

# Online Presence and Participation

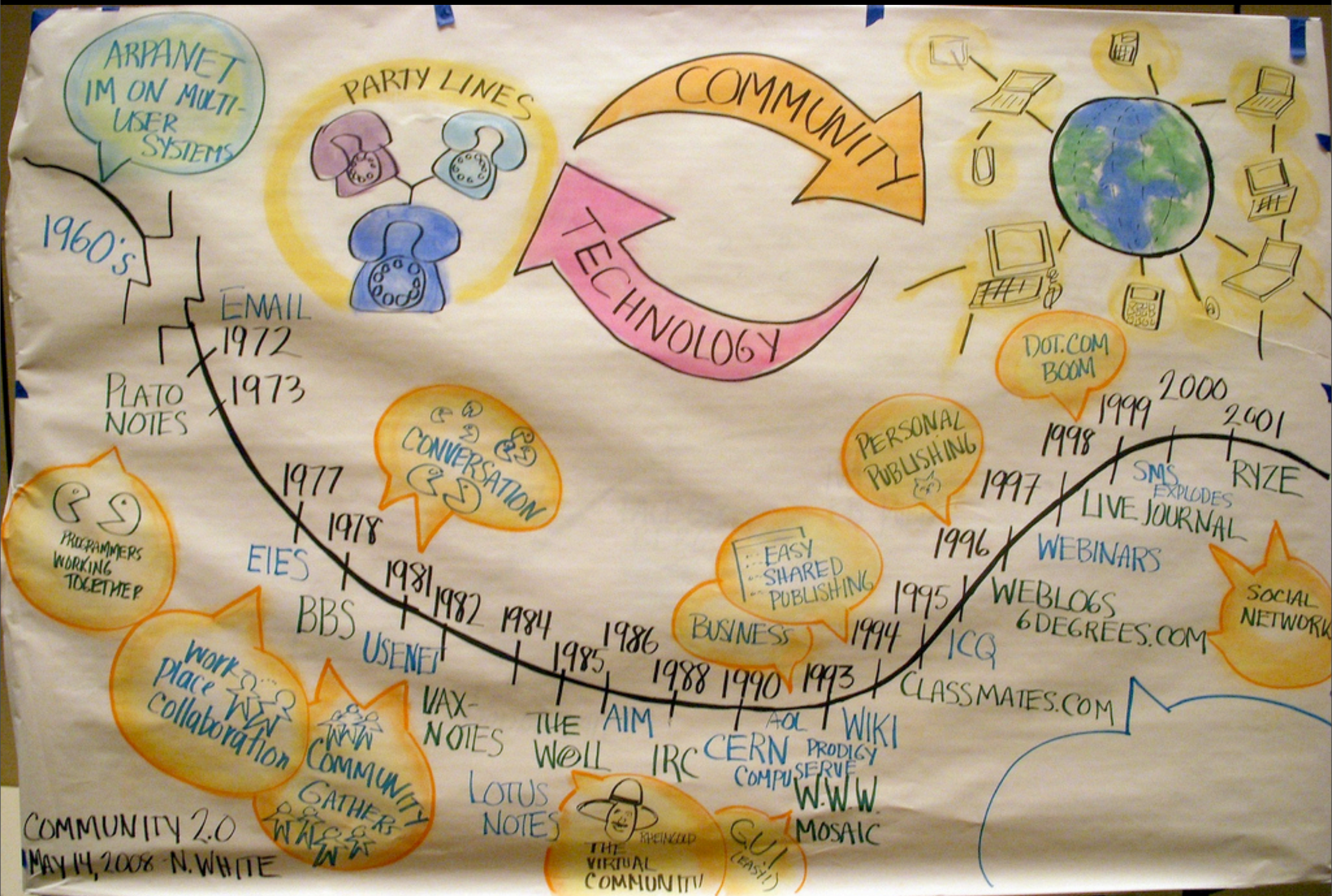
Mark A. Matienzo  
The New York Public Library

Society of American Archivists Annual Meeting  
Session 101 – August 13, 2009

# Disclaimer

The following presentation, while factual, expresses opinions of my own and not of my employer, my coworkers, etc.















