# Archiving Large Swaths of User-Contributed Digital Content: Lessons from Archiving the Occupy Movement

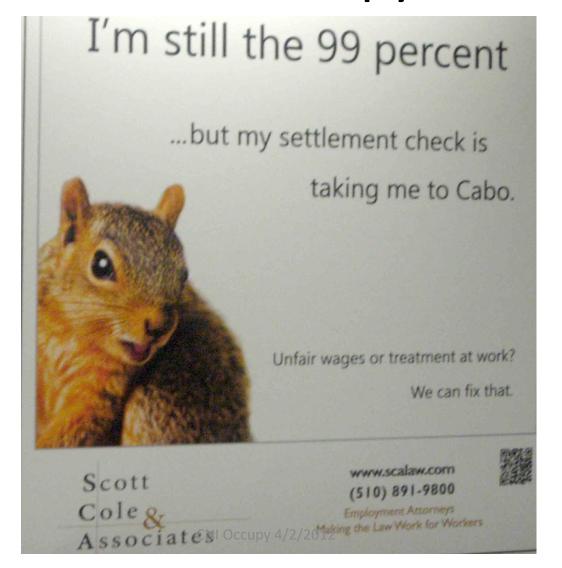
http://activist-archivists.org/wp/

# What's interesting/important about Occupy?

### March 24, 2012 Flickr stats

- "#Occupy" 632,089
- "Occupy Wall Street" 164,304
- "Occupy Protest" 179,454
- "Occupy Movement" 40,572
- "#OWS" 113,904
- "Occupy Oakland" 27,202
- "Zucotti Park" 9,164

# What's interesting/important about Occupy?



# How Occupy material resembles what we'll be facing in the future

- Vast quantity of user-contributed material
- No easy way to control for quality, file format, metadata
  - no enforcing guidelines as with organizational records
  - no semi-consistency as in a single individual's personal records
- Much of the material can most easily be found on Social Networks
- Need to find smart ways to harvest metadata and analyze files, as well as to influence behavior of potential contributors

# How this material may (or may not) be different

- Huge suspicion of conventional organizations (incl universities and libraries)
- DIY mentality—want to control own story
- Reluctance to sanction turning over material to an institution
- Consensus Process becomes more difficult when each meeting attracts a different set of people

# **OWS Archives Working Group**

#### **Archives Working Group**



Posted on January 5, 2012

Occupy Wall Street: Archive Mission Statement

The Occupy Wall Street Archive is an institutionally-independent collection of historical material created and distributed in and around Liberty Plaza of ephemera, signs, and audiovisual materials.

Occupy Wall Street Archives Working Group is created to ensure that the Occupy Wall Street movement will own its past. Its mission is to keep OWS historically self-conscious, and guarantee that our history will be accessible to the public.

Occupy Wall Street Archives Working Group is a collecting initiative, which preserves, safeguards, and makes accessible the records of OWS. It is the repository of the legacy which we will leave for the future generations.

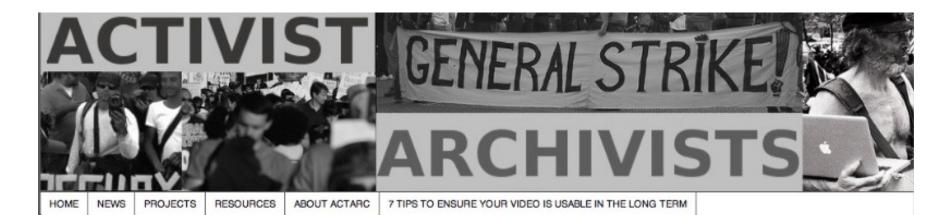
# What I know from my prior work

- InterPARES—If we hope to preserve electronic records, archivists need to be involved early in the life-cycle of that record, long before the record enters the archive
- Preserving Digital Public Television—Pushing metadata gathering upstream into the production cycle

