

## *Chapter 13*

---

# Museums, Archives, and Universities—Structuring Future Connections with Big Data

---

Jane Milosch, Michael J. Kurtz, Gregory J. Jansen, Andrea Hull, and Richard Marciano

### **Contents**

Introduction.....	160
New Strategies.....	161
New Efforts to Aggregate Collections Data.....	162
Development of World War II-Era Provenance Research.....	162
Identifying Key Challenges.....	164
New Partnerships and Big Data.....	165
The Enhanced “International Research Portal for Records Related to Nazi-Era Cultural Property” Project (IRP2): A Continuing Case Study.....	166
Provenance and Technical Challenges .....	167
Project Management.....	167
Technical Development .....	169
Continued Development and Implementation .....	170
Student Learning.....	171
Conclusion and Next Steps.....	172
Reference.....	172

## Introduction

Big data's value to the arts, humanities, and other disciplines is contingent upon the integrity, verification, and safeguarding of its information—its “provenance.” For art museums, provenance research on its collections and acquisitions attempts to establish an unbroken chain of ownership from an object's creation to its present owner. This information is traditionally either departmentally “siloe” within an institution or between institutions or, in Germany, for example, is the intellectual property of the researcher. With the advent of digital humanities, enabled by big data, the discipline of art history is encouraged to expand beyond its traditional bases in connoisseurship and history of collecting to encompass broader fields of study.\*

In Maxwell Anderson's 2015 essay, “The Crisis in Art History: Ten Problems, Ten Solutions,” first presented at a 2011 College Art Association of America meeting, he notes that art historians have become “a fragmented pool of experts,” and in order to “elevate the importance of their discipline in the hearts and minds of academics and non-academics alike,” it is necessary to build stronger connections between the “life of mind,” surviving works of art, and the public. He also encourages a new generation to “devote their lives to the study and care of cultural heritage.”<sup>1</sup>

Provenance research—fundamental to the histories of art and of collecting—might be the fulcrum on which to balance these new efforts Anderson calls forth. In the United States and in Germany, recent museum exhibitions drawn from their permanent collections have taken provenance as one of their guiding themes, organizing the presentation of the artworks to capitalize on the diverse stories provenance research can tell. These exhibitions have been popular with the public and critics alike, as they reveal new links between artists, collectors, dealers, and the markets in which they meet. The result of curators working with provenance researchers and archivists, they move compelling provenance stories from back of the house to the front of the house. Stories that engage the public and young scholars alike will become increasingly available when digital humanities and big data

---

\* In October 2016, The Phillips Collection in Washington, DC, and the University of Maryland in College Park, Maryland, organized a symposium, “Art History in Digital Dimensions,” aimed “to unite diverse audiences and practitioners in a critical intervention for the digital humanities and digital art history, providing a cogent and inclusive road map for the future.” In the public keynote lecture, “Digital Art History: Old Problems, New Debates, and Critical Potentials,” Paul B. Jaskot, professor of art history at DePaul University, gave a brilliant summary of the challenges and goals of digital humanities for art historians, which formed the backdrop of the lived-streamed discussions about “collaborative, trans-disciplinary models of research; the implications of data-driven approaches to art history and the humanities; legal and ethical obligations of scholars and museum professionals engaging art history in the digital world; and the innovative array of objects for study presupposed by digital art history,” which followed at the University of Maryland. See <http://dah-dimensions.org> (accessed April 24, 2017).

progressively inform museum exhibition practices. Because museums are unique in the scholarly world in that they engage extensively with the public, such exhibitions will help ensure the continued vitality of the discipline of art history.

## New Strategies

Museum curators who conduct provenance research on art objects in museum collections traditionally record their results in narrative format only and then turn over their texts to registrars and other data-entry specialists, who enter it into one of several different collection databases that museums use.