# UNCERTAIN INNS & FLEA MARKET

www.uncertaininn.com www.uncertainfleamarket.com www.uncertaingifts.com

Where The Hell Is Uncertain Texas ???

**I WAS HERE**  
**903-789-3292**





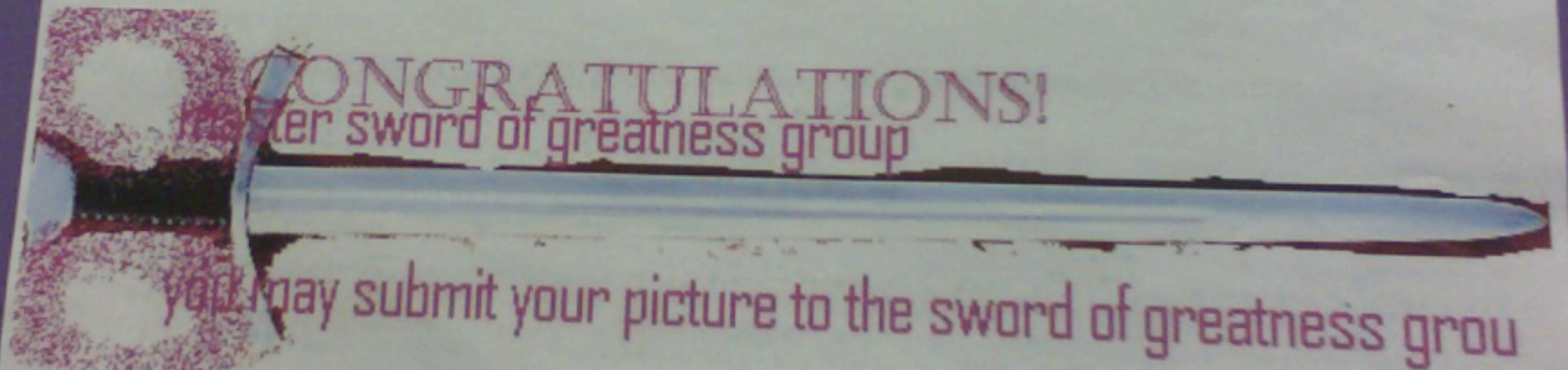


<http://www.flickr.com/photos/gorgeoux/2403605412/>









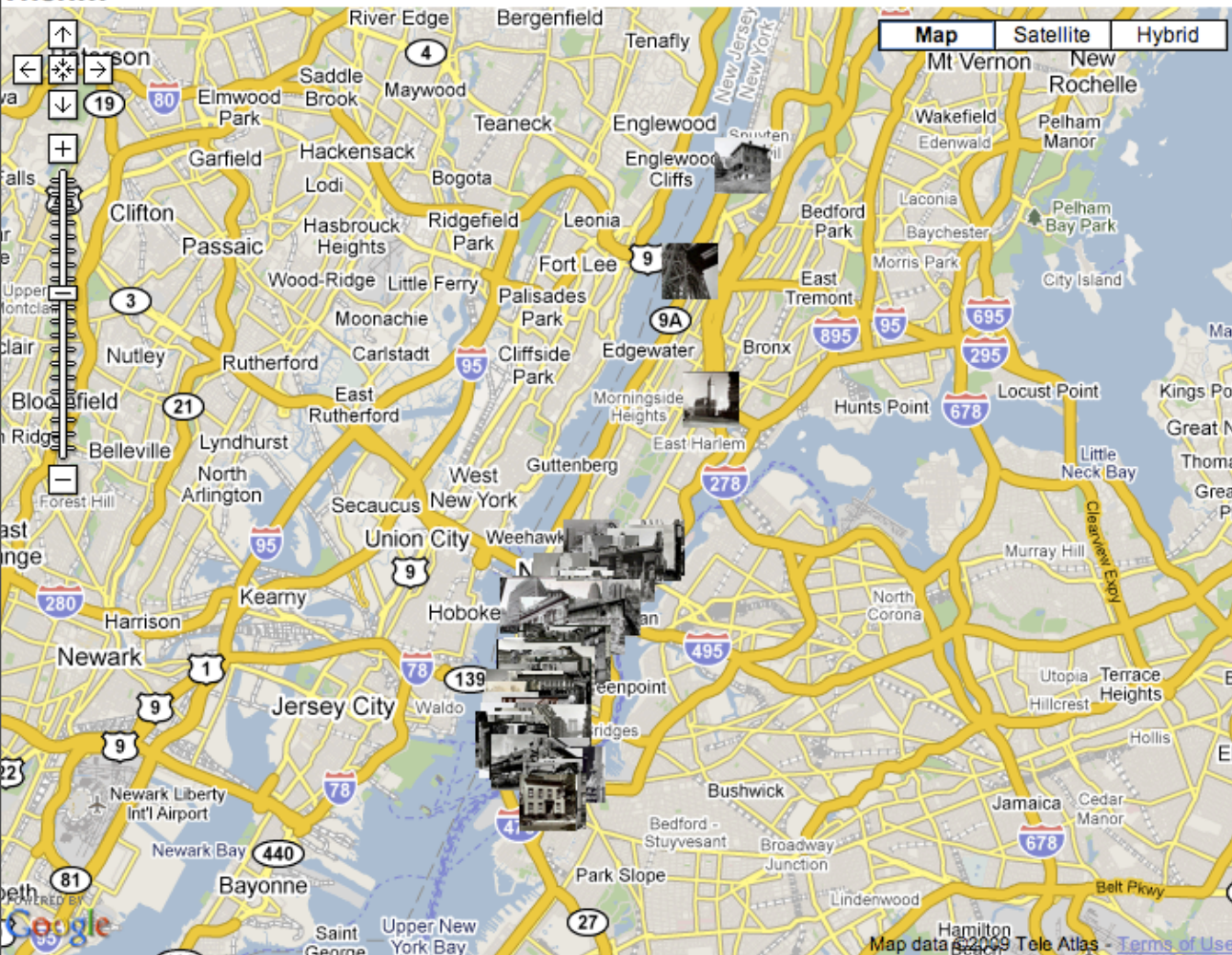


## Flickr and Google street view mashup



### New York Public Library

Then...



This application displays historic images from [Flickr Commons](#) and compares them with their respective modern day Google street view locations.

& now

Please select an image on the map





## Changing New York, 1935-1938

[Thumbnails](#) [Detail](#) [7 comments](#)

Photographer Berenice Abbott (1898-1991) proposed *Changing New York*, her grand project to document New York City, to the Federal Art Project (FAP) in 1935. The FAP was a Depression-era government program for unemployed artists and workers in related fields such as advertising, graphic design, illustration, photofinishing, and publishing.

Abbott's efforts resulted in a book in 1939, in advance of the World's Fair in Flushing



