



Old MAID/New MAID

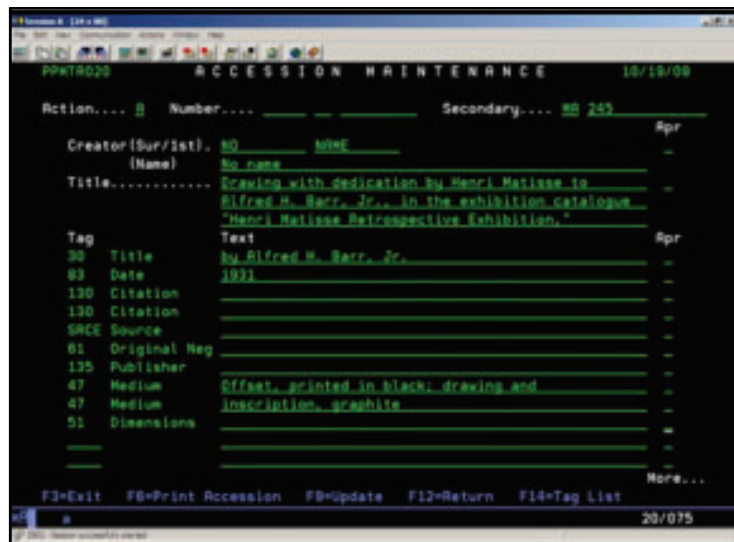
Migrating MoMA's Archive Image Database



Jonathan Lill

MoMA Archives

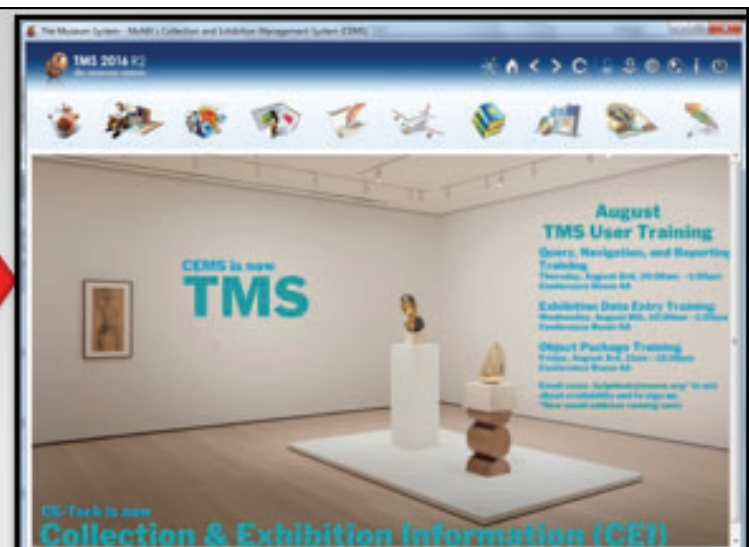
Good afternoon. It's a pleasure to talk to you today and to have the opportunity to tell you about the first time I was responsible for something many of you are familiar with, a mass metadata migration. And the focus of this story is how in a very short time frame we significantly enriched our data in several critical ways as it was moved from one system to another.



- IBM iSeries server software
- Backend database begun prior to 2004
- The name **Museum Archives Image Database (MAID)** created in 2005
- Contains records only for high res scans and born-digital installation photographs
- All physical installation images (~15,000) scanned and added in 2008

Jonathan Lill

MoMA Archives



- The Museum System
- Used by MoMA since ~2004
- Includes modules for objects, constituents, exhibitions, loans, and other functions
- Contains approximately 217,500 object records and 53,000 constituent records

2

Specifically, after years of delay, the MoMA Archives decided to migrate its image catalogue from a hopelessly antiquated system into The Museum System, the same place it takes care of all the Museum's artworks. We call it MAID, and today it contains the metadata for all the high resolution digital images in the Archives. Of 44,000 records at the time of migration, 33,000 were installation photographs, one of MoMA Archives key historical assets. The other 10,000 or so records were for other photographs, archival documents, and miscellaneous items from across the Archives that were digitized for publications or other use by request of staff and researchers. While the migration had been years in the making, my role in the work lasted only about two months.

