

October 2017

# Life Story: Creating Metadata for the Portrait File

Zoe Waldron

*The New York Public Library*, [zoewaldron@nypl.org](mailto:zoewaldron@nypl.org)

Follow this and additional works at: <http://online.vraweb.org/vrab>

---

### Recommended Citation

Waldron, Zoe (2017) "Life Story: Creating Metadata for the Portrait File," *VRA Bulletin*:Vol. 44: Iss. 1, Article 4.  
Available at: <http://online.vraweb.org/vrab/vol44/iss1/4>

This Feature Articles is brought to you for free and open access by VRA Online. It has been accepted for inclusion in VRA Bulletin by an authorized editor of VRA Online.

---

## Life Story: Creating Metadata for the Portrait File

### **Abstract**

The Print Collection Portrait File encompasses 71,557 portrait and biographical images of notable figures in American, British, and European history from ancient times to early 20th century, and is the largest digital image collection at the New York Public Library. The metadata creation practice for the Portrait File evolved in its approach over its long digitization history that spanned more than ten years, many metadata creators, one database migration, and the formation in 2013 of NYPL's Metadata Services Unit, which oversees metadata creation across the institution. This article demonstrates that metadata creation is an ongoing process, and explores enhancements that might be made to the metadata for Portrait File items in the future.

### **Keywords**

New York Public Library, Digital Collections, Miriam and Ira D. Wallach Division of Art Prints and Photographs, Images, Clippings, Metadata Creation, Data Enhancement, Data Quality, Archival Processing, Portraits, Biographical Images, Discovery, Metadata Best Practices, Metadata Standards, Metadata Practice, Digitization.

### **Author Bio & Acknowledgements**

Zoe Waldron is a Metadata Specialist in the Metadata Services Unit in the New York Public Library. She holds an MLS from the Graduate School of Library and Information Studies at Queens College and an MA in art history from Hunter College, both in New York City. She wishes to acknowledge her colleagues at NYPL past and present, and thank Griselda Steiner and Barbara Elam for their editorial contributions to this article.

## **Life Story: Creating Metadata for the Portrait File**

### **Introduction**

Compiled by librarians at New York Public Library (NYPL) in the early 20th century to meet the needs of picture researchers, the Print Collection Portrait File collection includes over 71,500 images of outstanding figures, mainly in American, British, and European history from ancient times through the time of compilation. Digitization of this historic collection of images of notable people began before 2005 and was completed in 2015. The collection is available for research and image download on NYPL's Digital Collections website.

What follows is an examination of the descriptive metadata creation practices for the Portrait File, which evolved over its long digitization history. I'll also discuss enhancements that might be made to improve the data quality of Portrait File items, and lastly, explore some next steps in working with the collection's metadata. This article is derived from a presentation I gave at the Visual Resources Association Conference in Louisville, Kentucky in March 2017. For ease of reading, I've capitalized the names of data elements discussed in this article.

### **Overview of the Collection**

NYPL, perhaps unusual for a public library, maintains museum-level collections overseen by curators in the various curatorial divisions of its research centers. Owned by the Print Collection of NYPL's Miriam and Ira D. Wallach Division of Art Prints and Photographs, the Portrait File is located in the Stephen A. Schwarzman Building's central research library on Fifth Avenue in Midtown Manhattan. The collection is stored in 239 approximately 17 x 11 ½ x 3 in. document boxes, in the balcony stacks and in five oversize boxes in the downstairs stacks in the staff workroom, adjacent to the Print and Photography Collections' reading room. As a Print Collection staff member noted: "[The Portrait File] shows us how we represented famous people before the digital age."

Collection items include original woodcuts, engravings, etchings, lithographs, and some photographic prints. However, the images are mainly clippings: pictures, prints, and illustrated articles, both photographic and non-photographic, taken from books, periodicals, and newspapers in English, German, French, Spanish, Italian, and other languages. While items date primarily from the mid-19th to early 20th century, the collection contains older materials as well, from the 18th century and earlier. Thus, the Portrait File documents the techniques, style, and distribution of printed portrait imagery in the United States and Europe.

The collection contains a wide array of biographical images related to its approximately 35,000 subjects. These include portraits of all types from bust to full-length; sculpture; coin and medallion reliefs; scenes from life; political caricatures and cartoons; family, friends and associates; homes and haunts; manuscripts, artwork, letters, and possessions; deaths, death masks, funerals, and burials; and memorials and commemorations.