With the coming of big data, many emerging art historians and museum professionals are working toward recording their research not only in narrative format but also in new formats that enable the sharing and exchange of these data across collections and institutions—they understand that it is no longer enough to work only in isolation and on a single track. Museums have much to gain from making their material attractive to aggregators, virtual museums, educators, and others, who are seeing that the benefits of collaborative scholarship enabled by digital humanities are increasingly valuable to better contextualize their objects and build richer stories around them.

American museums are ahead of most museums in Europe (let alone those of other geographical areas that were prolific creators of art, such as Africa, India, and China) in digitizing their collections. In most instances, their provenance information is automatically updated and shared with the public in real time and includes images of the objects. Concurrently, museums have a strong interest in protecting the integrity of their research resources and results from accident, corruption, and malfeasance and to illuminate gaps in provenance data that indicate need for further research. Many, if not most, objects have gaps in their ownership histories—particularly decorative and graphic art objects, often produced in multiples. And because provenance data and records are often dispersed over many collections, databases, and archival holdings siloed in institutions around the world, sometimes, establishing true ownership is quite complex.

Currently, linked open data (LOD) appears to offer the best path to address these issues. Through LOD, with widespread adoption of a single standard for recording provenance data, provenance information will, at some point, be viewable across institutions and collections, allowing museums to ask more nuanced questions about the lives of their objects—the people and places through which the object has traveled—telling a more complete history of art, and its place in a larger socio-historical narrative.\*

---

\* For more information about the launch of the “Art Tracks: Standardizing Digital Provenance Documentation for Cultural Objects” project at the Carnegie Museum of Art, visit <http://press.cmoa.org/2015/12/17/cmoa-neh-art-tracks/> (accessed April 24, 2017).

## New Efforts to Aggregate Collections Data

Although the road to get there may be a long one, several institutions are taking their first steps on the Research to Provenance Search path outlined in Chapter 4 of this book. The Smithsonian’s online search platform, Collections Search Center ([collections.si.edu](http://collections.si.edu)), provides a user-friendly search interface for navigating Smithsonian collections information and associated digital media. Smithsonian museums and research centers contribute data to this site by mapping new information to an Index Metadata Model, and Smith has launched new initiatives to help bring provenance data together in new ways.

## Development of World War II-Era Provenance Research

Let us now back up a moment—to 1997, when the Association of Art Museum Directors (AAMD)\* convened a task force to draft guidelines on how its members should handle art looted by the Nazis and not previously restituted. These guidelines informed the “Washington Conference Principles on Nazi-Confiscated Art” drafted by the 1998 Washington Conference on Holocaust-Era Assets. Forty-four governments and 13 nongovernmental organizations met to discuss how best to ensure that museums conscientiously research the provenance of artworks potentially looted during the National Socialist era, addressing Jewish losses in particular and concentrating on painting and sculpture, books, and archives.

In May 2001, the American Association of Museums (AAM) issued a series of guidelines for helping identify and publish potentially confiscated works held in American museums and in 2003 launched the Nazi-Era Provenance Internet Portal (NEPIP; <http://www.nepip.org/>) to provide a searchable central registry of these works and their histories, accessible to the public. After a decade of NEPIP, nearly 30,000 objects from 179 participating museums are listed there. During that time, many museums, archives, and research institutions rapidly digitized records to help identify objects in US collections that had changed hands in continental Europe between 1933 and 1945, to help fulfill the objectives of the Washington Principles.

Since 2000, to honor its commitment to the Washington Principles, the Smithsonian Institution has financially supported the work of provenance researchers at its museums and in 2004 launched the “Smithsonian Institution WWII Provenance Web Site,” which includes a publicly searchable object database and other resources.

In 2009, the Smithsonian Provenance Research Initiative (SPRI)† was founded as a pan-institutional program to advance the institution’s ongoing,

---

\* For more information, visit <https://aamd.org/object-registry/resolution-of-claims-for-nazi-era-cultural-assets/more-info> (accessed April 24, 2017).