Why we did it:

- The system was so old! Data was flat and inflexible.
- The Museum's IT department was desperate to get rid of it.
- The stand-alone image database was far away from where curatorial staff were looking for visual materials.
- Up to date tools for authority control and classification are necessary for future online exploitation of these resources.

The Museum System:

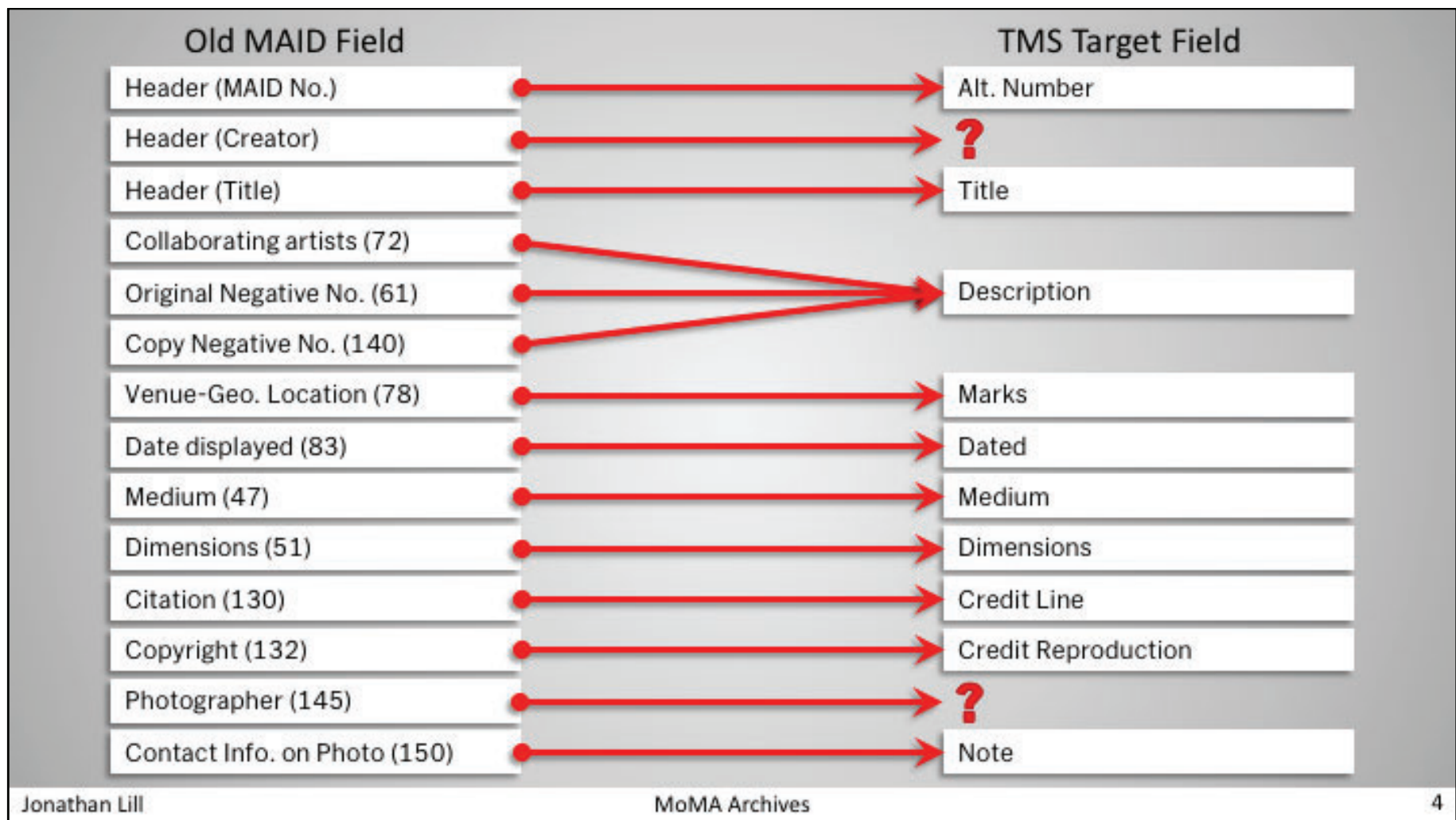
- Makes our materials vastly more visible to staff
- Enables better integration with MoMA's loan practices
- Makes data more accessible by MoMA's web team for use online
- *The Archives wouldn't have to pay for it.*

For the sake of time I'm going to pass over the various motivations for this move though you can see some here.

