Thinking Outside the Box: Migrating the Frick’s Photoarchive Collection into Arcade
Lily Pregill
ARLIS 2012, Toronto

The Prezi presentation that accompanied this talk can be viewed at:

http://prezi.com/eh0cg5pugefi/thinking-outside-the-box-migrating-the-fricks-photoarchive-collection-into-arcade/

Good morning, I’ll be speaking today about the work involved with migrating the Frick Art Reference Library’s Photoarchive Collection into Arcade, the shared catalog for the New York Art Resources Consortium or NYARC. NYARC consists of the Frick Art Reference Library and the libraries and archives of the Brooklyn Museum and The Museum of Modern Art. I will discuss mapping the Photoarchive’s data from their legacy system to MARC, what the online collection looks like today, challenges that we faced along the way, and share some of our future plans.

While I have the pleasure of being with all of you here today, many people contributed to this project. I would like to specifically mention my two partners on the migration team: Kerry Sullivan, Head of Photoarchive Records, and Julie Shean, Database Administrator at the Frick. This project could not have been accomplished without either of their expertise and encyclopedic knowledge of the Photoarchive collection data.

Image: Ms. Frick

Let me start with a brief introduction of The Frick Art Reference Library. The Library was founded in 1920 by Helen Clay Frick, daughter of industrialist and art collector, Henry Clay Frick. The library’s purpose was to provide a publically accessible research collection for “serious students of art history”. Here we see Ms. Frick traveling in Europe in 1920 to further research her father’s collection, study museum management and survey the devastation of WWI. It is during that trip that Ms. Frick visited Sir Robert Witt’s Library of Reproductions consisting of photographs of works of art throughout Europe. The Witt, which I’m sure many of you know, is now part of the Courtauld Institute of Art. Ms. Frick, who was astonished by the extent and research value of the Witt collection, upon seeing it purportedly turned to Sir Robert Witt and asked, “May I copy-cat your library?” The seed for the Frick Art Reference Library was firmly planted.

Image: Bowling Alley

The library got its start as a collection of clippings and photographs in the bowling alley of the Frick residence, but quickly outgrew this space.

Ms. Frick’s goals were threefold:
1) To gather photographs of paintings, drawings, sculptures, and illuminated manuscripts by European and American artists;

2) To collate with each photograph the facts relating to the history and location of a work of art; and

3) To create a comprehensive research center, including both text and images, for the study of Western art.

**Image: What is the Photoarchive?**

Today, the Photoarchive is a collection of over 1 million study photographs and reproductions of works of Western art spanning from the 4th to the mid-20th centuries. The collection is organized in files representing approximately 40,000 artists. The majority of the images were acquired or purchased from domestic and foreign photographers and museums, often through the help of scholars or other art professionals. The Library also received gifts of images from art historians and dealers in Europe and United States. Complementing the images gathered from outside sources are approximately 57,000 large format black and white glass plate and acetate negatives from the Library’s photography campaigns carried out between 1922 and 1967 in the US, England, and Italy.

The Photoarchive continues to grow and new photographs are purchased from independent photographers and small museums whose collections may never be published. Additionally, thousands of photographs and digital images are received from donors every year. Many museums have donated to the collection, including the National Gallery of Art, MFA Houston, the Art Institute of Chicago, and the Metropolitan Museum of Art. In 2001, MoMA donated its entire photoarchive of more than 47,000 photographs and reproductions, representing over 5500 artists, to the Frick, greatly augmenting their holdings in modern art.

**Image: What can you discover?**

Here are a few examples from the Photoarchive Collection:

**Image: Rubens (King Philip IV of Spain)**

The collection contains many reproductions of works that have been lost, stolen or destroyed.
Versions and copies are also collected for researchers to trace stylistic developments and influences.

The library acquires multiple images of a single work. This example demonstrates changes to the work over time and shows three states of a painting. 1920, 1930, 1980, with the figure disappearing in the 1930 image and was later restored.