# Archiving Occupy Media-

- Howard Besser, Activist Archivists
- David Millman, NYU Libraries
- Sharon Leon, George Mason U, Cntr for History and New Media
- Kristine Hanna/David Millman, Internet Archive Archive-It

### **Activist Archivist Website**



HOME \* CURRENTLY READING:

#### **Projects**

#### Report on Technical Metadata in Popular Video Sharing Sites

This report explores the embedded technical metadata from the born digital content created by several different sources (iPhone 4S, Motorola Droid 2, and Canon t2i). The resulting spreadsheet, which may be viewed **HERE** (note there are four different tabs, one for each device and the last a color key) charts how the metadata changes as it is first exported to three of the largest video sharing sites (YouTube, Vimeo, and Internet Archive) and then scraped or downloaded.

#### Occupy Preservation Week

Preservation Week is an American Library Association initiative that connects libraries and other institutions through events, activities, and resources that highlight what we can do, individually and together, to preserve our personal and shared collections. This year it will take place during the week of April 22nd, and we will be participating by hosting an event focused on our efforts, in collaboration with the OWS Archives Working Group, to archive and preserve material related to the OWS movement. The event will be free and open to the public, and will take place at 6 pm on April 24th at NYU's Michelson Theatre,

#### UNDER CONSTRUCTION

Please pardon our appearance as we put the site together

OCCUPY WALL STREET ARCHIVES AND RESOURCES

Archive-It: Occupy Movement Seed Collection "T collection documents and aggregates information related to the "Occupy Movement" starting in the Autumn of 2011 and continuing in 2012. In Septem of 2011, a demonstration was held in New York Cit calling itself "Occupy Wall Street".

Archvists Round Table of Metropolitan New York Inc. – Advocacy Committee The committee is currently addressing issues facing the Occupy Wal Street Archive Working Group.

Internet Archive A collection of media about Occu







## Activist Archivists Projects-

- "Why Archive" postcard
- 7 Tips to Ensure Your Video Is Usable in the Long Term
- Study of metadata loss through uploading to services
- Best Practices for Creators/Collectors
- "Toolkit" for Occupy archiving
- Coordinating discussions among various groups archiving different parts of Occupy
- Exploring methods for obscuring identities

# "Why Archive" postcard

- **ACCOUNTABILITY.** Archives collect evidence that can hold those in power accountable.
- **SELF-DETERMINATION.** We define our own movement. We need to create and maintain our own historical record.
- **SHARE.** Archives are a point of entry to our movement's rich record. We can use them to ensure transparency, generate discussion, and enable direct action.
- **EDUCATE.** Today's videos, flyers, web-pages, and signs are material for tomorrow's skill-shares, classes, and mobilizations.
- **CONTINUITY.** Just as past movements inspire us, new activists will learn from the experiences we document.
- RECORD & COLLECT what's happening around you.
- PRESERVE the record.

# 7 Tips to Ensure Your Video Is Usable in the Long Term

- Collect details while filming
- Keep your original raw footage, unaltered
- Make your video discoverable
- Contextualize it
- Make it verifiable
- Allow others to collect and archive
- Or archive it yourself