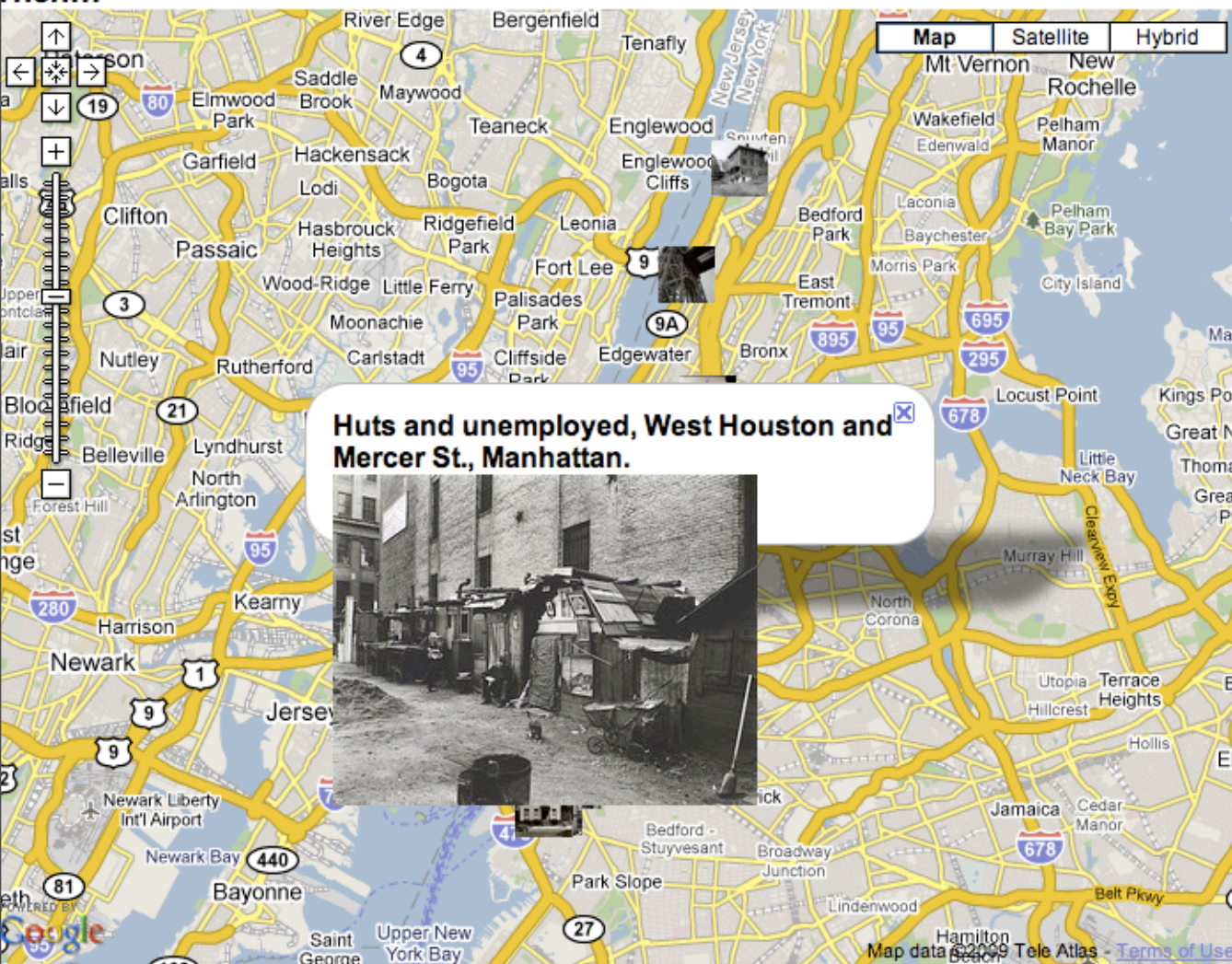


## Flickr and Google street view mashup



New York Public Library

Then...



This application displays historic images from [Flickr Commons](https://www.flickr.com/photos/nypl/) and compares them with their respective modern day Google street view locations.

& now





# cursive buildings

by joshua heineman

July 14



Reaching for the Out of Reach 57:

A party game of croquet, Georgia, circa 1885.

[ [more](#) from this [project](#) ([nypl](#) [permalink](#)) ]