The basic migration workflow was that our IT department would extract the data from the legacy software and hand it to me, I would do what needed to be done to get it ready for the new system, and MoMA's TMS team would import it into The Museum System.

The old database was a black box to me and to our IT department. For a while IT wasn't sure they could export any data out of the system in a coherent way. But to everyone's relief, with some work, they were able to produce a single spreadsheet with one image record per line.

Batch importing records into TMS isn't that easy. But a while ago TMS's parent company Gallery Systems had created a custom script for MoMA that essentially allowed a one click upload from an excel spreadsheet.



So when our IT department handed over a 44,000 line spreadsheet the first and easiest task was to re-shape the data to fit into the import spreadsheet. This mainly involved renaming some of these spreadsheet columns and merging a few different fields into one.

The more important and complex question was how quickly and to what extent could we enrich these records, adding metadata that would be both valuable to our users and helpful to us in integrating these resources into the larger data environment. What fields were necessary to add, what were possible within this one-click uploader script, and what had to be considered off the table?

Basic Metadata Improvements

More rational object numbers

Range of mystery abbreviations

IN636.43



Reduced to two types of numbers

IN636.43

MA1407



ARCH.2008

PA598



ARCH.3287

AP40



ARCH.5583

EX106.4.17



ARCH.2133.4.17

MS138



ARCH.5232

Addition of Classifications/Object Types

Classification

Occurrences in MAID

Photograph

36158

• Document

647

• Correspondence

364

• Publication

343

Ephemera

163

• News Clipping

86

• Notebook

79

Work on Paper

53

• Poster

25

Textile

6

• = New term established for the Archives

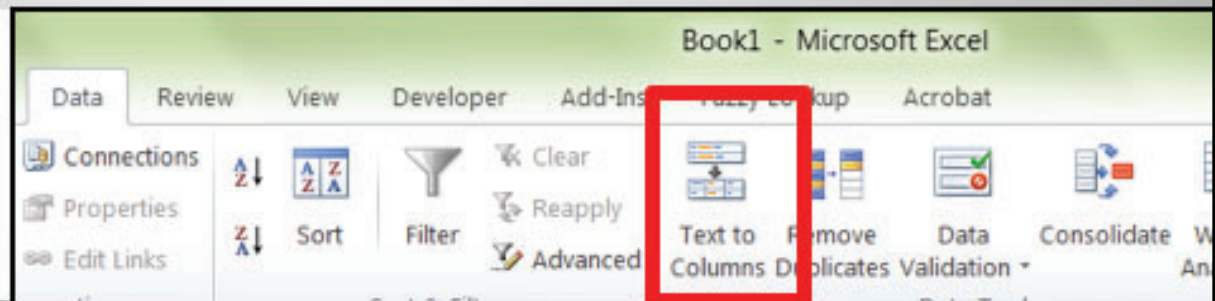
The uploader allowed for one unique identifier or “object number” but it also allowed for one “alternate number”. So we took advantage of this by retiring our current object number to the “alternate number field” and inventing a new slightly more rational system of object numbers in its place.

Next, TMS relies on a classification field to very basically categorize objects. We never had this capability before, it’s very, and the uploader allowed us to add it. MoMA uses a fairly limited list of classification terms and adding or changing any causes a lot of work as it affects numerous reports and functions downstream. So we couldn’t go wild in choosing new classification terms. Ultimately we negotiated six new terms as well as using a few already present in the system.