# Study of metadata loss through uploading to services

iPhone Direct Upload YouTube Scrape		iPhone Comp to YouTube Scrape	iPhone Direct Vimeo Source	iPhone Direct Vimeo Scrape HD		
General			The Contract			
Count	282	282	286	27		
Count of stream of this kind	1	1	1			
Kind of stream	General	General	General	General		
Kind of stream	General	General	General	General		
Stream identifier	0	0	0			
Count of video streams	1	1	1			
Count of audio streams	1	1	1			
Video_Format_List	AVC	AVC	AVC	AVC		
Video Format WithHint List	AVC	AVC	AVC	AVC		
Codecs Video	AVC	AVC	AVC	AVC		
Audio_Format_List	AAC	AAC	AAC	AAC		
Audio Format WithHint List		AAC	AAC	AAC		
Audio codecs	AAC LC	AAC LC	AAC LC	AAC LC		
Complete name	/Users/rufus/Desktop/TEST/Test1 direct iPh		/Users/rufus/Desktop/TEST/74338889.n	/Users/rufus/Desktop/TEST/[Vimeo-]		
Folder name	/Users/rufus/Desktop/TEST	/Users/rufus/Desktop/TEST/IPhone	/Users/rufus/Desktop/TEST	/Users/rufus/Desktop/TEST		
File name	Test1 direct iPhone	iphone computer to youtube		[Vimeo-32757424] test iPhone 4s dire		
File extension	moli	(7)04	mov	TIDA		
Format	MPEG-4	MPEG-4	MPEG-4	MPEG-4		
Format	MPEG-4	MPEG-4	MPEG-4	MPEG-4		
	smp4 m4v m4a m4b m4p 3gpp 3gp 3gpp2 3	mp4 m4v m4a m4b m4p 3gpp 3gp 3gpp2 3	mp4 m4v m4a m4b m4p 3gpp 3gp 3gpg	mp4 m4v m4a m4b m4p 3gpp 3gp 3gp		
Commercial name	MPEG-4	MPEG-4	MPEG-4	MPEG-4		
Format profile	Base Media / Version 2	Base Media / Version 2	QuickTime	Base Media		
Internet media type	video/mp4	video/mp4	video/mp4	video/mp4		
Codec ID	mp42	mp42	ot .	Isom		
Codec ID/Url	http://www.apple.com/guicktime/download	http://www.apple.com/quicktime/download	http://www.apple.com/guicktime/downle	http://www.apple.com/guicktime/dow		
Codec	MPEG-4	MPEG-4	MPEG-4	MPEG-4		
Codec	MPEG-4	MPEG-4	MPEG-4	MPEG-4		
Codec/Extensions usually us	mp4 m4v m4a m4b m4p 3gpp 3gp 3gpp2 3	mo4 m4v m4a m4b m4p 3gpp 3gp 3gpp2 3	mp4 m4v m4a m4b m4p 3gpp 3gpp 3gpp	mp4 m4v m4a m4b m4p 3gpp 3gp 3gp		
File size	9235939	21361913	92250801	1017402		
File size	8.81 MiB	22.3 MB	88.0 MIB	9.70 MIB		
File size	9 M/6	22 M B	88 MIB	10 M/B		
File size	8.8 M/B:	22 M/8	88 MIB	9.7 M/8		
File size	E.B. MIB	22.3 N/B	88.0 M/B	9.70 MIB		
File size	8.806 M 8	22.28 Mi8	87.98 MIB	9.703 14.8		
Duration	38700	30750	30682	-3075		
Duration	30s 700ms	30s 750ms	30s 682ms	30s 750ms		
Duration	30s-700ms	30s 750ms	30s 682ms	30s 750ms		
Duration	30s 700ms	30s 750ms	30s 682ms	30s 750ms		
Duration	00:00:30:200	00:00:30.750	00:00:30.682	00:00:30.250		
Overall bit sate mede	Name :	Linn	001001001000	ninn		

#### **Best Practices for Content Collectors**

- Security
  - Sensitive material
  - Scraping for content
- Content Search
  - Internet Archive, Archive-It, YouTube, Vimeo, Bit-Torrent
- Receiving Content
- Metadata Extraction
- Copyright

#### **Best Practices for Content Creators**

- Security
  - Hidden camera laws, parties' consent laws
- Capturing Content
  - Highest quality, set date and time-stamps, note location
- Offloading Content
  - Raw files directly onto computer, keep material organized
- Uploading Content
  - Importance of tagging, review of diff services
- Depositing with an Archive
- Copyright

## Occupy Archiving Kit

- Why Archive?
- What is an "archive"? How do I create an archive?
- Creating archiving-friendly content
- How can I collect materials for the archive?
- What should I save?
- How should I organize my materials? How do I get it into the archive?
- Description/Metadata
- Media Management
- Storage & Preservation
- Access
- Exhibition and Presentation/Outreach
- Rights and Re-Use

# Skill-shares for Occupiers



Collage by Taeyoon Choi

At this free event, the public will be invited to explore and archive the many views of democracy. Anyone wishing to present their work in art and activism can sign up for a short time slot. Email taeyoon [at] eyebeam [dot] org to reserve your slot now.