† For more about SPRI, <http://provenance.si.edu/jsp/spri.aspx> (accessed April 24, 2017).

serious commitment to provenance research of its collections—prioritizing objects in its collection that might have been misappropriated during the Nazi era—and to the preservation of cultural property and heritage, both considered integral parts of the Smithsonian’s mission of increase and diffusion of knowledge. In service to the field of museum provenance generally, SPRI is leading the construction of an international network of museum professionals and researchers who deal with various aspects of provenance research (concentrated on World War II [WWII]-era, but expanding to include postcolonial and antiquities), training emerging museum professionals who will take up the challenges of twenty-first century provenance, and assisting the development of new technologies that will enable provenance research, so long isolated in individual institutions, to enter the arena of large-scale social investigations enabled by big data and the digital humanities.

In 2011, in recognition of the need to build comprehensive international exchanges on provenance research, the National Archives and Records Administration (NARA), under the direction of Michael Kurtz, launched the International Research Portal for Records Related to Nazi-Era Cultural Property (Portal)\* to fulfill the objectives of the 1998 Washington Conference on Nazi-Confiscated Art, the 2000 Vilnius Declaration, and the 2009 Terezin Declaration to make all relevant records and information publicly available. The Portal is a collaboration of 18 national and other archival institutions with records that pertain to Nazi-era cultural property working with other expert national and international organizations to extend public access to widely dispersed records through this single Internet Portal.

In conjunction with the launch of the Portal, NARA partnered with SPRI and, working with the AAMD and the AAM, hosted a two-day seminar in which museum professionals and others interested in provenance learned about new electronic tools, collaborative projects, and strategies for research. Experts guided discussions about using these resources and shared results of recent and ongoing research projects. A total of 150 people attended, including 35 guest speakers and representatives from 60 museums, 34 US states, and 10 European countries.†

---

\* For more information about the International Research Portal for Records Related to Nazi-Era Cultural Property (Portal), visit <https://www.archives.gov/research/holocaust/international-resources> (accessed April 24, 2017).

† For more information about the World War II Provenance Research Seminar, May 2011, visit <https://ww2provenanceseminar.wordpress.com> (accessed April 24, 2017).

## Identifying Key Challenges

In the aftermath of the Portal launch and symposium, two observations stood out: one, the realization that when results of their provenance research cannot be broadly shared, nor resources accessed easily throughout the community of professional researchers and interested parties, museums' scarce financial and human resources are squandered. When each museum must constantly "reinvent the wheel," with exhaustive duplication of efforts in a proliferation of digitized provenance information, this is not sustainable in the twenty-first century. The second observation is that, although it was a monumental accomplishment to bring these 18 organizations together on the Portal, it was not searchable in a federated, or other meaningful, way, thus limiting its utility to museums and the field in general.

Around this time, the Freer|Sackler recognized that their provenance research and digitization efforts needed to be synchronized more effectively, to better share and exchange research and collection data with other Asian art museums and collections. The Freer and Sackler galleries, together considered one of the world's finest collections of the arts of the Far and Near East, have a long history of working to ensure complete access to their provenance research, beginning with its founder Charles Lang Freer's insistence on documentation to accompany his acquisitions.

The Freer|Sackler's provenance project, supported by SPRI, brings to the forefront scholarly interest in Euro-American aspects of the Asian art market before, during, and after WWII. In 2014, after the Freer|Sackler completed extensive research on nearly 400 Chinese objects with WWII-era provenance gaps, the museum quickly realized that this information was still of limited use because the information was neither contextualized nor searchable.