Medium Metadata Improvements

Out of 44,000 records:
5000+ had blank date values
33,000 were values patterned like this:

	A	B
1	May 26, 1972 through September 11, 1972	
2		



Presto!
Normalized dates!

	A	B	C
1	5/26/1972	through	9/11/1972
2			

Jonathan Lill

MoMA Archives

6

The next major metadata enhancement involved dates. It was late breaking news to us that the data importer allowed for standardized date/time fields in addition to a free text display date field. When I heard this I think I jumped out of my chair. We had never been able to search or sort on dates in our old image system, only by keyword. But this meant we had to convert forty thousand display dates into normalized start and end dates. Let me tell you, this isn't as hard as it looks but requires somewhat advanced Excel skills, or in my case a lot of googling "how do I do that?" Briefly put, the key to this kind of activity is to recognize the patterns the specific data falls into, use that pattern to divide your dataset into smaller manageable groups, and tackle each group one by one. For best throughput I tend to start with the easiest largest groups first and save the most unique and tedious issues for last. For example, out of 44,000 records; there are 5000+ records where the value is blank. 75% of the rest feature two date phrases separated by the word through. We can convert them using one transformation. And before you know it, every record has a normalized date. Except, y'know, for everywhere a month was misspelled or there was some other abnormal permutation.

Advanced Metadata Improvements

MA2386.1		Correspondence	
Title	Letter from Alfrd H. Barr, Jr. to Daniel-Henry Kahnweiler regarding Angry Arts petition to American artists requesting Pablo Picasso remove "Guernica" (1937, May-early June) from the Museum in protest of Vietnam war	Constituent1	Alfred H. Barr, Jr.
		Role1	Correspondent
		Constituent2	Daniel-Henry Kahnweiler, Pa
		Role2	Correspondent
Photographer		<input checked="" type="checkbox"/> More Co Add?	
MA2902		Work on Paper	
Title	Sketch with instructions from Diane Arbus to John Szarkowski regarding gallery title wall for the exhibition, "New Documents"	Constituent1	Diane Arbus
		Role1	Artist
		Constituent2	John Szarkowski
		Role2	Correspondent
Photographer		<input type="checkbox"/> More Co Add?	
MA3022.2		Work on Paper	
Title	Untitled collage (verso) from Ray Johnson to Robert Rauschenberg, verso marked in colored pencil: "Bob Raushenberg 278 Pearl St. nyc, ny."	Constituent1	Ray Johnson
		Role1	Artist
		<input type="checkbox"/> More Co	
Jonathan Lill			
MoMA Archives			
7			

Most importantly, the TMS uploader allowed us to add two name entries to each record. Records in the old system were very flat and there was only one uncontrolled field for "artist" that was barely used and another for photographer. By adding and populating these additional name fields to a line in our spreadsheet, the upload script would connect that object record to existing authority records in TMS. This was a more laborious process than our other actions, but it was also of more critical importance for connecting our archival records to all other objects in the system through related names.

I'm a big proponent of Microsoft Access as a tool and it enables this type of complex data entry to be simplified and streamlined. I imported the Excel spreadsheet of MAID records into Access. Then I had the TMS admin send me a spreadsheet of all constituents name records in the system and imported that also into Access. In Access I could build a simple custom interface where each MAID record could be more simply displayed. The new added name fields are on the right.

Advanced Metadata Improvements

Advanced Metadata Improvements			
Pane	MA2386.1	Correspondence	
	Title	Letter from Alfrd H. Barr, Jr. to Daniel-Henry Kahnweiler regarding Angry Arts petition to American artists requesting Pablo Picasso remove "Guernica" (1937, May-early June) from the Museum in protest of Vietnam war	<div>Constituent1</div> <div>Role1</div> <div>Constituent2</div> <div>Role2</div>
	Photographer		<input checked="" type="checkbox"/> More Co Add?
	MA2902	Work on Paper	
	Title	Sketch with instructions from Diane Arbus to John Szarkowski regarding gallery title wall for the exhibition, "New Documents"	<div>Constituent1</div> <div>Role1</div> <div>Constituent2</div> <div>Role2</div>
	Photographer		<input type="checkbox"/> More Co Add?
	MA3022.2	Work on Paper	
	Title	Untitled collage (verso) from Ray Johnson to Robert Rauschenberg, verso marked in colored pencil: "Bob Raushenberg 278 Pearl St. nyc, ny."	<div>Constituent1</div> <div>Role1</div>
			<input type="checkbox"/> More Co