OWS Archive Share Day is presented in collaboration with OWS Archives members Anna Perricci and Christine O'Heron, in conjunction with 'Speakers' Corners,' an exhibition by Taeyoon Choi CNI Occupy 4/2/2012 currently on view at Evebeam.

# OWS Archive Share Day used variety of tools

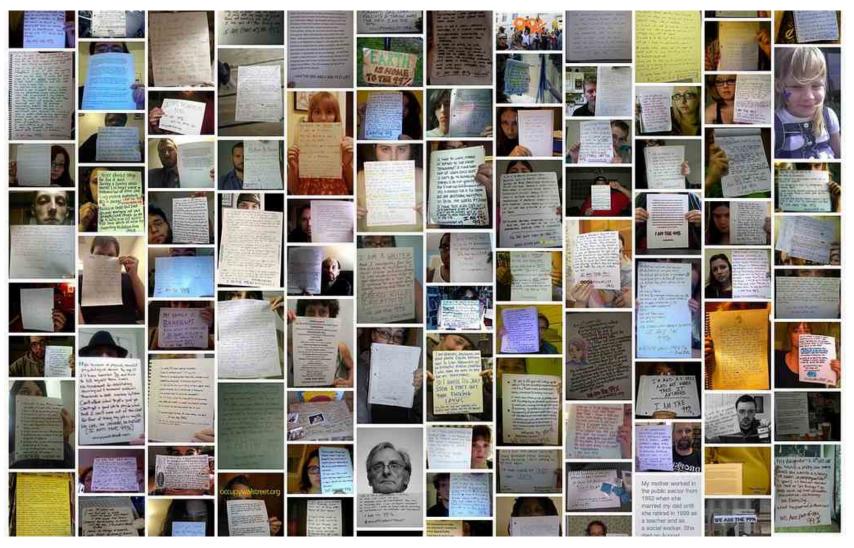
Bulkr in just over 1 minute



### Bulkr

- Batch download from FLICKR
- Set selection filters (#OWS, Creative Commons)
- Set what elements you want to grab (essence, EXIF metadata, tagged-text metadata

# Downloaded FLICKR image



### Downloaded EXIF metadata

```
Elickr Photo Metadata
   Created by Bulkr on Mar 31, 2012
   http://clipyourphotos.com/bulkr
  INFO |
+----
Photographer : blulaces
Photo URL : http://www.flickr.com/photos/blulaces/6209835586/
License : Creative Commons (http://creativecommons.org/licenses)
Taken Date : Man Oct 3 19:13:88 GMT-8488 2811
Upload Date : Mon Oct 3 22:13:08 GMT-0400 2011
Geotag Info : Latitude:40.706286, Longitude:-74.009528
             : 152
Views
             : 0
Connents
  TITLE
WallpaperForThe1Percent
   DESCRIPTION
A New Desktop Wallpaper for the 1%
<a href="http://wearethe99percent.tumblr.com/archive"</p>
rel="nofollov">wearethe99percent.tumblr.com/archive</a>
<a href="http://blulaces.com/" rel="nofolloy">blulaces.com</a>
+----
 TAGS
occupy occupywallst occupywallstreet occupyeverywhere wearethe99percent
wearethe99% 99percent "99% ipercent" 1% unemployment poverty hunger inequality
tumblr wallpaper ows classwar posters signs blulaces jayshells
CNI Occupy 4/2/2012
```

# Other Archive Share-Day and Hackathon activities

- Re-mixing of older footage
- Creating a visual timeline
- Mining material for data (eg. number of colocations of an officer's name with "pepper spray")