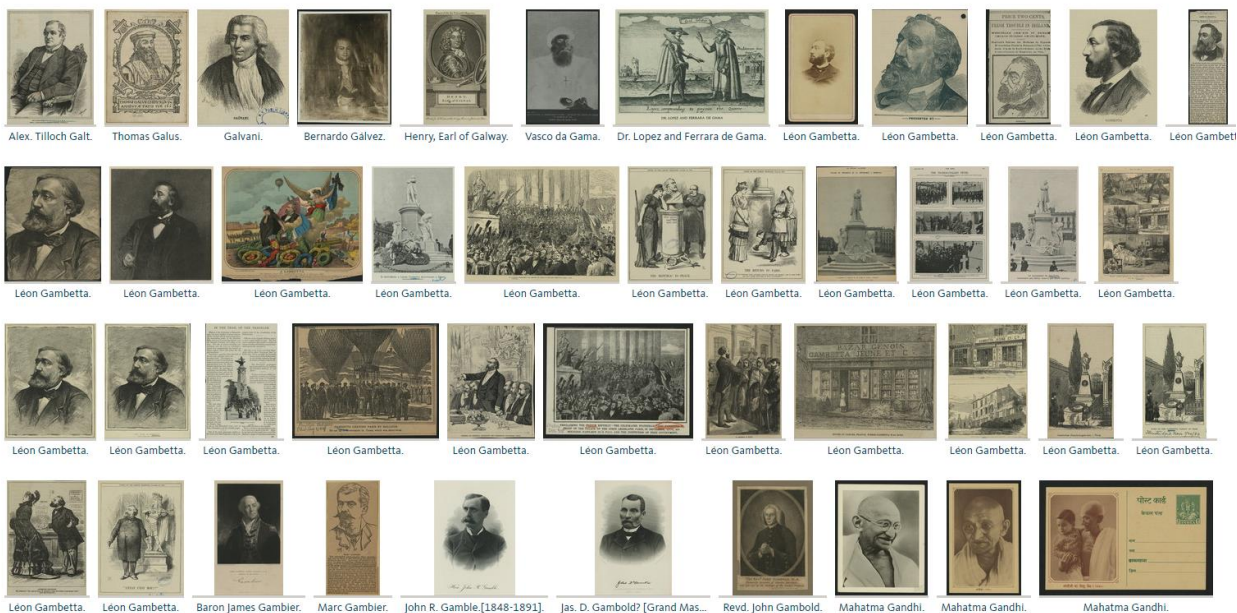


Figure 1. Screenshot from NYPL’s Digital Collections website showing the variety of subjects and biographical image types in the Portrait File collection.

Through its imagery, the Portrait File tells the stories of activists, artists, authors, the clergy, explorers, freedmen, mathematicians, monarchs, military leaders, the nobility, presidents, poets, musicians, scientists, slaves, soldiers, and many other outstanding people. The collection provides strong coverage of the British and European monarchy, nobility, and military leaders, and includes American presidents and politicians, the personalities of New York City politics, as well as the luminaries of arts, culture, and science in North America and abroad. In telling the stories of its many subjects, the Portrait File covers major milestones of European and United States history, giving researchers myriad biographical, historical, and pictorial currents to follow.

While collection subjects are largely white and male, women, frequently as holders of hereditary titles, as authors, or as spouses or family members of other portrait subjects, make up a significant portion of the collection. The collection also documents a diverse global population including people of African, Asian, and Latin American descent.

### Organization of the Portrait File

The physical collection is organized alphabetically by portrait subject, which may be further divided into topical sub-categories. For example, the 1,359 items documenting Napoleon I are separated into 58 categories such as “As a young officer,” “On horseback,” and “In imperial robes.” By contrast, the Danish poet and novelist Jeppe Aakjær, is represented with just one image. This hierarchical subject-based organization of the physical collection was maintained in the metadata structure: There is a record for the collection at top; followed by a container for the letter of the alphabet corresponding to the Subject’s name; followed by a sub-container for the Subject, e.g., Napoleon I; followed by optional topical sub-container e.g., “As a young officer.” These are followed finally, by the item and its digital capture (Fig.2).

