 annotations for a selection of manuscripts chosen in the context of thematic pathways. In this platform (available at: <<https://spotlight.vatlib.it>>) the content of all the annotations is indexed along with the metadata, thus constituting a semantically enriched system that allows scholars to query an integrated search of all the available contents of a resource.

The project aimed to demonstrate, among the advantages of the IIIF for manuscripts, how the annotation level is a fundamental innovation for the study of contents: transcriptions, comments, comparative analysis of texts and images.

Thanks to the funds received, the Library has implemented a workflow using Mirador with scholarly analysis in order to tell scholarly narratives.

The Vatican Library has intended to engage the visitors to its website on the possibilities for using annotated manuscripts in IIIF, according to specific themes, by providing tools for discovering and comparing digital materials.

The deep analysis of contents of manuscripts entails the understanding of the "pre-print" world in which the manuscript is born. This implies a knowledge pertaining to the history of the man-

uscript, its origin, provenance as well as other circumstances of the production of a manuscript; identifications of dates, scribes, artists; discussions about the intellectual content and descriptive discussion on paleographic matters.

In its essential lines, a thematic pathway is composed by three different kinds of information:

- A general description (introduction, historical information, etc) of the chosen theme, it represents the “Story”;
- Descriptive and structural metadata and a curatorial narratives for each manuscript;
- Annotations, comments, in-depth analysis about detailed parts of a manuscript (e.g. texts, comments, illuminations, etc.) and transcriptions of units of information.

## The four thematic pathways

1. The first one is about *Courses in Paleography (Greek and Latin, from antiquity to the Renaissance)*

The rich collection of manuscripts preserved in the Library makes it possible to follow the evolution of the Greek and Latin scripts all the way from antiquity to the Renaissance.

The availability of on-line images of manuscripts, together with the possibilities offered by the IIF APIs, allows a complete transformation of teaching practice in this field.

For each of the sections (Greek and Latin) of this thematic path, a set of complete digitized manuscripts, chosen to illustrate the phases in the development of the script from the fourth to the sixteenth century, is provided. From each manuscript, chosen pages with a paleographical and codicological description and a diplomatic transcription is also made available.

2. The second one is about *The evolution and transmission of texts of specific works: Latin Classics*

The Vatican Library owns one of the most important collections of manuscripts with texts by Classical Latin authors, many of them richly illustrated.

The aim of this pathways is to describe 81 manuscripts directly from the original codices: metadata and annotations pertaining to the study of texts and illuminations have been provided. The work throws light not only on the illustrations of the texts but especially on the relationship between text, illuminations, comment and the gloss.

The importance of this project lies in the remarkable variety of typologies of the Classical world.

3. The third one is about *Vatican Palimpsests: Digital Recovery of Erased Identities*

The Vatican Library has identified more than 380 manuscripts in its own collections, which include palimpsests, erased and then recycled parchment folios. This pathway intends to present this rich and scarcely explored material to the public by making an in-depth archaeological research on the palimpsests of twenty-four select manuscripts and recover their lost identities with the help of IIF technology.

Making accessible hardly legible images to the public is a challenging task because the



actual method of publication has been designed to typical objects. By the pathway, digital reconstruction makes four palimpsests accessible both by their upper and lower scripts, a condition which the actual conservation of these manuscripts and the normal method of publication do not allow.

Erased texts are often very old and significant witnesses of a lost past but they are difficult to access for the naked eye. They need an expert interpreter and highly special photographic and post-processing technologies and especially the flexibility of presentation offered by IIF APIs which can turn erased texts more accessible online than in their physical existence.

4. The last one is about The humanist prince's library: Federico da Montefeltro and his manuscripts

The library of Federico da Montefeltro, Duke of Urbino (since 1474), is known as a typical humanist collection.

The collection was outstanding not only for its substance (the amount of volumes as well as the quality, in relation with other libraries of that age), but for the value of each manuscript partly acquired from antique market, many commissioned by Federico and realized by refined copyists and greater artists of that time. The manuscripts were produced in two main locations: Florence and Urbino.

In the first years, Federico preferred to buy or order manuscripts in Florence (both in writing and in illumination), later he preferred Ferrara or Padoan artists and scribes active in Urbino.

This pathway points out the characteristics of the two schools, very different in style, and the most important artists (half of the chosen manuscripts is representative of the Florentine school while the other half of the Ferrara and Padoan schools).