#### **Creative Commons Guidance**

- Creative Commons lets you mix-and-match four different conditions:
  - Attribution: You let others copy, re-use and distribute your video, but they must credit you.
  - Share-Alike: You let others copy, re-use and distribute your video, only if they
    do the same with the work they create.
  - Non-Commercial: You let others copy, re-use and distribute your video for non-commercial purposes only.
  - No Derivative Works: You let others copy and distribute your video, but not to create new works using it.
- You can use these conditions in different combinations to share your work in a controlled way. Creative Commons licenses are legal tools that depend on pre-existing copyright laws. Having a Creative Commons license on your work may give you legal recourse, but it may not actually prevent people from downloading and re-using your video illegally.

## Marking Creative Commons licenses

- There are a few ways to mark your video with a Creative Commons license. One way is to include a Creative Commons "bumper" or text card in your video. Creative Commons has created some with graphics that you can download from their website. This method is useful if your video is going to be shared offline (e.g. on DVD, live screenings), as the license information is attached to the video itself.
- Another way to mark your video with a Creative Commons license is to publish your video on platforms that are Creative Commons-enabled, such as YouTube, Vimeo, or Internet Archive. These platforms allow you to easily select a license during the upload process. This method is useful because the license is machine-readable. A search engine, for example, can detect the license.

### Promoting ObscuraCam

- "ObscuraCam is a visual privacy app for photo and video, that gives you the power to better protect the identity of those captures in your photos, before you post them online"
- Developed by Guardian Project in conjunction w/Human Rights group WITNESS-

### ObscuraCam



#### **Activist Archivist Collaborations-**

- With Occupy Wall Street groups
  - Helping store and manage Media Working Group's media streams
- With Tamiment Collection
  - Methods for collaborative selection of YouTube videos
  - Methods for extracting metadata from recordings

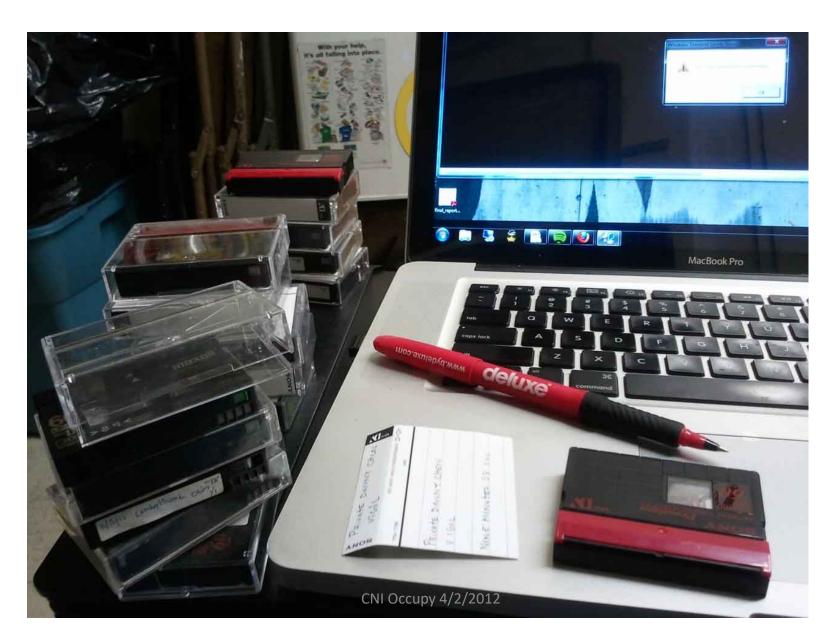
# Occupy's Media Working Group— Streaming Services (1/3)

- livestream, ustream—almost impossible to hack into downloads
- One of Occupy Oakland's streamers uses the streaming service as his fileserver.
  - does no recording onto local media (to avoid confiscation of equipment)
  - when he wants to edit, he downloads from streaming service, edits, then uploads it back
  - trusts the streaming service, and thinks that it will be around forever

