The New York Public Library

Digital Research Strategy

2021–24

nypl.org/digitalresearch
Vision

The New York Public Library's Stephen A. Schwarzman Building was erected on the site of the Croton Distributing Reservoir. This vast four-acre structure provided the city with drinking water throughout the nineteenth century. That history remains a part of the Library today: the physical vestiges of the reservoir are embedded in the Library's foundation, visible in the building's walls.

The reservoir’s purpose informs our present-day work. It pooled a vital resource—20 million gallons of water—and provided freely to New Yorkers. NYPL has gathered an equally critical resource—a body of knowledge spanning 50 million research collection items—for the benefit of the public. But access to this asset has been limited by our very edifice. Archives like ours have functioned as reservoirs without conduits, sites of pilgrimage for those seeking the resource they hold but with limited means of sharing that resource outside their walls.

It is only now that digital tools allow us to share our collections more freely. The ways in which people interact with and access information have fundamentally changed to prioritize digital experiences. The global pandemic has only intensified trends that existed long before society was shuttered. We believe now that we must bring forth a world in which access to the collections we steward is no longer predicated on physical presence in our research libraries.

NYPL has long championed the digital revolution in libraries. We have transferred collections from the stacks to server racks, harnessed crowdsourcing to extract machine-readable data, and led the development of e-reading platforms. This work has transformed our collecting strategies, preservation efforts, and access services. Our Digital Research Strategy 2021–24 presents a set of ambitious priorities that NYPL will pursue to continue and quicken this digital transformation: to seek radical access, to democratize our collections, to deepen engagement, and to reinforce systems and infrastructure.

We will continue to care about and for paper and petabytes in the coming years, but as we move forward we will prioritize digital delivery as our primary method of access. Being “digital-first” will open our collections to global audiences, allowing us to reach millions affordably and without restrictions of time or space, or a patron's abilities. It will make material easier to discover, sample, use, and reuse in more creative ways than are possible now. With this approach, we can fully embrace our mission as both a public library and a research archive.
Strategy: At a Glance

The New York Public Library’s Digital Research Strategy 2021–24 focuses on four strategic principles:

1. Seek Radical Access
   We will eradicate physical barriers to information. We will leverage technology’s capacity to enhance discovery efforts. And we will prioritize global access, making as much material public as possible and expediting the sharing of our collections. We will consider:
   - Descriptive Practices
   - Remote Access
   - At-Risk Digitization
   - Copyright Parsing

2. Democratize Our Collections
   We will develop platforms, digitize collections, and pursue projects that serve an audience which reflects the diversity of our city, nation, and world. This is a fundamental rethinking of how we must continue to adapt to serve our public. We will consider:
   - Diversity, Equity, and Inclusion
   - Access-Level Digitization
   - Enhanced Discovery Capability
   - User Research

3. Deepen Engagement
   We will create online collection experiences that rival and surpass the quality of studying in our grandest reading rooms. We will invest in best-in-class viewing platforms, support inventive reuse and remixing of our digital collections, and train our staff to be fluent in digital work. We will consider:
   - Digital Collections
   - Projects and Experiments
   - Staff Development
   - Communications

4. Reinforce Systems & Infrastructure
   We will expand storage and digitization systems and increase our capability to steward and serve born-digital formats. We will grow and support the teams that undertake this work and develop strong governance and oversight structures. We will consider:
   - Digital Preservation
   - Digital Storage
   - Born-Digital Collections
   - Governance & Oversight
Strategy

The Digital Research Strategy focuses on four strategic principles, each with four initiatives.

1 Seek Radical Access

We will eradicate physical barriers to information. We will leverage technology’s capacity to enhance discovery efforts. We will prioritize global access, making as much material public as possible and expediting the sharing of our collections.

- Descriptive Practices
  Make material available more quickly by refining descriptive practice

  Description democratizes access. Our patrons can find library resources because they are cataloged, and they can only find our digital collections when they are adequately described. We are working to remove obstacles to descriptive processes, making time-based media available online at a pace not previously possible. These changes will benefit all our descriptive workflows and shape a more open and engaged culture around description with our research centers. Work here includes finding more efficient ways to update archival description in our digital collections access platforms, new automated communications that will allow us to prioritize new description, and new forms of communication and collaboration among our curatorial divisions and production teams.