The multiple images of a single work may also include details and color reproductions. Additionally, there are images of works in private collections, small historical societies and museums that just are not reproduced anywhere else.

This is an example of mount. While the images themselves are a tremendous resource, the data accompanying each image offers researchers primary source material documenting the work.

The cataloging data describing the work of art includes:

- Information about the collection history
- Former attributions
- Exhibition record
- Condition history
- Biographical information about portrait sitters

And more.

This is an incredible data set! The cataloging is not merely descriptive, but each record is a piece of research documentation on the art object. And this documentation, gathered from scholars, oral history and published resources continues to be updated by the Library.
A card catalog in the Reading Room provides item-level access to slightly less than ½ the collection. The rest of the collection is represented by collection-level artist records. The card catalog remains an essential access point used by researchers.

**Image: STAR**

In 1996 the library began to develop a database to improve access to the collection with funds from the Henry Luce Foundation. At that time, the STAR database management system was chosen as the best solution. STAR contained variable length fields, repeatable field and subfields, linked lookups in the database and data entry controls.

At the same time the Book Department at the Frick was bringing up their first OPAC. The decision was made to not use the OPAC for the Photoarchive because of the limitations in indexing and in the additional fields needed to accommodate the Photoarchive data. There was also concern about integrating image and text results in a single search, with the fear that there would be many unwanted hits in the OPAC if the item-level work records were introduced, so a separate system was put into place.

**Image: Collection-level record in FRESCO**

In 1998, collection-level records began to be added to the Library’s OPAC, which we call FRESCO, and contributed to WorldCat. Here’s an example of the collection-level record for the most widely held Canadian artist in the collection.

Unfortunately, STAR did not have a public interface. It was always the intention to develop a web-based interface in-house to search the collection, but it just never happened in STAR.

In 2010, the department was considering a move from STAR to a FilemakerPro database that an interface could be built on top of, and at the same time a proposal to recon a portion of the American school images was being prepared. In thinking about a recon project, MARC was being considered as a data structure. I was brought into the conversation to talk about MARC mapping and by the end of the meeting with Julie and Kerry, we were talking about dumping the data into Arcade. We made the recommendation and we were on our way.

This may seem like an obvious move to some - sure why not integrate the image data with the text data? But this was a bold move by the Photoarchive department and I would say that was also true or the NYARC partners. When you say you’re “thinking outside the box”, you need to define that box locally – what may seem conservative to some, could be revolutionary to others. A migration is always difficult and a leap of faith even under the best circumstances. For us, this was a bold move. The Photoarchive department was giving up the autonomy of using their own system to join a consortium catalog with all of its complexities. They also needed to learn a new language (MARC) and completely transform their workflows.
For the NYARC partners, their catalog was going to grow enormously by 180,000 image records and 70,000 local authority records; although the Photoarchive was part of the Frick, it was like absorbing another library. But it made so much sense (we didn’t need another data silo; we could share expertise if we were all working in the same system, and we did it for a low cost, etc.). Additionally, everyone was excited to see how the collection would complement the existing Arcade data, and we were confident we could overcome any barriers that came our way.

**Image: Arcade**

So the decision was made, so what is Arcade? Arcade was launched in 2009 as the joint catalog for Brooklyn Museum, the Frick, and MoMA. Prior to the Photoarchive migration, the system contained almost 800,000 bibliographic records. The platform we use is Innovative Interface’s Millennium integrated library system. A search in Arcade returns results on the joint collection, and each library has their own scope, or subset, that can be searched against. As I mentioned, FRESCO is the name of the Frick’s scoped interface and the Photoarchive is a sub-scope of the Frick’s collection.

**Image: Project Phases**

Here are the project phases for the migration to give you a sense of what was involved. Planning started in January 2010, but the bulk of work happened January 2011-May 2011.

**Project preparation:** Project management activities, investigation into Millennium system options, research into data structures.

**Data preparation:** data cleanup, seized opportunities to bring the data in line with AACR2 and MARC.

**Millennium/Arcade training:** Photoarchive staff was new to MARC, new to Millennium, new to having an online interface.