# Occupy's Media Working Group— Streaming Services (2/3)

- OWS keeps masters for their streams and other media locally, and as numbers grow, they're having trouble with managing the files and storage and want a DAM, but...
  - Activist Archivists (AA) brokered a deal to put copies of their content on a reliable archival service
  - AA helped them select open source tool to use for cataloging (Omeka), and AA volunteer is currently cataloging their mini DV collection of footage from all General Assemblies and Spokes Councils from September 17th to date--

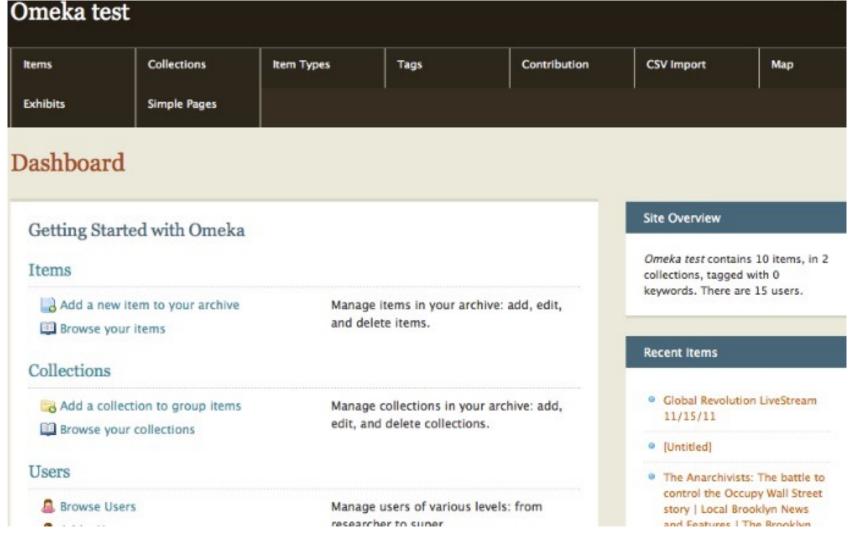
# Cataloging GA/Spokescouncil mini-DVs



### Metadata for OWS Core google doc

A	В	С	D	E	F	G	Н	1	J
Timestamp	Title	Description	Name	OWS Committe or Group	Location	Filename or Link	Keywords or Hashtags	Name of Catalouger	Location
3/29/2012 20:56:33	East New York Foreclosure Arrests and Protests	The first half of the tape is footage from protests against the foreclosure of Quincy Drayton's home. Footage shows police officers arresting occupiers and one man was gashed in the head as officers broke down the door of the home.  The second half the the piece is Elizabeth and another occupier recounting the experience later at the Occupy offices.		Global Revolution	Global Revolution Office			Kelly Haydon	East New York, NJ
3/29/2012 20:02:43	March & Vigil for Danny Chen	Vigil and March for Danny Chen at night in Columbus Park, Chinatown.		Global Revolution	Global Revolution Office		DannyChen	Kelly Haydon	New York, NY
3/3/2012 17:43:51	December 15th Put CHen	This might be a general assembly tape, not sure.		Global Revolution	Global Revolution Office			Kelly Haydon	New York, NY
3/3/2012 17:38:12	March on NDAA			Global Revolution	Global Revolution Office			Kelly Haydon	New York, NY
3/3/2012 17:37:37	March on NDAA			Global Revolution	Global Revolution Office			Kelly Haydon	New York, NY
3/3/2012 17:36:28	March on NDAA			Global Revolution	Global Revolution Office			Kelly Haydon	New York, NY
3/3/2012 17:30:38	NYC March + Police Issues	There is some metadata on here that I can't read"hony pory?"		Global Revolution	Global Revolution Office			Kelly Haydon	New York, NY