- Remote Access
  Create secure distanced access to rights-restricted items

  We steward one of the world’s greatest collections of time-based media, from Lou Reed’s demo recordings to Sonny Rollins’s rehearsal tapes and from Martha Graham’s dances to rare newsreels and films documenting the 1960s Civil Rights Movement. While we can provide remote access to paper and still image collections through our services, we have no means of providing access to the vast majority of our audio and moving image collections at a distance because of copyright law.

  This challenge is not unique to NYPL; many academic libraries struggle with serving time-based media to remote users. We are working with our peers to develop a legal framework and leverage new technologies that together will provide patrons with secure access while protecting the rights of rights holders. We will create a new “Virtual Reading Room” that allows limited access to users with bona fide research needs—providing a window to collections that have heretofore been available only at our research centers.

Many of our audio and moving image collection items, like this Annabella Whitford Moore dance of 1897 captured and hand-tinted by Thomas Edison, are unavailable onsite because of copyright restrictions. By developing a Virtual Reading Room, qualified researchers may gain access to restricted collected items.

Learn more about this image.
• **At-Risk Digitization**  
*Renew strategy for undigitized unique at-risk holdings and for new acquisitions*

In 2014, NYPL launched its Audio and Moving Image (“AMI”) Initiative, a large-scale effort to preserve, digitize, and create access to time-based media that is at risk of permanent loss through media degradation and technology obsolescence. As of December 2020, the Library has identified more than 400,000 rare, unique, or at-risk items, and inventoried 315,000 and digitized over 200,000 of those items.

We have established a new Media Preservation Services team to manage pre- and post-production activities and expanded our in-house conservation and digitization capabilities for new media formats. Parallel transfer workflows will support the Library’s long-term commitment to AMI preservation. Critical investments will support the continued evolution of the AMI Initiative from project-based to programmatic services as we look to maintain our ability to conserve and digitize new acquisitions of at-risk media for years to come.

NYPL holds more than 400,000 at-risk audio or moving image collection items, held on dozens of formats spanning nearly every type of optical, magnetic, grooved, and film media. Beginning in 2014, the Audio and Moving Image Initiative has digitized over 200,000 time-based collection items. But continued investment is needed to continue this work as we focus on expanding in-house conservation and digitization capabilities and maintain our ability to respond to new acquisitions of at-risk media. Photo by Martha Swope.  
[Learn more about this image](#).

• **Copyright Parsing**  
*Identify the copyright status of collection items by transcribing and parsing the Catalog of Copyright Entries*

Since 2017, we’ve received grants to transcribe the class A registrations from the 1923–77 Catalog of Copyright Entries, a list of all copyrights published by the US Copyright Office. By transcribing the 87,920 pages of these records—and making that data set publicly available and user-friendly—we can identify and make freely available thousands of works that are in the public domain because they are out of copyright.
Democratize Our Collections

We will develop platforms, digitize collections, and pursue projects that serve an audience which reflects the diversity of our city, nation, and world. This is a fundamental rethinking of how we must continue to adapt to serve our public.

- Diversity, Equity, and Inclusion
  Prioritize digitization that champions diversity, equity, and inclusion

  Digitization is a choice and a responsibility, through which we can shine a spotlight on missing voices and narratives that celebrate the diversity in our holdings. It is also a uniquely powerful tool by which we can influence the collections that are accessed, used, and drawn upon to create new knowledge. Mindful of this potential, NYPL is committing all discretionary digitization capacity—our capacity beyond that which is committed to public orders and programmatic needs—to missing and underrepresented voices. For the next three years, we will focus specifically on African American, African Diaspora, and African experiences and culture.

The Schomburg Center for Research in Black Culture holds the largest collection of Green Books, the annual guidebooks published in the Jim Crow era that helped Black travelers navigate the country safely by listing the hotels, restaurants, beauty salons, bars, and gas stations where they would be welcome. Digitization of this resource helped spur the study and examination of the history of Black travel in the U.S. and contributed to the production of books, articles, news stories, documentaries, and even artworks. Our new digitization commitment means that we will be dedicating more resources to bringing critical resources like these to the public, ensuring that our digital collections reflect the diversity of the collections we steward and the public we serve.