May 23



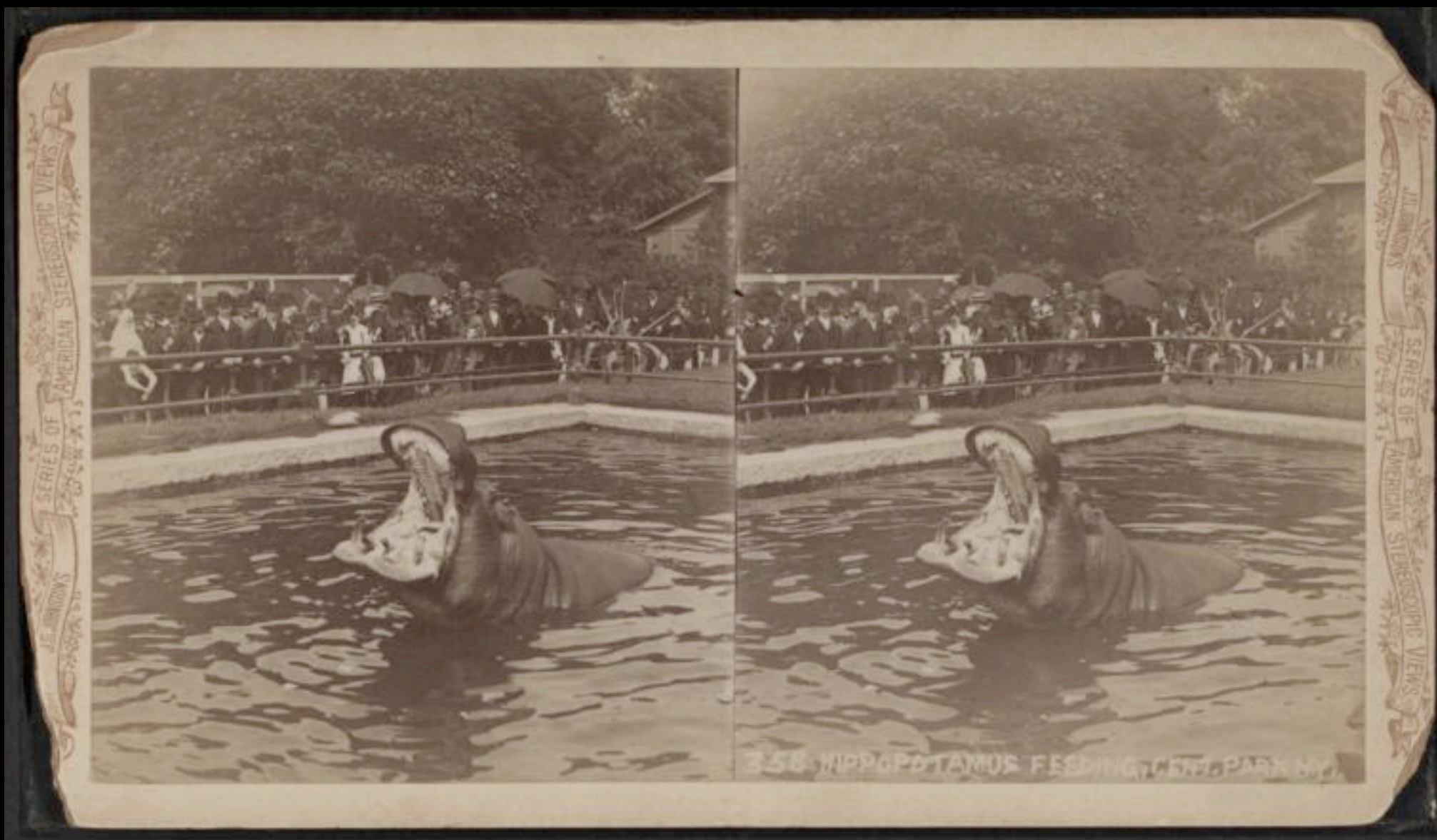
[ [subscribe](#) ]  
[ [frontpage](#) ]



*"deeply into whatever"*

**ahhhhhmega-zine**  
current issue - [no. 5](#)  
[no. 4](#), [no. 3](#), [no. 2](#), [no. 1](#)





















WIKIMEDIA  
COMMONS

Search

Go

Search

Navigation

- [Main Page](#)
- [Welcome](#)
- [Community portal](#)
- [Village pump](#)

Participate

- [Upload file](#)
- [Recent changes](#)
- [Latest files](#)
- [Random file](#)
- [Help](#)
- [Contact us](#)
- [Donate](#)

Toolbox

- [What links here](#)
- [Related changes](#)
- [Special pages](#)
- [Printable version](#)
- [Permanent link](#)
- [Subpages](#)

[Log in / create account](#)

[project page](#)

[discussion](#)

[edit](#)

[history](#)

# Commons:Batch uploading/NYPL Digital Gallery

[< Commons:Batch uploading](#)

## Images from NYPL Digital Gallery

[\[edit\]](#)

Will be great if we batch upload PD-images from NYPL Digital Gallery - <http://digitalgallery.nypl.org/nypldigital/index.cfm> *NYPL Digital Gallery provides free and open access to over 685,000 images digitized from the The New York Public Library's vast collections, including illuminated manuscripts, historical maps, vintage posters, rare prints, photographs and more.* --[Butko](#) ([talk](#)) 14:45, 14 April 2009 (UTC)