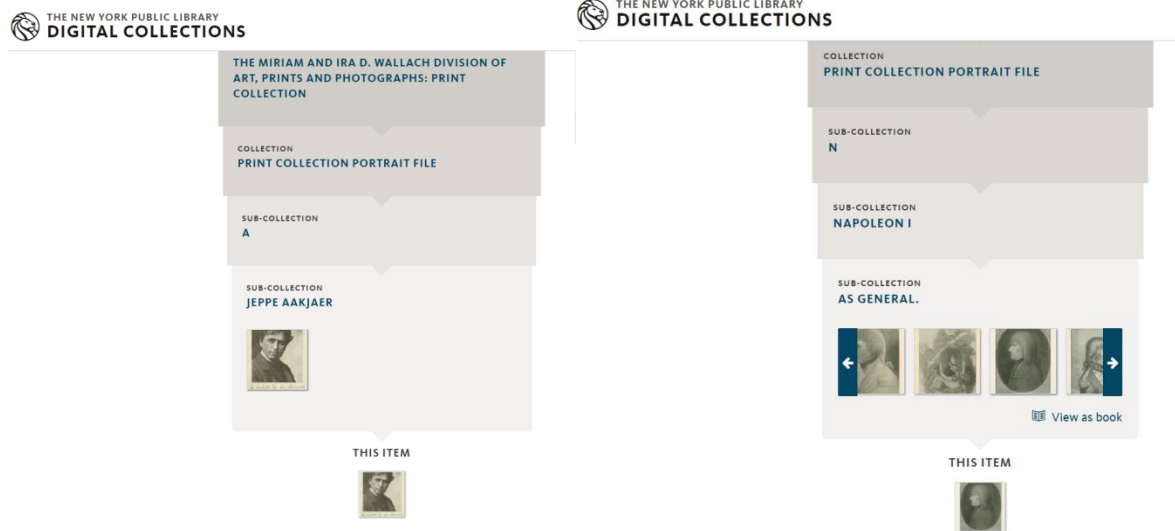


Figure 2. Screenshots of NYPL’s Digital Collections website showing Portrait File collection metadata structure for two items: portraits of Jeppe Aakjær and Napoleon I respectively. The library division owner is shown above the portrait of Jeppe Aakjær.

### Processing the Collection and Creation of Metadata

Metadata creation for the Portrait File began prior to the launch in 2005 of the Digital Gallery, NYPL’s first online image library. The institutional goal was to get hundreds of thousands of items online, and metadata creation practices for collection items reflected the need to add content quickly. Titles were inherited from parent to child records; Subject names and creator Names were entered inconsistently; Genre terms were not entered; Notes citing publication sources of an item were not often made; and Dates were not entered for all items. While these practices speeded work in the early days of the project, they created problems that magnified as the collection metadata grew.

Since the compilers of the collection did not have the benefit of online research to unambiguously identify the subjects of thousands of images, items were misfiled, and the physical collection was further disorganized after years of use. In the early phase of the project, metadata was created for the items in the order they were found in Portrait File boxes, with little apparent sorting or outside research conducted beforehand. As a result, in the metadata and capture organization, Portrait File Subjects were misidentified, conflated, and misplaced.

With experience as an item-level cataloger of archival materials, I began creating metadata for the Portrait File in 2008. I worked under NYPL’s then metadata manager who gave me the leeway to establish parameters for the project. It soon became clear that, in order to create reliable metadata, the physical items first needed to be sorted by subject and by visual content. An archivist colleague showed me how to re-house items in boxes in white acid-free paper folders, replacing yellowed folders that were possibly more than a hundred years old. With a metadata production assistant joining the project in 2012, we moved forward in earnest, sorting the Portrait File items box by box, and then entering metadata.