**STAR programming:** data reformatting, data mapping, export procedures and routines written.

**STAR export:** routines to cross-load records, convert to MARC, tested using MarcEdit.

**Millennium import prep:** defined codes in system, created new scopes, expanded indexing rules, created 2 new authority indexes, created bib templates.

We also had NYARC committee discussions – we have both a User Group and a Technical Contacts group that meets regularly within the consortium and these groups reviewed and provided feedback on integrating the Photoarchive’s data into the system.

**Millennium web interface redesign:** edit bibliographic field labels, customize Photoarchive scope page, new help, new search scripts.
**Millennium load tasks:** write load profiles, create bib, item, authority templates, prep file and fix validation errors in MarcEdit, run load and indexing routines, load 6 files (3 authority, 2 works, and 1 negative), run diagnostics.

**Post load tasks:** run data cleanup routines, authorities processing, load patches

**Image: Data mapping**

When the Photoarchive moved to STAR, existing and emerging standards in the Visual Resource community were examined. Fields that were used were developed from in-house standards in relation to the core record structure as defined in the Categories for the Description of Works of Art, a standard developed by the Art Info Task Force, sponsored by CAA and the Getty. VRA core version 1 was also used as a primary guide for defining fields after it was released in 1996. The Inventory of American Paintings at the National Museum of American Art was looked at as the most important existing example in place at that time, and documentation was shared to help guide the Photoarchive’s work. We had an excellent metadata foundation for this work.

We did the same examination when we embarked on this migration, and consulted with our colleagues. The National Gallery of Art Image Collections generously shared their annotated data maps and cataloging manual with us. Also colleagues from the Morgan Library provided valuable advice and feedback on initial mappings.

I must say that in migrating data we used the “no data left behind approach”; we wanted to carry over all the data and needed to be creative to provide each piece of data a home in MARC.

**Image: Mrs. Daniel Denison Rogers**

Let’s take a look at some data through an example. Meet Mrs. Daniel Dennison Rogers. This painting is part of the University of North Carolina’s Ackland Art Museum.

**Image: Mrs. Daniel Denison Rogers - STAR Screen 1 through Screen 5**

Here is the record for this artwork as it appeared in STAR:

**Name:** Here we learn that this woman has a name of her own – Elizabeth Bromfield

Take a look at the **classification:** Local iconographic subject headings are used in the Photoarchive collection, so we needed to accommodate local subjects.

**Provenance** – extensive provenance note.
Subject/Biography: Here we learn a bit about the portrait sitter.

Sources: Everything is sourced; all the research is cited.

This is a tremendous amount of information, and, as I said, we didn’t want to lose any of it by bringing it over into the MARC environment.

Image: Mrs. Daniel Denison Rogers Record in Millennium

And this is the record as it appears today in Millennium.

Image: 590s Detail

While we mapped much of the data to MARC, not everything fit nicely. We ended up defining a series of 590 fields using indicators to refine their meaning. This approach allowed us to assign custom field labels in the OPAC based on the indicator and didn’t exhaust our 590s for future use by other consortium partners. Here you can see examples of 590s and what didn’t fit nicely into existing MARC fields:

5901 Current Repository
5902 Sourced Main Information
5904 Sourced Color Notes
5907 Sourced Subject Information
5909 Reproductions

Additionally defined 590s not in use in this record include:

5903 Sourced Condition History
5905 Sourced Attribution History
5908 Sourced Pendants/Related

Image: Mrs. Daniel Denison Rogers Record – OPAC View

We also have local 650s, 790 fields for local name forms, and 9xx fields for local/internal info.
Along with the work or image records, we also mapped and migrated 70,000 local authority records. The Photoarchive maintains a local authority file for artist names. Sometime they agree with the Library of Congress heading and sometimes there are differences. Here’s an example. While we didn’t employ RDA in our bibliographic records, we made use of some of the new RDA fields in our authority mapping.