- This collection turned out to be more promising than I supposed. They use LizardTech ContentServer to serve up their images, whose API is [described here](#) . Here's how you extract original TIFFs at full size: first use a "browse" query to obtain some XML including the image dimensions, like this one [\[1\]](#) . The folder name and image name can be obtained from URL of the zoom view. Then, use a getimage query like this one [\[2\]](#) to get the full size TIFF, specifying the dimensions from the previous query. Tada. Close examination shows no artifacts in the TIFF - these are original scans (internally, they are SID images). The first one I extracted was 3845 × 4947, about 60 MB as a TIFF, and 27 MB as a PNG (which you can preview [here](#) ). They throttle you at 80 KB/s per transfer, but they *do* allow simultaneous transfers; any way you look at it though it would take a long time to fetch all the images we need. In light of the long download time per image, we're going to want to license filter *before* downloading. [Dcoetzee](#) ([talk](#)) 07:00, 15 April 2009 (UTC)
- Update: their complete collection of high-resolution images is browsable [here](#) . This can be used to easily obtain a list of folder-name pairs. I'll presently begin downloading. [Dcoetzee](#) ([talk](#)) 06:23, 16 April 2009 (UTC)
- Update: a better way to download these is to use the "getfile" function to get the raw .sid files, which are highly compressed (as in [\[3\]](#) ) and then use LizardTech's command-line decoder to convert to TIFF ([\[4\]](#) ). This is a quicker download and doesn't even require the dimensions. [Dcoetzee](#) ([talk](#)) 22:12, 16 April 2009 (UTC)
- I'm still in the middle of grabbing these. Enumerating IDs turned out to be trickier than I thought, because the folders are so large the browse interface times out on them. I ended up enumerating them instead using wildcard searches on single letters. Even just looking at the high res images, it's a *lot* of data. All told we're talking at least 100 GB in PNGs, and I'm pretty sure all of the high-resolution images are public domain works, although that will require further confirmation. It's an excellent source. [Dcoetzee](#) ([talk](#)) 06:59, 22 April 2009 (UTC)
- Update: I've enumerated about 65000 high-res images, and am in the process of downloading and converting them to PNGs, slow enough to not overwhelm their bandwidth. So far I've retrieved about 17250, occupying 323 GB. I'm also in the process of generating image descriptions of them based on NYPL metadata. I've created [Category:New York Public Library Digital Gallery](#) and plan to start uploading some of them soon. [Dcoetzee](#) ([talk](#)) 13:18, 16 May 2009 (UTC)
- Update: I've had contact from a representative of the NYPL, who has been very helpful in furnishing IDs and sanctioning the sharing of their public domain images. He gave me a list of about 40,000 stereographs which I can begin uploading immediately as soon as I put together a suitable fully-automated upload tool for the task. [Dcoetzee](#) ([talk](#)) 21:43, 25 June 2009 (UTC)
  - Great work. I think this is good news and I'm very happy that someone over there is nice enough to help out.--[Diaa abdelmoneim](#) ([talk](#)) 20:46, 27 June 2009 (UTC)

[http://commons.wikimedia.org/wiki/Commons:Batch\\_uploading/NYPL\\_Digital\\_Gallery](http://commons.wikimedia.org/wiki/Commons:Batch_uploading/NYPL_Digital_Gallery)





WIKIMEDIA  
COMMONS

Search

Go

Search

Navigation

- Main Page
- Welcome
- Community portal
- Village pump

Participate

- Upload file
- Recent changes
- Latest files
- Random file
- Help
- Contact us
- Donate

Toolbox

- What links here
- Related changes
- Special pages
- Printable version
- Permanent link
- Subpages

Log in / create account

project page

discussion

edit

history

# Commons:Batch uploading/NYPL Digital Gallery

[< Commons:Batch uploading](#)

## Images from NYPL Digital Gallery

[\[edit\]](#)