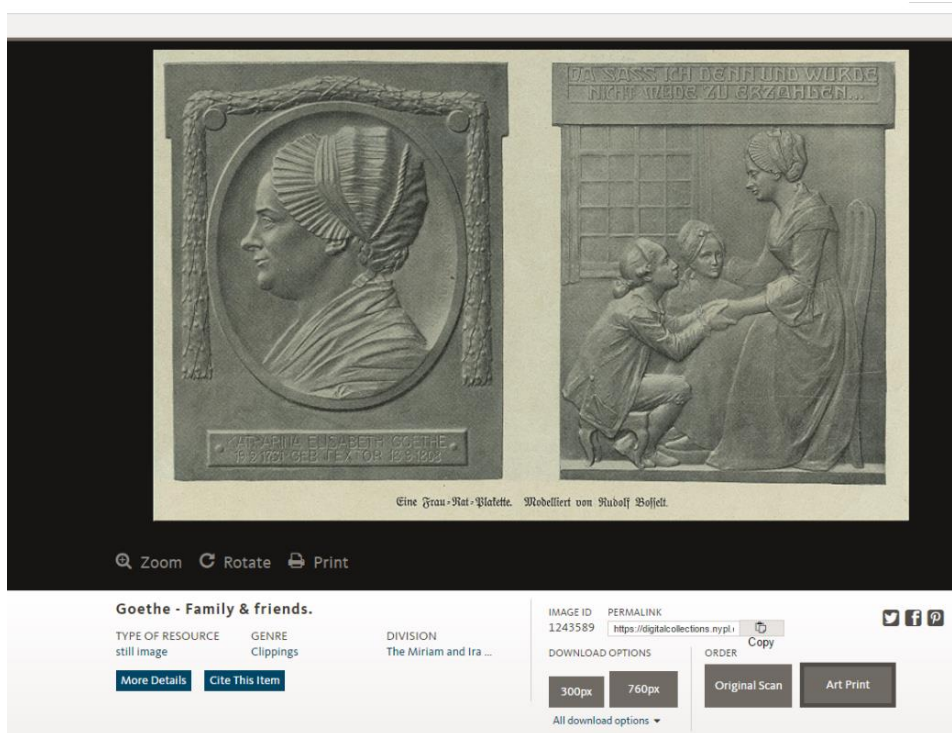
Metadata staff worked on an in-house built Oracle relational database, entering data in a form that accommodated MARC data standards. For guidance in entering descriptive metadata, we referred to the published cataloging standards *Anglo-American Cataloguing Rules (AACR2)* and *DCRM(G): Descriptive Cataloging of Rare Materials (Graphics)*. For indexing terminology,

we used the Library of Congress's Name Authority File (NAF), its Subject Headings (LCSH), its Genre/Form Terms (LCGFT), and its Thesaurus for Graphic Materials (TGM). We also used the Getty Research Institution's Union List of Artist Names (ULAN) and its Art & Architecture Thesaurus (AAT). With few other formal guidelines for our work, best practices for entering descriptive metadata were conveyed mainly in conversation and via email.

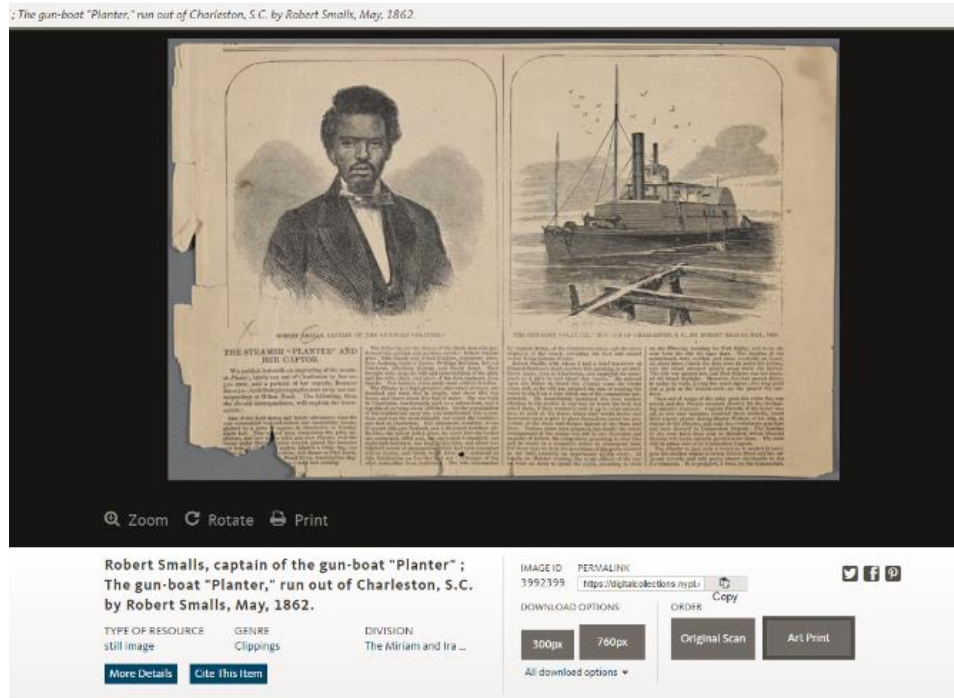
Since the main access point to the collection's items is by Name subject, researching and identifying personal subjects of Portrait File images unambiguously, with their full name and life dates, became the main priority of our work. We conducted online research on websites such as Wikipedia, the National Portrait Gallery (UK), The Peerage, and Find A Grave. We consulted catalog records in the union cataloging database OCLC Connexion client to find name headings not available in LCNAF. We researched biographies and obituaries in newspaper databases. TinEye Reverse Image Search pointed us to images matching Portrait File items, which led us to data about the subject of the image. With this research, we were able to positively identify about 99 percent of Portrait File subjects, so we could match them with their published Name authority heading, or if necessary create a local heading, and enter it in metadata records.