# Omeka for Cataloging/Managing



# Occupy's Media Working Group— Streaming Services (3/3)

- Currently working on plan to turn over the Global Revolution streaming collection to NYU's Tamiment
  - OWS will NOT sign a donor agreement, as they see that as conferring exclusivity to a bureaucratic organization
  - Instead, OWS will execute a Creative Commons license letting anyone else use the material, and AA volunteers will make sure that the corpus is transferred to NYU-

#### YouTube

- Tamiment has been selectively browsing through YouTube Occupy videos, trying to choose which ones to keep, then cataloging them with
  - Title, Creator, Creation Date, Upload Date,
     Description, URL, Youtube Username, License,
     Format, Codec, Source Media, On Internet
     Archive, CC License type

## Tamiment YouTube Cataloging

Title	Creator	Creation Date	Upload Date	Description
Right Here All Over (Occupy Wall St.)	Directed by Alex Mallis	?	5-Oct-11	Artistic coverage of early occupatio
Where Do We Go From Here? Occupy Wall St.	Director: Ed David, Prod	October 18?	19-Oct-11	Selected protestor statements in Zu
Occupy Wall St. Continues	RT	21-Sep	22-Sep-11	Coverage from Russia Today on 21
Michael Moore @ Occupy Wall Street	?	26-Sep	26-Sep-11	Documentary filmmaker Michael M
NYPD Cop Punches Protester at Occupy Wall Street, 10 14 11	ReasonTV	14-Oct	14-Oct-11	New York police officer assaults pro
Occupy Wall St General Assembly Oct.18	?	Oct 18?	18-Oct-11	Live recording of "people's mic" del
Occupy Wall Street Soros? AdBusters? somebody on drugs? Declaration	?	15-Oct	18-Oct-11	15 Octover, 2011 reading of Declera
kalle Lasn Adbusters is trying to control Occupy Wall Street	Judy Lapointe	?	19-Oct-11	Internet personality "Judy101101"
Hate at Occupy Wall Street	Emergency Committee f	?	13-Oct-11	Public service announcement, prod
Mayor Mike Bloomberg, Wall Street Bankster, Meets the 99% at #OWS	?	12-Oct	12-Oct-11	Bloomberg walks through crowd at
#OccupySF - Police Violence - 2011-10-05	?	5-Oct	6-Oct-11	Live hand held recording of police r
New York City Protest - Occupy Wall Street - 4:45 PM	?	17-Sep	17-Sep-11	Hand held livestream recording of p
OCCUPY WALL STREET PROTESTERS' ARRESTS	?	20-Sep	20-Sep-11	Police arrest Jason Ahmadi, Eric Ge
Anonymous - Occupy Wall Street	Anonymous	?	10-Sep-11	Animation discussing preliminary o
Occupy Wall Street: Police brutality, media blackout	RT	?	27-Sep-11	Russia Today discusses police agres
Occupy Wall Street Protests	RT	?	17-Sep-11	Russia Today covers protests on 17
PEACEFUL FEMALE PROTESTORS PENNED IN THE STREET AND MACEDI- #OccupyWallStreet	The Other 99%	24-Sep	24-Sep-11	Hand held live recording of 5 female
cops beating people up at occupy wall street	?	5-Oct	5-Oct-11	Hand held recording of New York pr
tear gassing of the kitchen Violent arrests pepper spray #OWS #occuppywallstreet	?	15-Nov	15-Nov-11	Hand held recording of forced evict
Keep wall street occupied		?	27-Oct-11	Internet personality outlines mail b
20,000 Protestors To Occupy Wall St.		?	2-Sep-11	Pre-demonstration narration promo
Peter Schiff Speaks for 1 Percent at Occupy Wall Street	ReasonTV	?	27-Oct-11	Investment broker and author Pete
I AM NOT MOVING	?	?	11-Oct-11	Short film juxtaposing US foreign po
'Occupy' protesters angered by bank profits	Aljazeera	?	17-Oct-11	17 October, 2011 Aljazeera report o
An Open Message to the 99% (Occupy Wall Street)	?	?	26-Oct-11	Narrated explanation of debt as sla
New York City Protest - Occupy Wall Street - Video and Photo	?	17-Sep-11	17-Sep	Montage of September 17th march
What We Saw at the Occupy Wall Street Protest	ReasonTV	?	7-Oct-11	4 October, 2011 interviews in Zucco
#OWS Protesters March On The Brooklyn Bridge 700+ Arrested 10-1-11	?	1-Oct-11	2-Oct-11	Hand held footage of October 1st n
Military at #occupywallstreet	We Are Change	?	3-Oct-11	Luke Rudkowski of We Are Change
WOccupyWallStreet: Man assaulted, hidden from Press, denied lawyer, 10.14.11	1	14-Oct-11	14-Oct-11	Hand held footage of events of 140
#ows #occupywallstreet general assembly facilitation	?	8-Oct-11	9-Oct-11	Independent footage of lecture on