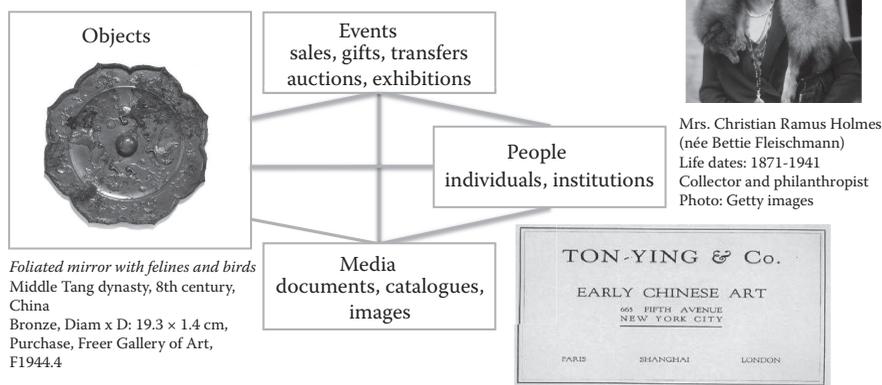
To contextualize its research of these objects, Freer|Sackler devised an approach to link 50 WWII-era Asian art collectors and dealers biographies with art objects and associated archival data. Funded by a grant from the David Berg Foundation, this project became known as the Asian Art Provenance Connections Project (Figure 13.1)\* and was launched in the spring of 2016. Its web pages (<http://www.freersackler.si.edu/collections/provenance-research/>) hyperlink objects with the collector/dealer biographies and with other Smithsonian-wide archival documentation. By creating a search interface that exposes relationships articulated through the data, the project promotes international awareness, supports more effective research methodologies, and facilitates international collaboration and information exchange between museums and the public concerning Asian art provenance.

---

\* For more information about the Freer Gallery of Art and Arthur M. Sackler Gallery of Art's provenance research efforts and resources, including the Asian Art Provenance Connections Project, visit <http://www.freersackler.si.edu/collections/provenance-research/> (accessed April 24, 2017).

Asian Art Connections Project artwork

The Asian Art Provenance Connections Project makes provenance information both searchable and linked to supporting resources



**Figure 13.1** For more information about the Asian Art Provenance Connections Project, <http://www.freersackler.si.edu/collections/provenance-research/>

## New Partnerships and Big Data

Recently, to move to the next step, the federated searches of various databases, the Freer|Sackler began investigating provenance information expressed as LOD. In Chapter 4, Smith discusses the Smithsonian’s collaboration with the Carnegie Museum of Art (CMOA) and the Yale Center for British Art, funded by a grant from the National Endowment for the Humanities. This collaboration uses the CMOA’s Art Tracks project to develop a program to convert provenance information expressed as narrative text into structured data, as a first step to LOD. Smith’s section articulates these recent developments, which herald the enhanced ability of art museums to develop and share provenance data and results.

In 2016, when SPRI learned of the University of Maryland’s Digital Curation Innovation Center’s (DCIC’s) efforts to enhance federated search of the NARA Portal and to develop visualization tools for data representation, SPRI partnered with DCIC to help them test the effectiveness of their prototyping. The following section of this chapter, by Michael Kurtz, author of the 2006 book *America and the Return of Nazi Contraband: The Recovery of Europe’s Cultural Treasures*, outlines a project employing students working with a professional software developer to map metadata for each of the 18 museums, libraries, and archives that are part of the Portal and to enhance its linking ability. Kurtz provides a case study of how big data and federated search capabilities are yielding exciting results and—Maxwell

Anderson would approve—are bringing students into the learning and producing process.\*

## The Enhanced “International Research Portal for Records Related to Nazi-Era Cultural Property” Project (IRP2): A Continuing Case Study