### Refining Metadata Creation Practices

As we worked, we continued to refine the metadata creation practices. We discontinued inheriting titles from higher-level record to items, and began transcribing the unique titles or captions of each item, working in many languages and deciphering handwritten text and signatures (Fig. 3).







Figures 3 and 4. Screenshots from NYPL’s Digital Collections website featuring two items from the Portrait File collection. The item at left, “Goethe—Family and friends” has a Title inherited from a higher level metadata container, while the Titles of the two images for the item at right, “Robert Smalls, captain of the gun-boat ‘Planter’; The gun-boat ‘Planter,’ run out of Charleston, S.C. by Robert Smalls, May, 1862” were transcribed directly from the image captions on the item.

Taking advantage of data inheritance capability of the database, which allows data elements at a higher level to be inherited to records below, we applied the Genre term “Clippings” at the collection level and inherited it down to all the items in the collection. We entered additional genre terms such as, “Portraits,” “Cartoons,” and “Silhouettes,” in sections of the collection where they applied, also employing data inheritance when feasible. We entered topical subjects like “Childhood & youth” “Family,” “Sculpture,” and “Homes & haunts,” at the container level and inherited them down to items when possible.

We also began citing the sources of published items in notes, e.g., “Published in The daily graphic, March 23, 1874, page 75.” We gleaned this information from handwritten citations made on the clippings by the compilers of the collection, or from information imprinted on the item in running heads or elsewhere.

For a time, we entered names of creators in the Name element, but we ended up abandoning this time-consuming practice in favor of transcribing the data in a Note with a statement of responsibility attribute, e.g., “Photograph by C.M. Bell, Washington, D.C.” We did not enter headings for gender or occupation of portrait subjects since this data was not always applicable to items, and because it can be accessed in NAF records for Portrait File subjects.

Metadata creation for the Portrait File continued through the migration in 2012 to the Metadata Management System or MMS, NYPL's homegrown platform for managing metadata, imaging, and rights workflow. The system uses a MODS (Metadata Object Description Schema) data schema. In order for records to validate and save, the MMS required at minimum five data elements: Title, Identifier, Type of Resource, Date, and Location (of original), denoted by a red star in the data entry form. Genre later became a system-required data element. These developments coincided with the adoption of Resource Description and Access (RDA), which replaced AACR2 as the cataloging standard at NYPL, The Library of Congress, and the library community at large.

With a date now required for item-level records, date research became a major part of metadata creation for Portrait File items, many of which lacked dates. To assign a probable date or date ranges of issue, we examined items carefully, evaluating periods of activity of the image creator, the run dates of a given source periodical, and the date of invention of reproductive technologies—for example, the reproduction of photographic images in print was not possible much before 1895. We considered the life dates of the subject and dates of events cited in the image caption to determine when the item may have been published. We also consulted outside sources like the British Book Trade Index for publishers' dates of activities, and Google Books for digitized periodical issues. This work necessitated the creation of the in-house "Portrait File Cheat Sheet," a nine-page reference document listing major periodical runs mainly in Great Britain and the United States and image creator names and their life dates, among other data. Since clippings are published, we employed the date type "Date issued," selecting the qualifiers "approximate," "inferred," or "questionable" when needed, and entered a date or date range in every item-level record.