### But this won't scale!

#### March 24, 2012 YouTube stats

- "#Occupy" 169,000
- "Occupy Wall Street" 98,400
- "Occupy Protest" 70,500
- "Occupy Movement" 54,800
- "#OWS" 50,300
- "Occupy Oakland" 13,400
- "Zucotti Park" 6,690

# Alternative approach to YouTube Selection process

- Develop categories of important YouTube videos
  - Celebrity visits, Internal workings (library, kitchen, media), Confrontations with police, Labor, Housing, etc.
- Have Occupiers fill in an online form listing the
   5 most important videos in each category

# Advantages of YouTube Collaborative Filtering Selection Process

- Scalable and manageable
- Consistent with Occupy ideas of inclusiveness and of managing own story
- Tamiment can still choose to be selective in collecting only a portion of what is voted in, but the total set for review is a manageable scale

# Social Networks starting to police their "Terms of Use"

News > Technology > Facebook

# Facebook privacy chief fires warning shot at bosses who demand access

Social networking site warms against employers who ask for the passwords of job applicants, saying legal action is possible

Dominic Rushe in New York



guardian.co.uk, Friday 23 March 2012 15.56 EDT

### YouTube User Agreement

• 5B "You shall not download any Content unless you see a 'download' or similar link displayed by YouTube on the Service for that Content."

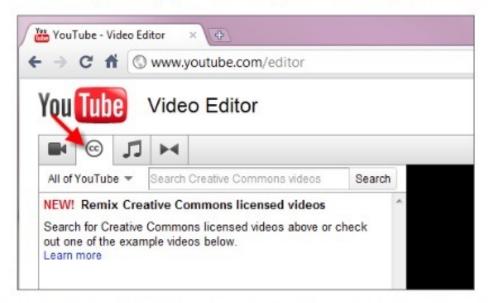
But even 1916 Public Domain material on Creative Commons YouTube Channel don't have "download" button



# YouTube sees Creative Commons as only for Re-Mix, not archiving

#### How do Luse a Creative Commons video?

- Enter the YouTube Video Editor (accessible by clicking your username in the top right corner of any page, selecting My Videos from the drop-down menu, and selecting Video Editor from the options up top.
- 2. Click the CC tab located under the YouTube logo.
- Either enter the URL of the video you wish to use in the search bar, or enter keywords to search through Creative Commons.
- 4. Click the "+" sign in the top right corner of the clip you want to add it to your video for editing.



Please note: Only videos that have been tagged Creative Commons will display in the search results. Additionally, advertisements may be displayed on videos created using the YouTube Video Editor.

#### Things you should know

The ability to mark uploaded videos with a Creative Commons license is only available to users whose account is in god standing. You may check the status of your account settings.

# But using YouTube's video editor to view and "Re-Mix" exact copies is probably legal

### **Archiving Occupy Media-**

- Howard Besser, Activist Archivists
- David Millman, NYU Libraries
- Sharon Leon, George Mason U, Cntr for History and New Media
- Kristine Hanna/David Millman, Internet Archive Archive-It