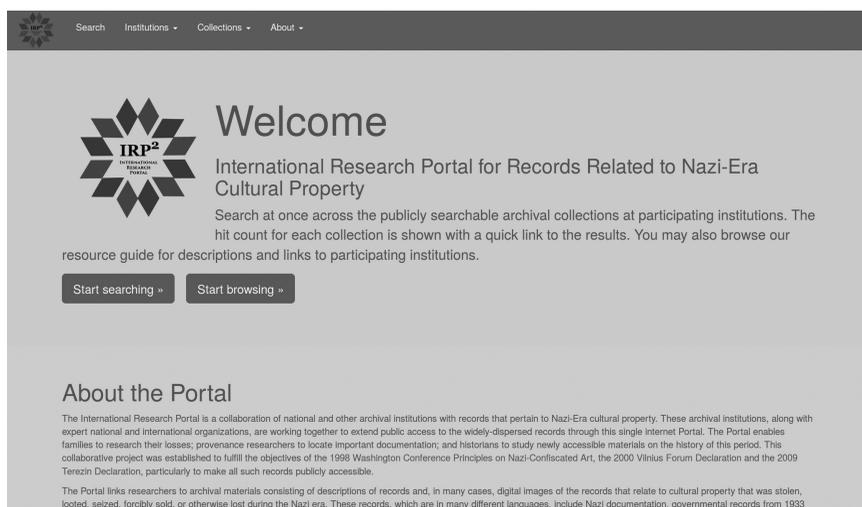
In February 2015, the DCIC at the University of Maryland’s College of Information Studies (iSchool) initiated the “International Research Portal for Records Related to Nazi-Era Cultural Property” project (IRP2), a research project to enhance information search and retrieval capabilities for archival, library, and museum collections maintained online on The International Research Portal for Records related to Nazi-Era Cultural Property (Portal). Dr. Michael Kurtz, associate DCIC director, is the project director, assisted by DCIC software developer Gregory Jansen and a team of iSchool students. The students are enrolled in three of the iSchool’s degree programs: Master of Library and Information Science (MLIS), Master of Information Management (MIM), and Master in Human–Computer Interaction.†

The Portal, launched in May 2011 and currently hosted by the US NARA, is a collaboration of 18 national and other archival institutions, libraries, museums, and research centers, with the intent of enabling researcher access across the distributed collections of the participating institutions. The Portal was established to fulfill the objectives of the 1998 Washington Conference on Nazi-Confiscated Art, the 2000 Vilnius Declaration, and the 2009 Terezin Declaration to make all relevant records and information publicly available (Figure 13.2).

---

\* The author would like to thank her colleagues from SPRI and Freer Gallery of Art and Arthur M. Sackler Gallery for their collaborative efforts that have greatly contributed to the ideas and project presented in this essay: Elizabeth Duley, head of collections, Freer|Sackler; Jeffrey Smith, assistant registrar, Freer|Sackler; and David Hogge, head of archives, Freer|Sackler, as well as SPRI project consultants (unless otherwise noted, currently working on projects): Lynn Nicholas, SPRI advisor; Laurie A. Stein, LLC, SPRI senior advisor; Dorota Chudzicka, Provenance research associate (2008–2013); Nick Pearce, Richmond Chair of Fine Art, University of Glasgow, UK and SPRI senior research fellow; Samantha Viksnins, project coordinator, Freer|Sackler (2015–2016); Andrea Hull, SPRI Communications and Grants Associate; Johanna Best, program manager/ACLS postdoc fellow; and Colleen Carroll, program assistant.

† Student volunteers included the following: Karishma Ghiya (MIM), Allison Gunn (MLIS), Torra Hausmann (MLIS), Weidong Li (MIM), Lisa Rogers (Master of Human–Computer Interaction), Sohini Sarkar (MIM), Sanjna Srivatsa (MIM), Jennifer Wachtel (MLIS), and Melissa Wertheimer (MLIS).



**Figure 13.2** The Portal is designed to look right on mobile and desktop browsers.

## *Provenance and Technical Challenges*

The Portal links researchers to archival materials consisting of descriptions of records and, in many cases, digital images of records that relate to cultural property that was stolen, looted, seized, forcibly sold, or otherwise lost during the Nazi era. These records, which are in many different languages, include Nazi documentation, governmental records from 1933 onward, and inventories of recovered artworks, postwar claims records, and auction house and art dealer records. Cultural property in these records covers a broad range of objects, from artworks to books and libraries, religious objects, antiquities, sculptures, musical instruments, and more (Figure 13.3).

Provenance issues abound. As is usually the case, archival information in the Portal collections is heterogeneous, the archival context ambiguous, and the original context difficult to reconstruct. Other problems hinder the effective use of the Portal. There is the problem of the sheer volume of information—hundreds of records series (complete and partial) and millions of discrete items of information, such as property cards of looted objects created by the Nazis and by the US Army during the postwar restitution program are siloed within the individual institutions, where it is not now possible to conduct searches and use search terms across the Portal.