- Alfred H. Barr, Jr.
- Mark Baron
- Ronald Baron
- Samuel Baron
- Baron/Boisanté Editions, New
- Barone Gallery
- Daniele Baroni
- Baroni e Marangon
- Sergio Baroni
- Vittore Baroni
- Jeane Baronin von Oppenheim
- Baroody Music (BMI), Santa M
- Martin Barooshian
- Joan and Richard Barovick
- Alfred H. Barr, Jr.
- Clay Barr
- Elsa Barr
- Artist

The entire list of 53,000 existing constituent names populated these drop down menus.

Advanced Metadata Improvements

Object ID	Category	Title	Photographer	Constituent1	Role1	Constituent2	Role2	More Info
MA2386.1	Correspondence	Letter from Alfrd H. Barr, Jr. to Daniel-Henry Kahnweiler regarding Angry Arts petition to American artists requesting Pablo Picasso remove "Guernica" (1937, May-early June) from the Museum in protest of Vietnam war		Alfred H. Barr, Jr.	Correspondent			<input checked="" type="checkbox"/> More Info Add?
MA2902	Work on Paper	Sketch with instructions from Diane Arbus to John Szarkowski regarding gallery title wall for the exhibition, "New Documents"						<input type="checkbox"/> More Info Add?
MA3022.2	Work on Paper	Untitled collage (verso) from Ray Johnson to Robert Rauschenberg, verso marked in colored pencil: "Bob Raushenberg 278 Pearl St. nyc, ny."						<input type="checkbox"/> More Info Add?

Jonathan Lill

MoMA Archives

9


Each name was assigned a relator term from a controlled list.

Advanced Metadata Improvements

MA#	Type	Title	Photographer	Constituent1	Role1	Constituent2	Role2	More Co Add?
MA2386.1	Correspondence	Letter from Alfrd H. Barr, Jr. to Daniel-Henry Kahnweiler regarding Angry Arts petition to American artists requesting Pablo Picasso remove "Guernica" (1937, May-early June) from the Museum in protest of Vietnam war		Alfred H. Barr, Jr.	Correspondent			<input checked="" type="checkbox"/>
MA2902	Work on Paper	Sketch with instructions from Diane Arbus to John Szarkowski regarding gallery title wall for the exhibition, "New Documents"						<input type="checkbox"/>
MA3022.2	Work on Paper	Untitled collage (verso) from Ray Johnson to Robert Rauschenberg, verso marked in colored pencil: "Ray Rauschenberg 278 Pearl St. nyc, ny."						<input type="checkbox"/>

Simple instructions: *If names are here* *Then add them here*
If the names aren't present in the drop-down menu then ignore

Most of this work was done by an intern so instructions were simple. Pick two names from the description, always include the photographer if there is one, and select them from the drop down menu. Describe them with the appropriate relator term. If no individual names are mentioned, move to the next record. If the names aren't already in the drop down menu, move to the next record. The emphasis was on speed and results while avoiding getting bogged down in problematic edge cases. Ultimately through this process and some mass updating of installation photo data, we added nearly 50,000 name authority connections to these records. 95% of all the records in the migration have at least one name link. This work took about four weeks. Once all this work was done it was uploaded to the new system by our TMS team.



ALEXANDER CALDER photo: James Soby, summer 1957, Long Island.