Learn more about this image.
• **Access-Level Digitization**  
*Support a digital-first collections strategy for print collections by developing access-level digitization*

Our current digitization practices for print collections are built to serve high-value, high-prestige special collections materials. While this expert approach to imaging is necessary for everything from digital preservation of highly fragile materials to high-resolution capture for commercial reproduction, it does not meet the needs of a more typical library patron looking for expansive and fast off-site access to research collections, particularly printed volumes. Building on the Scan & Deliver service developed during the pandemic, we will pilot expanded Digital Imaging Unit services focused on access-level service for general print collections, building new workflows that prioritize faster, scalable, high-quality image capture, and delivering materials to patrons worldwide for free. This investment will allow us to meet the needs of more patrons, serve more materials, and embrace accessibility.

![Dante and the early astronomer: science, adventure, and a Victorian woman who opened the heavens / Tracy Daugherty.](image1.jpg)

A Scan and Deliver request button now appears on over 2 million NYPL records in the catalog.

• **Enhanced Discovery Capability**  
*Enhance discovery, description, and accessibility of print materials by implementing optical character and handwritten text recognition*

We will continue utilizing advanced technologies to improve the patron experience. This includes implementing optical character recognition and handwritten text recognition in order to enhance the discovery, description, and accessibility of print materials. Removing the barrier between information trapped in an image format and machine readability will help democratize our collections, making the information contained easier to discover for all, and, for the first time, give our patrons the ability to search within the volumes we have digitized.

![By investing in optical character and handwritten text recognition workflows, we will unlock data held in our digitized collections.](image2.jpg)

By investing in optical character and handwritten text recognition workflows, we will unlock data held in our digitized collections. Texts, like this 1914 double inscription, from Irish revolutionary Pádraic Pearse to W.B. Yeats, who then inscribed it to Lady Gregory, will be made more easily discoverable, ensuring a patron’s searches reveal more and more relevant collection items.

*Learn more about this image.*
• **User Research**  
*Undertake primary research to better understand users and to inform planning*

We are undertaking primary research to ensure our efforts and attention are focused on meeting all our users’ needs and desires. This work focuses not only on researchers using our tools to further their scholarship, but all patrons accessing our collections, whether it be for a creative project or to pursue their curiosity. This work also focuses on ensuring that our platforms meet the needs of the diverse public that we serve.

User engagement will make sure that we understand how our patrons are viewing and using items in our collections, like this illuminated 15th-century manuscript miniature depicting the creation of Eve from the body of sleeping Adam from *Des cleres et nobles femmes* by Giovanni Boccaccio. Learn more about this image.

---

3 **Deepen Engagement**

We will create online collection experiences that rival and surpass the quality of studying in our grandest reading rooms. We will invest in best-in-class viewing platforms, support inventive reuse and remixing of our digital collections, and train our staff to be fluent in digital work.

• **Digital Collections**  
*Redevelop Digital Collections as an accessible platform that engages the curious*

Our Digital Collections portal launched in 2005 as our Digital Gallery, a place where patrons could see a selection of images that had been digitized by the Library. Over the years, the gallery grew in scale and morphed in function: it now serves over 900,000 digitized collection items and is a digital reading room for our digitized content.

This platform is in need of modernization. The ability to wander and discover serendipitously is challenged by the layout of the item pages and the sheer quantity of items now available. Tools are required that can allow curators to contextualize collection items. Better bridges to our catalog and archives portal are required, as are faceted and complex search functions so users can make sense of our ever-growing digitized collections. Improvements are also needed to ensure that the platform meets the accessibility needs of all our patrons.
Investments in the Digital Collections portal will make the platform more accessible and engaging. Tools will allow curators to give more context to key collection items, such as this sketch by Ludwig van Beethoven for the scherzo from his “Archduke Trio”.

Learn more about this image.

• Projects & Experiments

*Launch projects and support experimentation that showcases the reuse and remixing of digital collections*

We have a deep history of creating and supporting digital humanities explorations of our public domain collections. We will continue providing support for projects that show the many ways our digital collections can be utilized, in order to encourage more people to engage creatively. This includes collaborations with partner universities and developing future scholarships for digital humanities study at the Library.