With the formation of the Metadata Services Unit (MSU) in 2013, a now a six-member team which oversees metadata activities institution-wide, and with the hiring of its manager in 2014, we began to more deeply formalize descriptive metadata creation practices at NYPL. One of the manager's first tasks was to create a data dictionary which spells out the data elements of the MMS metadata profile in MODS. The MSU also began a documentation effort which included the creation of metadata guidance and best practices as well as a metadata creation training program.<sup>1</sup> In 2014, NYPL launched its Digital Collections website, a major overhaul of the first digital library with a new interface and features that aid searching and discovery.

### **Recommendations for Enhancement of Portrait File Metadata**

With the last box, labeled "XYZ," metadata creation for Portrait File was completed in October 2015. Immersed in the new data quality and enhancement effort in the MSU, my thoughts turned to the quality of Portrait File metadata my colleagues and I had spent many years creating. I drafted a report outlining areas of the metadata that needed enhancement. Below are some highlights from this report:

- Enter the title as found on the item. Image captions provide rich data that can be discovered by keyword search. Entering distinct and meaningful titles by transcribing the title as found on the item will greatly enhance discovery of the collection. Title is mandatory for metadata created in the MMS and parent titles are not inherited.
- Ensure that each item has an encoded date or date range. Dates are mandatory in MMS

---

<sup>1</sup> Averkamp and Rubinow 2017; <http://bit.ly/NYPLmetadataMETRO17>



and for baseline discovery of the collection.

- Name subjects and their associated URIs (Uniform Resource Indicators) should be entered in all items when available. Use of the Subject Name element is recommended by the MSU to enhance discovery, and URIs allow linked data, which will allow the exposure, connection, and linking of Portrait File data on the Semantic Web.
- Enter Names of creators, selected if possible from an established authority, and their associated URIs in Portrait File item records. This is recommended by the MSU to enhance discovery of the collection.
- Genre terms allow users to filter search results by content type, a major help when searching vast digital collections. The inclusion of at least one genre term is mandatory in metadata for item-level records. While the presence of “Clippings” in all items fulfills the requirement, adding additional terms, like “Photographs,” “Prints,” and “Political cartoons” to item-level metadata will allow users to more deeply refine searches by content type (Fig. 4).
- Publication data provides users with a context for learning more about Portrait File items. Notes should be made citing published sources of items, when they are known. The format of Note text strings should be standardized so that data can be parsed and easily used later, for linking it to source periodicals, for example.

The screenshot displays the NYPL Digital Collections website. On the left, there is a navigation sidebar with sections for 'TOPIC', 'NAME', 'COLLECTION', 'PLACE', 'GENRE', 'PUBLISHER', 'DIVISION', 'TYPE', and 'DATE RANGE'. The 'GENRE' section is expanded, showing a list of genres and their counts: Clippings (73543), Portraits (72337), Political cartoons (335), Caricatures (337), Illustrations (84), and a 'Less' link. A 'Filter genre by keyword' table is also present, listing genres and their counts: Clippings (71,343), Portraits (71,337), Political cartoons (335), Caricatures (337), Illustrations (84), Photographic prints (43), Cartoons (Commentary) (39), Sculpture (58), Leaflets (27), Stats (22), Photocopies (21), Photographs (18), prospectuses (15), Postcards (12), Advertisements (12), Catalogs (12), Title pages (11), Etchings (10), Montages (10), Correspondence (7), Medals (6), Engravings (6), Forms (Document) (6), busts (6), Book covers (5), Frontispieces (5), Manuscripts (5), Negatives (5), and Engraving (4).

The main content area shows a search bar with the text 'Search keyword(s)' and a 'Browse' button. Below the search bar, there is a grid of search results, each consisting of a thumbnail image and a caption. The captions include names and titles such as 'Raphael Sabatini', 'Comtesse de Sabran', 'Ettore Sacchi (Giustizia)', 'Giuseppe [i.e. Giuseppe] Sa...', 'Signora Sacconi-Ricci', 'Hans Sachs', 'Hans Sachs, 1494', 'Hans Sachs, Geb. d. Nov. 14...', 'Hans Sachs (aged 81); Sach...', 'George Sackville', 'No. X. The lovely Emily. No...', 'The hon. Lionel Sackville-W.', 'Lord Sackville retires', 'Ph. Sadee', 'Señor Sagasta, ex-prime min...', 'Praxedes Mateo Sagasta. 1827', and 'Señor D. Praxedes Mateo Sag...'.