046 - Special Coded Dates (R) (birth, death, start, end)
372 - Field of Activity (R) Scope and date
373 - Associated Group (R) National School

Once the data was mapped and loaded into Arcade, we continued to work to refine the interface.

This is the Photoarchive’s scoped page in FRESCO. There are four primary indexes here. Note that the Artist/Collection index is a combined index of Library of Congress and local authority name forms.

Also note that the 180,000 records only represent 10% of the works in the collection!

We use material icons in the catalog and we worked to make it clear to users what was being described. For the collection, we use an archival box and an image, and for the mount, we use a single image.

As in 1996, there was some concern about dumping 180,000 image records into Arcade and whether they would confuse users and overwhelm search results. We discussed whether the image records should only appear in the Photoarchive scope and not be included in a general Arcade search. But we wanted users to discover these records, especially those researchers who might not realize the Photoarchive collection contained materials that could assist their research. Spontaneous discovery was what we were hoping for, so we ultimately decided to include them in a general Arcade search, but created an exit strategy for excluding them if we really did have a problem with confusing or overwhelming search results.

Actually, what we’ve found is that the data complements the book collection in really fascinating ways:
This is a truncated result set for a search on the 1931 Stroganoff sale at Lepke’s Auction house in Berlin.

**Image: Head of Voltaire**

The results offer the auction catalog record for that particular sale, which is held in the Frick’s Book Collection, along with the image records for objects sold there, contained in the Photoarchive Collection.

**Image: Head of Voltaire – Frick Digital Image Archive**

Another feature that I wanted to point out is the integration of 10,000 links to the Frick’s Digital Image Archive to connect users directly to the image when possible. We continue to digitize images from the collection and links will continue to be added to Arcade. Here you can see the record for this image in the Digital Image Archive and the reciprocal linking to the complete record in Arcade.

**Image: Young Woman with Black Collar Record**

In this example you can see that the Photoarchive data is continually updated as we learn new information about the art objects represented in our collection. This is from the same 1931 Lepke sale and we learn that this painting was looted by the Nazis in 1940 and restituted to the family in 2006. The Photoarchive records are chock full of these stories, and is an amazing tool for provenance research.

**Image: Challenges**

- Migrated into a live system; all configurations and load happened while live and supporting the work of three active libraries.

- A parent/child record structure would have been a better model for the metadata, but unavailable in the Millennium system.

- Indexing constraints – We wanted a provenance index, but a limit of the system is that we couldn’t build a keyword index to search just the 561 field.

- Local authorities; exit strategy – account for all options – 3 authority files were created to cover all bases: Library of Congress file, local Photoarchive file, and combined Library of Congress and Photoarchive file; we have since deleted the local only authority file.

- Mapping - We needed to get creative with 590 fields and 9xx fields
**Image: Future**

**XINET:** Streamline sharing data with Xinet, the Frick’s Digital Asset Management System or DAMS, where the digital images live. We would also like to dynamically add thumbnails from Xinet into Arcade.

**ARTstor:** We have embedded links to approximately 10,000 digital images available in the Digital Image Archive into records in Arcade, but we would also like to link to the 25,000 images we have contributed to ARTstor. Ideally, we would like to do this through an API and have approached ARTstor about this possibility.

**WorldCat:** While the collection-level Photoarchive records have been contributed to WorldCat, we have not yet contributed the item level catalog records to WorldCat. We are investigating a workflow to make that happen.

**Image: Google Result**

**Google:** We have been engaged in a pilot project to expose this data to Google — here’s one example. The sitemapping and display of the data in Google is not perfect and we are still learning how to improve on this, but the results that can be found in Google are making new users, primarily genealogy researchers, aware of the collection.

And, we hope to collaborate and share data with our colleagues. We were recently contacted by the Rijksmuseum, who are now thinking about moving their image documentation into their Koha catalog.

**Image: Photoarchive Box**

Thank for your time this morning, and I encourage all of us to keep thinking outside the box.