### 3. IIF Discovery for Humans Community Group

IIF enables the creation of rich digital collections that bring together content distributed among cultural heritage institutions. With image viewers, one is able to analyze works held in physically different locations side-by-side or overlaid within a web browser. However, in order to take advantage of the research tools afforded by IIF, a user must be able to find IIF resources.

Interoperable objects are of no use if one cannot find them, particularly if relevant objects reside in servers in many different institutions. Discovery in this case means human searching, browsing and finding of IIF resources across institutions. To be successful, IIF discovery must be user-focused and meet defined users' concrete needs.

To meet these needs, the IIF Discovery for Humans Community Group was recently organized. This group aims to go beyond specification work to promote implementations that enroll experts in research, content, user experience, metadata, and various technologies. In order to advance discovery in the LAM space, this group will foster user-focused approaches enabling the targeted discovery, spanning institutional and domain silos, of IIF resources.

These aims are different from and complementary to the approach of the IIF Discovery Tech-

nical Specification Group, which is chiefly concerned about providing the technical means for locating and finding updates about IIIF resources, as a prerequisite for harvesting and indexing metadata for searching within and across these institutional collections.

If we focus again on the two questions arisen in this paper about:

- Which are the requirements that make a web platform a discoverable digital library service in the light of the IIIF;
- How is it possible to discover IIIF-compliant content through current web platforms.

We may say that both are of fundamental interest to this group and they are closely related to the purposes of the initiatives conducted by the Group, aimed at:

- Gather problem statements and use cases to understand needs for user-focused discovery of IIIF resources
- Develop specifications for metadata attributes and crosswalks to enable discovery of LAM IIIF content across institutions and domains
- Create and maintain a list of metadata profiles in use by IIIF-supporting institutions to promote consistency in semantic description and its consumption
- Frame small-scale experiments that work towards live discovery implementations
- Provide a venue for demonstrations of applicable discovery applications and technologies
- Maintain a registry of existing discovery efforts
- Build on and amplify the ongoing work of the Discovery Technical Specifications Group and the IIIF Technical Community
- Communicate and disseminate the work of the group to the larger IIIF community, as well as allied professional communities.

One of the recent activities of this group was to **collect a list of discovery features** in order to:

- Provide examples of features as implemented
- Extract a comprehensive list of discovery features
- Develop a feature typology
- Identify IIIF-specific discovery affordances
- Build a feature checklist for self-evaluation
- Inform development of other discovery platforms
- Showcase exemplary feature implementations

First of all this task has provided a collated list of discovery features “in the wild,” to better understand the current landscape of IIIF resource discovery. It was useful as a basis for compiling a wide-ranging list of possible discovery features. This was further condensed and organized to derive broader categories for these features, and to identify which features were specifically tied to IIIF rather than associated discovery more generally. From this analysis we were able to get a sense of which features are broadly implemented and which are rarer. Based on this work we developed a feature checklist that may be used to identify which core discovery features are and are not available on a given site.

Metric for discovery features was the first important milestone planning the group’s commitments as an important first step to face the second question: How is it possible to discover IIIF-compliant content through current web platforms, the thread underlying this paper.



## References

Crane, Tom. 2017. *An Introduction to IIIF*. Digirati. Accessed April, 15 2021. <http://resources.digirati.com/iiif/an-introduction-to-iiif/>.

Crane, Tom. 2021. *What is IIIF Content State?* Accessed April 15 2021. <https://tom-crane.medium.com/what-is-iiif-content-state-dd15a543939f>.

Manoni, Paola. 2020. "L'adozione del IIIF nell'ecosistema digitale della Biblioteca Apostolica Vaticana." *DigItalia* 2: 96-105.

Manoni, Paola - Ponzi, Eva. 2020. "Thematic Pathways on the Web: IIIF Annotations of Manuscripts from the Vatican Collections: il "Progetto Mellon" della Biblioteca Vaticana". *Rivista di Storia della Miniatura* 24: 211-216.

Salarelli, Alberto. 2017. "International Image Interoperability Framework (IIIF): a panoramic view". *JLIS* 8, no. 1. Accessed April,15 2021. doi: 10.4403/jlis.it-12090.

Zeng, Marcia Lei. 2019. "Semantic enrichment for enhancing LAM data and supporting digital humanities: Review article" *El profesional de la información* 28, no 1.