## *Project Management*

During the first phase of the IRP2 project, MLIS students, with the active involvement of the project director and software developer, mapped metadata for each

The screenshot shows a search portal with a dark header containing navigation links: Search, Institutions, Collections, and About. The main content area is titled "Search" and includes a "Keywords" field with the text "Blau vase". Below this is a "Translate Keywords" section with buttons for "French" and "German", and "Start Search" and "Reset" buttons. A note states: "PLEASE NOTE: Searches do not include all participating institutions." A "Show Advanced Fields" section is expanded, showing input fields for "Artist" (with "artist name" as a placeholder), "Location" (with "location" as a placeholder), and "Technique" (with "technique" as a placeholder). A note below these fields says: "Only some collections support the advanced search fields. Other will use them as keywords." Below the search fields is a table with the following data:

Collection	Features	Dates	Locations	Hits
Art Database of the National Fund (Austria) National Fund of the Republic of Austria for Victims of National Socialism		1940 - 1944	<b>Austria</b>	

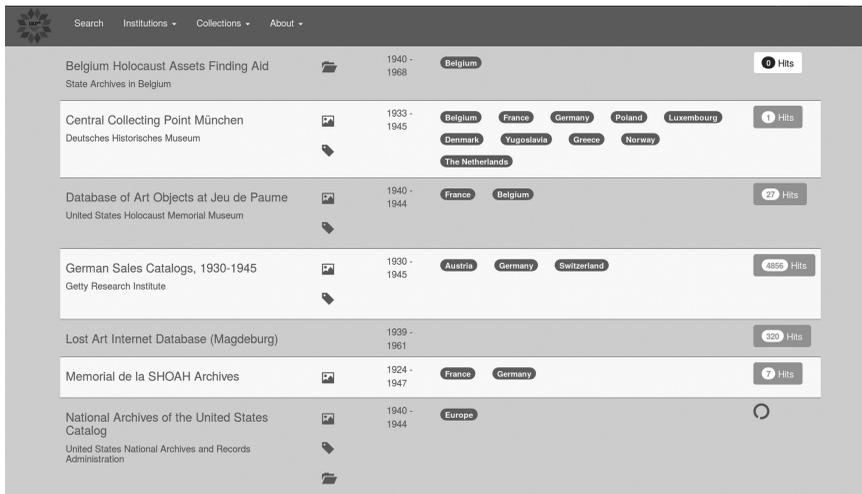
**Figure 13.3** Portal searches may include advanced fields and translation of terms.

collection to identify relevant and comparable provenance data to the greatest extent feasible. To increase the expertise available to the project team, an international advisory board of experts provided content and technical guidance. In addition, art provenance research experts provided informal feedback when technical demos were presented at various stages of the project.\*

On a parallel track, two MIM students, using the Portal as the basis for their required master's-level research projects, worked with the software developer to mock up a federation of catalogues to enable automated searches across the Portal's collections and to retrieve and consolidate information on individuals, organizations, and cultural objects related to Nazi-looted cultural assets. DCIC presented a proof of concept demonstrating a web-based federated search approach at an Advanced Art Provenance Research Workshop hosted by the Association of Art Museum Directors (AAMD) at the National Archives in Washington, DC, on November 3, 2015.

Building on the proof of concept, the project team has developed a federated search function for the Portal that provides tools for researchers with divergent needs, such as heirs and families of Holocaust victims searching for lost property,

\* Advisory Board members include the following: Association of Art Museum Directors: Anita Difanis; EHRI Project: Conny Kristel; King's College London: Tobias Blanke and Michael Bryant; National Archives and Records Administration: Chris Naylor; National Gallery of Art: Nancy Yeide; Smithsonian Institution: Jane Milosch; US Holocaust Memorial Museum: Michael Levy and Michael Haley Goldman; Yad Vashem: Haim Gertner.