Will be great if we batch upload PD-images from NYPL Digital Gallery - <http://digitalgallery.nypl.org/nypldigital/index.cfm> *NYPL Digital Gallery provides free and open access to over 685,000 images digitized from the The New York Public Library's vast collections, including illuminated manuscripts, historical maps, vintage posters, rare prints, photographs and more.* --[Butko](#) (talk) 14:45, 14 April 2009 (UTC)

- This collection turned out to be more promising than I supposed. They use LizardTech ContentServer to serve up their images, whose API is [described here](#) . Here's how you extract original TIFFs at full size: first use a "browse" query to obtain some XML including the image dimensions, like this one [\[1\]](#) . The folder name and image name can be obtained from URL of the zoom view. Then, use a getimage query like this one [\[2\]](#) to get the full size TIFF, specifying the dimensions from the previous query. Tada. Close examination shows no artifacts in the TIFF - these are original scans (internally, they are SID images). The first one I extracted was 3845 × 4947, about 60 MB as a TIFF, and 27 MB as a PNG (which you can preview [here](#) ). They throttle you at 80 KB/s per transfer, but they *do* allow simultaneous transfers; any way you look at it though it would take a long time to fetch all the images we need. In light of the long download time per image, we're going to want to license filter *before* downloading. [Dcoetzee](#) (talk) 07:00, 15 April 2009 (UTC)
- Update: their complete collection of high-resolution images is browsable [here](#) . This can be used to easily obtain a list of folder-name pairs. I'll presently begin downloading. [Dcoetzee](#) (talk) 06:23, 16 April 2009 (UTC)
- Update: a better way to download these is to use the "getfile" function to get the raw .sid files, which are highly compressed (as in [\[3\]](#) ) and then use LizardTech's command-line decoder to convert to TIFF ([\[4\]](#) ). This is a quicker download and doesn't even require the dimensions. [Dcoetzee](#) (talk) 22:12, 16 April 2009 (UTC)
- I'm still in the middle of grabbing these. Enumerating IDs turned out to be trickier than I thought, because the folders are so large the browse interface times out on them. I ended up enumerating them instead using wildcard searches on single letters. Even just looking at the high res images, it's a *lot* of data. All told we're talking at least 100 GB in PNGs, and I'm pretty sure all of the high-resolution images are public domain works, although that will require further confirmation. It's an excellent source. [Dcoetzee](#) (talk) 06:59, 22 April 2009 (UTC)
- Update: I've enumerated about 65000 high-res images, and am in the process of downloading and converting them to PNGs, slow enough to not overwhelm their bandwidth. So far I've retrieved about 17250, occupying 323 GB. I'm also in the process of generating image descriptions of them based on NYPL metadata. I've created [Category:New York Public Library Digital Gallery](#) and plan to start uploading some of them soon. [Dcoetzee](#) (talk) 13:18, 16 May 2009 (UTC)
- Update: I've had contact from a representative of the NYPL, who has been very helpful in furnishing IDs and sanctioning the sharing of their public domain images. He gave me a list of about 40,000 stereographs which I can begin uploading immediately as soon as I put together a suitable fully-automated upload tool for the task. [Dcoetzee](#) (talk) 21:43, 25 June 2009 (UTC)
  - Great work. I think this is good news and I'm very happy that someone over there is nice enough to help out.--[Diaa abdelmoneim](#) (talk) 20:46, 27 June 2009 (UTC)

[http://commons.wikimedia.org/wiki/Commons:Batch\\_uploading/NYPL\\_Digital\\_Gallery](http://commons.wikimedia.org/wiki/Commons:Batch_uploading/NYPL_Digital_Gallery)





WIKIMEDIA  
COMMONS

Search

Go

Search

Navigation

- Main Page
- Welcome
- Community portal
- Village pump

Participate

- Upload file
- Recent changes
- Latest files
- Random file
- Help
- Contact us
- Donate

Toolbox

- What links here
- Related changes
- Special pages
- Printable version
- Permanent link
- Subpages

Log in / create account

[project page](#)

[discussion](#)

[edit](#)

[history](#)

# Commons:Batch uploading/NYPL Digital Gallery

[< Commons:Batch uploading](#)

## Images from NYPL Digital Gallery

[\[edit\]](#)