NYPL has long supported projects and experiments that reuse and remix digital collections. This early plan of Central Park by George Hayward was part of Map Warper, a tool we developed focused on rectifying historical maps to match today's more precise maps. We will continue to invest in and support partnerships in the digital humanities that draw on our digital collections.

Learn more about this image.
• **Staff Development**  
*Facilitate staff development, including training staff on our Digital Collections API*

In order to empower our patrons to access and utilize our collections, we also need to empower our staff to understand our Digital Collections platform and tools so that they can provide reference assistance to patrons who wish to do the same. We are creating new training and documentation that will allow staff members to engage more deeply with our digital collections platform and API (application programming interface), building their comfort and familiarity with the increasingly digital nature of our collections.

Collection items, like this Bernard Schadt image created during the Works Progress Administration, are available to patrons on our Digital Collections site, or directly through the site’s API. New staff training will make sure our librarians are ready to assist patrons with any machine-based inquiry drawing on that API.  
[Learn more about this image.](#)

• **Communications**  
*Introduce the public to our strategy through a proactive communication plan*

We want to keep you informed! We have created this strategy to share the work we are doing now and our plans over the coming years. We will share updates as we make progress on our goals. We will also seek forums to discuss projects at key inflection points, ensuring that our communication is two-way and that we make opportunities to hear from you.

### 4 Reinforce Systems & Infrastructure

We will expand storage and digitization systems and increase our capability to steward and serve born-digital formats. We will grow and support the teams that undertake this work and develop strong governance and oversight structures.

• **Digital Preservation**  
*Improve preservation, security, and access by delivering new repository software*

Over the past decade, our digital storage and preservation needs have rapidly changed. We have developed and deployed systems to ensure that the millions of files held in our repository are safely stowed and easily accessible. As part of that evolution, we now look to implement new digital preservation software so we may improve our ability to provide a secure and trustworthy digital preservation environment for all digital files. Once implemented, this new system will provide better support for and access to all digitized and born-digital formats.
• **Digital Storage**  
*Evolve our digital storage plan and expand storage to meet needs*

Our digital collections are growing rapidly as we acquire new collections and digitize and preserve analog formats. We will continue to invest in expanded storage to ensure that our work is not limited by our ability to hold our collections. We currently manage ten petabytes of content and we expect to grow our storage by at least one petabyte per year in the coming years. In addition, we plan to shift our storage to prioritize cloud-based archiving, creating a future-oriented storage environment that provides improved access and security.

---

[Image: Continued investment in digital storage will ensure that we continue to have adequate space as we collect and digitize. With this investment we can guarantee that our collection items, like this image of Maya Angelou and Charles Gordone from the 1961 stage production of *The Blacks*, taken by Martha Swope, remain safely available to our patrons.](Learn more about this image)

---

• **Born-Digital Collections**  
*Implement preservation and access pathways for born-digital materials*

We are implementing preservation and access pathways for born-digital materials, collection objects that are digital at acquisition and will never have a physical form. As these will continue to make up an ever-increasing volume of our new collections, it is essential that we improve our born-digital media ingest process and create seamless means of access. Our current focus is on web-based and email archives, as well as born-digital time-based media.

• **Governance & Oversight**  
*Enhance our governance and oversight, and ratify our first digital preservation policy*

As digital assets make up more and more of our collections, our sophistication and processes around digital preservation must continue to evolve. We have long stewarded digitized and born-digital collections, but we face an inflection point as we focus on our digital transformation. We are creating new ways to ensure that stakeholders across our production teams and curatorial divisions are consulted and informed in our preservation work, and that responsible and accountable parties' responsibilities are made explicit. As part of this work, we will publish our first digital preservation policy, a standalone policy focused on our standards and workflows that ensures appropriate stewardship of our collections.
About

The Digital Research Strategy 2021–24 is overseen by NYPL's digital research team. This group manages the processes and practices involved in the acquisition, creation, description, storage, discovery, reuse, and preservation of NYPL's unique digital assets.

Contact us: DigitalResearch@nypl.org

Learn more: npl.org/digitalresearch