**Figure 13.4** Search results are presented for various collections.

lawyers, investigators, provenance research experts, archivists, museum specialists, librarians, and interested members of the general public (Figure 13.4).

### *Technical Development*

Since the proof of concept, the software has matured into a more stable, more modular codebase, which provides the basis for additional student developers to work on the project and for enhanced server operations. As a federated search site, the Portal must perform many search tasks at once for a single user. It also must interface with a variety of different institution search sites that have different technical implementations, including various data formats, search fields, query operators, and more. Over time, the software has been redesigned to address these issues more easily, and the ability to add new search sites has been enhanced.

Major development phases, much of it student led, has pushed the Portal to enhanced levels of function and usability. These include the following:

- Federated search of participating institutions
  1. The ability to navigate, using a federated search strategy, across selected Portal collections with relevant keyword search terms to identify, consolidate, and present pertinent information.
  2. Local search implantation of a Belgian finding aid search, converted from the original PDF file into an XML-based search.
- Translation of query terms
  1. Description language for each collection was catalogued.

2. Options were added for translation of English search terms into French and German.
- Advanced search fields were
    1. Added to the Portal site for artist name, location, and year.
    2. Mapped accordingly onto participating search sites.
    3. Enabled to reuse an advanced search term as a keyword when participating search sites do not have an applicable field.
  - Autosuggestion of terms
    1. Authority lists were generated from the Getty Institutes' Union List of Artist Names, Cultural Objects Name Authority, and Thesaurus of Geographic Names.
    2. Authority lists were indexed with a Solr search engine.
    3. Advanced search fields were enhanced with "autosuggest as you type" terms.
    4. Autosuggest terms include hints that help disambiguate similar terms with relevant details, for instance, showing brief biographical information for an individual.
  - Personalization for researchers, who can
    1. Sign up and log in to the Portal site.
    2. Save useful search queries for later use.
    3. Review and manage saved searches.
  - Redesign of Portal web site
    1. A logo was designed and incorporated into the site.
    2. The site was migrated to the commonly used Bootstrap template system.
    3. The site was designed to adapt to mobile and desktop platforms.

## ***Continued Development and Implementation***

Students from the iSchool, other departments of the University of Maryland, and the Museum Studies program at Johns Hopkins University participated in testing and provided feedback for the federated search capability. Student feedback identified continuing challenges with linking search terms with relevant records. In response, the software developer redesigned the portal interface to emphasize the searching and the browsing of the extensive collection descriptions that make up the existing portal. The new design puts search and browse on equal footing and gives the user many more pathways into the browsing portion of the website (Figure 13.5).

The final phase of the project (begun in spring 2017) focuses on using structured data, drawn from the US Army property cards mentioned earlier to develop new research capabilities. Graph database and visualization technologies will be used to create linked relationships with people (looters, victims, and collectors), artists, works of art, dates, and locations of events. These relationships will help address the provenance issues that arise from heterogeneous collections with records and information melded from a variety of sources.

The screenshot shows a web interface with a dark header containing navigation links: Search, Institutions, Collections, and About. Below the header is a table of contents with the following entries:

- National Archives and Records Administration of the United States
- Bundesarchiv (Federal Archives of Germany)
- The National Archives of the United Kingdom
- France Diplomatie
- Central State Archives of Supreme Bodies of Power and Government of Ukraine (TSDAVO)
- Lost Art Internet Database (Magdeburg)
- State Archive in Belgium
- U.S. Holocaust Memorial Museum and Claims Conference
- Deutsches Historisches Museum
- Mémorial de la Shoah
- NICOD Institute for War, Holocaust, and Genocide Studies
- The Italian Directorate General of Archives
- Getty Research Institute
- Landesarchiv Berlin
- National Fund of the Republic of Austria for Victims of National Socialism and the General Settlement Fund for Victims of National Socialism

To the right of the table, there are detailed descriptions for three specific collections:

- Ardelia Hall Collection: Records Concerning the Central Collecting Points**: As the Nazis took over Europe in WWII, they looted everything from major artwork pieces from national museums to ceremonial and personal items from the Jewish people. The Ardelia Hall is a record collection of these recovered items and the work to restore these stolen items to their proper owners. This is one of the most popular collections at the National Archives and provides a different picture of the Holocaust that many may not know about.
- Miscellaneous Property Reports, 1945-1948**: Record Group 260; Descriptive Pamphlet #M1940 (6 vols). Description available in NARA's online catalog. Digital images available on FOLIO.
- OMGUS Headquarters Records, 1938-1951**: Record Group 260; Descriptive Pamphlet #M1941 (45 vols). Description available in NARA's online catalog. Digital images available on FOLIO.
- Offenbach Archival Depot, 1946-1951**: Record Group 260; Descriptive Pamphlet #M1942 (13 vols).

**Figure 13.5** Institution and collection descriptions presented with a table of contents.

Software development will continue to focus on the design of visual and graph-based discovery features to enable lateral browsing across linked topics and facts. Based on the property card graph database feature, researchers will be able to reconstruct the journeys of looted assets across Europe and beyond. In this phase, the project director and the software developer will engage a visualization design expert to assist in graphically illustrating the linked relationships, which are often not possible to identify in traditional, hierarchical archival finding aids. They expect new relationships to be revealed, leading to fresh insights from a research perspective.

A significant issue in implementing the federated search capability for the Portal was the need for ongoing stewardship to ensure that the upgrades and maintenance required to keep the Portal vital and usable are properly managed. The IRP2 project team and the NARA officials responsible for the stewardship of the Portal decided that the European Holocaust Research Infrastructure (EHRI), which includes 16 of the 18 Portal contributing institutions in Europe, would be the best repository for and ongoing steward of the redesigned Portal. The goals of EHRI, a European Union-funded project, are to identify the approximately 1800 repositories (mostly in Europe) that have Holocaust-related documentation and to catalogue the records with the ultimate aim of providing the greatest digital access possible. The redesigned Portal is at <http://irp2.ehri-project.eu>

### Student Learning

The combination of professional development support with student-led feature development is a strong model for interdisciplinary, collaborative student learning.

In the IRP2 project, students were free to experiment and prototype, receiving much-needed support from the professional software developer. Later, their efforts were consolidated, refactored, and merged into the evolving project. Project leaders carefully evaluated the abilities of team members so that all the students worked on features appropriate for their available time and skill levels.

Some features in any software development project will be beyond the skills of most graduate students outside computer science. In these cases, it is more efficient to draw on the expertise of a software developer. The balance of professional and student-led work ultimately depends upon the project's requirements. With the IRP2 project, the balance was 80/20 student to professional software development and metadata identification in the first year; in the second year, the breakdown was 60/40.

## Conclusion and Next Steps

The provenance challenges reflected in the Portal will be much more manageable with the full implementation of the prototype described earlier. Major gains in researcher access can be appropriately anticipated. However, overall major provenance issues remain in the management and exploitation of big data in archives, special collections in libraries, and in museums. The DCIC and the SPRI, along with other national and international research collaborators, intend to explore archival and museum provenance issues with the goal of integrating archival research data and user-contributed data, at scale, with cyberinfrastructure to generate new forms of analysis and engagement.

Specifically, the DCIC and SPRI will use Dras-tic (Digital Repository at Scale that Invites Computation) to improve collections software currently used to research, record and securely share big data, including provenance data. Dras-tic is an open-source community software project that evolved from a \$10.5 M National Science Foundation grant involving the University of Maryland and was developed in collaboration with Archival Analytics Solutions, Ltd., a software development firm in the United Kingdom.

The ability to accurately determine the provenance of archival and museum collections is at the heart of the Dras-tic enterprise: its goal is to build out a horizontally scalable archival framework to serve the academic, museum, archival, library, and scientific management communities as a credible solution for big data management in the cultural heritage community.

## Reference

1. Anderson, M. L. 2011. The crisis in art history: Ten problems, ten solutions. *Visual Resources: An International Journal of Documentation*, 27, no. 4: 336.