Will be great if we batch upload PD-images from NYPL Digital Gallery - <http://digitalgallery.nypl.org/nypldigital/index.cfm> *NYPL Digital Gallery provides free and open access to over 685,000 images digitized from the The New York Public Library's vast collections, including illuminated manuscripts, historical maps, vintage posters, rare prints, photographs and more.* --[Butko](#) (talk) 14:45, 14 April 2009 (UTC)

- This collection turned out to be more promising than I supposed. They use LizardTech ContentServer to serve up their images, whose API is [described here](#) . Here's how you extract original TIFFs at full size: first use a "browse" query to obtain some XML including the image dimensions, like this one [\[1\]](#) . The folder name and image name can be obtained from URL of the zoom view. Then, use a getimage query like this one [\[2\]](#) to get the full size TIFF, specifying the dimensions from the previous query. Tada. Close examination shows no artifacts in the TIFF - these are original scans (internally, they are SID images). The first one I extracted was 3845 × 4947, about 60 MB as a TIFF, and 27 MB as a PNG (which you can preview [here](#) ). They throttle you at 80 KB/s per transfer, but they *do* allow simultaneous transfers; any way you look at it though it would take a long time to fetch all the images we need. In light of the long download time per image, we're going to want to license filter *before* downloading. [Dcoetzee](#) (talk) 07:00, 15 April 2009 (UTC)
- Update: their complete collection of high-resolution images is browsable [here](#) . This can be used to easily obtain a list of folder-name pairs. I'll presently begin downloading. [Dcoetzee](#) (talk) 06:23, 16 April 2009 (UTC)
- Update: a better way to download these is to use the "getfile" function to get the raw .sid files, which are highly compressed (as in [\[3\]](#) ) and then use LizardTech's command-line decoder to convert to TIFF ([\[4\]](#) ). This is a quicker download and doesn't even require the dimensions. [Dcoetzee](#) (talk) 22:12, 16 April 2009 (UTC)
- I'm still in the middle of grabbing these. Enumerating IDs turned out to be trickier than I thought, because the folders are so large the browse interface times out on them. I ended up enumerating them instead using wildcard searches on single letters. Even just looking at the high res images, it's a *lot* of data. All told we're talking at least 100 GB in PNGs, and I'm pretty sure all of the high-resolution images are public domain works, although that will require further confirmation. It's an excellent source. [Dcoetzee](#) (talk) 06:59, 22 April 2009 (UTC)
- Update: I've enumerated about 65000 high-res images, and am in the process of downloading and converting them to PNGs, slow enough to not overwhelm their bandwidth. So far I've retrieved about 17250, occupying 323 GB. I'm also in the process of generating image descriptions of them based on NYPL metadata. I've created [Category:New York Public Library Digital Gallery](#) and plan to start uploading some of them soon. [Dcoetzee](#) (talk) 13:18, 16 May 2009 (UTC)
- Update: I've had contact from a representative of the NYPL, who has been very helpful in furnishing IDs and sanctioning the sharing of their public domain images. He gave me a list of about 40,000 stereographs which I can begin uploading immediately as soon as I put together a suitable fully-automated upload tool for the task. [Dcoetzee](#) (talk) 21:43, 25 June 2009 (UTC)
  - Great work. I think this is good news and I'm very happy that someone over there is nice enough to help out.--[Diaa abdelmoneim](#) (talk) 20:46, 27 June 2009 (UTC)

[http://commons.wikimedia.org/wiki/Commons:Batch\\_uploading/NYPL\\_Digital\\_Gallery](http://commons.wikimedia.org/wiki/Commons:Batch_uploading/NYPL_Digital_Gallery)



A photograph of a workshop or garage. In the foreground, a red forklift is partially visible, holding a black sign with white text. In the background, a white van with its hood open is parked inside the workshop. The workshop has a large open bay door and various tools and equipment are visible inside.

KEEP OUT  
EXPERIMENT  
IN  
PROGRESS



# Thank You

[mark@matienzo.org](mailto:mark@matienzo.org)

<http://matienzo.org/>

<http://twitter.com/anarchivist>