ID	Object ID	Format	Type	Name	Life Dates	Description
1	ARCH 6668	ARCH	Photograph	Alexander Calder	American, 1898-1976	Photograph of Alexander
1	ARCH 6669	ARCH	Photograph	Hugo Herdeg	Swiss, 1909-1953	Photograph of Alexander
1	ARCH 6670	ARCH	Photograph	Fred Stein	American, 1909-1967	Photograph of Alexander
1	ARCH 6671	ARCH	Photograph	Herbert Matter	American, born	Photograph of Alexander
1	ARCH 6672	ARCH	Photograph	Alexander Calder	American, 1898-1976	Photograph of Alexander
1	ARCH 6673	ARCH	Photograph	Herbert Matter	American, born	Photograph of Alexander
1	ARCH 6674	ARCH	Photograph	Alexander Calder	American, 1898-1976	Photograph of Alexander
1	ARCH 6675	ARCH	Photograph	Alexander Calder	American, 1898-1976	Photograph of Alexander
1	ARCH 6676	ARCH	Photograph	James Thrall Soby	American, 1906-1979	Photograph of Alexander
0	EL 1936.61 a-g	P&S	Sculpture	Alexander Calder	American, 1898-1976	Mobile
1	EL 1945.22	P&S	Sculpture	Alexander Calder	American, 1898-1976	Figurine of a Dancer
1	EL 1945.23	P&S	Sculpture	Alexander Calder	American, 1898-1976	Figurine of a Dancer
1	P75	A&D	Design	Alexander Calder	American, 1898-1976	Jean Vilar presenta Nuclea
0	SC42.1952	DRAW	Drawing	Alexander Calder	American, 1898-1976	Pattern for a Stable
10	SC78.1958.1-9	P&S	Sculpture	Alexander Calder	American, 1898-1976	Candelabrum
2	SC133.1957	DRAW	Drawing	Alexander Calder	American, 1898-1976	Sheep
1	SC436.1994	P&S	Sculpture	Alexander Calder	American, 1898-1976	Untitled
1	SC1121.1964	P&S	Sculpture	Alexander Calder	American, 1898-1976	For 3 Years of Fairly Good
1	SC1574.1968	DRAW	Drawing	Alexander Calder	American, 1898-1976	Curt Valentin

Photograph of Alexander Calder
Photographer: James Thrall Soby

1957

Role Type
Object Related

Role	Name	Date Range	Displayed	Active
1 Photographer	James Thrall Soby, American, 1906-1979		<input checked="" type="checkbox"/>	
2 Subject	Alexander Calder, American, 1898-1976		<input checked="" type="checkbox"/>	

Add Delete Move Up Move Down

Jonathan Lill

MoMA Archives

11

What was the result? Firstly, I can't overemphasize what a big step it was for us to have our image records in a modern system.

Secondly, we can for the first time search by format or within a date range. More importantly the materials are fully integrated with artworks and other collection objects across the museum. These are search results for any object with Calder as a related name. In the highlighted record he's only the subject of the photograph. But you can see in the search results how works from the Archives appear next to sculptures, drawings and items from other parts of the Museum.

There is of course much work that can still be done. More names to add, legacy errors and omissions to correct, and finding new ways we can exploit the new system we're in.

Photograph of Alexander Calder

Title:	Photograph of Alexander Calder
Dated:	1957
Venue/Location:	Long Island
Person(s):	James Thrall Soby (Photographer) Alexander Calder (Subject)
Classification:	Photograph
Medium:	Gelatin Silver Print
Dimensions:	9 7/8 x 8 1/16" (25.1 x 20.5 cm)
Description:	
Object Number:	ARCH.6676
Credit Line:	Photographic Archive, Artists and Personalities. The Museum of Modern Art Archives, New York
Copyright:	Must be reviewed by Archives before publication.
Contact Information on	Must be reviewed by Archives before publication.



To request this image for publication or other uses, please contact [Art Resource](#) (for publication in North America) or [Scala](#) (for publication in all other geographic locations). All questions or comments.

I haven't today talked at all about front ends. But let me mention that months after our data migration, MoMA's TMS team were able to customize an off the shelf TMS web access module to create a new user interface that, for so that for the first time all of our digital images are freely accessible on the web.



jonathan_lill@moma.org

moma.org/archives
moma.org/history
maid.moma.org



Jonathan Lill

MoMA Archives

13

Thank you.