Figure 4. Screenshot NYPL’s Digital Collections website search and discovery feature, showing some of the Genre terms by which Portrait File items can be filtered.

The 2016 report, “Aggregating and Representing Collections in the Digital Public Library of America” emphasized the need to create context especially in the form of collection

descriptions to improve user access.<sup>2</sup> At the collection level, the data in the Description (or abstract) element should be enhanced to provide users with a clear understanding of the background, content, and organization of the collection. Doing so would support the ongoing work in the MSU to enhance collection-level records and their Descriptions.

The enhancements described above can be implemented in targeted metadata remediation projects. These can be done manually by staff, via crowdsourcing, or by making automated global updates in the database.

### **Next Steps in Improving Metadata**

Additional steps can be taken to improve Portrait File metadata. Site analytics, for example, can be used to make correlations to data quality. With 71,557 items, the Portrait File collection is the largest collection in NYPL's Digital Collections website, but in early 2017 it was only the sixth most used over the previous two years. Why is this and how does this relate to data quality for the collection? It would be interesting to run a comparative audit of the top three most used collections to see what inferences could be made. User feedback is another area that could be fruitful to study. While the MSU hasn't done formal studies of Portrait File user behavior, user feedback from our website often concerns incorrect and missing data which suggests the need for metadata cleanup. Curatorial staff can likewise contribute valuable feedback on users' experiences with NYPL's digital collections, which can point out areas in which to improve collection metadata.

### **Conclusion**

Metadata creation practice for the Portrait File evolved as metadata practice at NYPL developed, leading to the formation of the MSU and the establishment of formal standards for the creation of metadata institution-wide. My goal in relating this history is to provide a rationale for the implementation of some of these standards; show how they can be used to evaluate data quality of Portrait File items, and to present some possibilities for the enhancement of the collection's metadata in the future. The Portrait File is a record of our culture, our history, and ourselves. Portrait File metadata is likewise a record of our legacy as digital librarians. By opening a window on our past, we throw a light on our present, and from here we can have a clearer view of the future.

### **References**

Averkamp, Shawn and Sara Rubinow. "Measuring Metadata for the Masses: Starting a Data Quality Program from Scratch." Slideshare of presentation given at the METRO Conference, New York, NY, January 11, 2017. Accessed June 28, 2017.

<http://bit.ly/NYPLmetadataMETRO17>

Digital Public Library of America. November 2016. "Aggregating and Representing Collections in the Digital Public Library of America." Accessed June 26, 2017. <http://bit.ly/dplaCollections>

Waldron, Zoe. "Life Story: Creating Metadata for a Portrait File" Slideshare of presentation given at the Visual Resources Association Conference 2017, Louisville, KY, March 30, 2017. Accessed June 26, 2017. <http://bit.ly/2t5pSil>

---

<sup>2</sup> DPLA 2016. <http://bit.ly/dplaCollections>

## Links

The Print Collection Portrait File: <https://digitalcollections.nypl.org/collections/print-collection-portrait-file#/?tab=navigation>

NYPL Digital Collections: <https://digitalcollections.nypl.org>

Miriam and Ira D. Wallach Division of Art, Prints and Photographs:

<https://www.nypl.org/about/divisions/wallach-division>

Anglo American Cataloguing Rules (AACR2): <http://www.aacr2.org>

DCRM(G): Descriptive Cataloging of Rare Materials (Graphics): <http://rbms.info/dcrm/dcrmg>

Library of Congress Name Authority File (NAF): <http://id.loc.gov/authorities/names.html>

Library of Congress Subject Headings (LCSH): <http://authorities.loc.gov/cgi-bin/Pwebrecon.cgi?DB=local&PAGE=First>

Library of Congress Genre/Form Terms (LCGFT): <http://id.loc.gov/authorities/genreForms.html>

Thesaurus for Graphic Materials (TGM): <http://www.loc.gov/pictures/collection/tgm>

The Art & Architecture Thesaurus (AAT):

<http://www.getty.edu/research/tools/vocabularies/aat/index.html>

The Union List of Artist Names (ULAN):

<http://www.getty.edu/research/tools/vocabularies/ulan/index.html>

Metadata Object Description Schema (MODS): <http://www.loc.gov/standards/mods>

RDA Resource Description and Access Toolkit: <http://www.rdatoolkit.org>