#### Normalizing Data as Part of an ArchivesSpace Migration



Students in front of Main Building, 1943, photographer unknown

Adam Strohm

Director, University Archives and Special Collections Illinois Institute of Technology

Archives \* Records 2018 August 17, 2018

## University Archives and Special Collections

- Part of Paul V. Galvin Library at Illinois Institute of Technology
- Staff of 4
- ► ca. 1,500 collections



Paul V. Galvin Library, photographer: Adam Strohm

#### Where We Started

Multiple metadata creators over

many years

- Inconsistencies & idiosyncrasies
- Emphasis on keyword search

# Re:discovery Proficio

Collection Nhr: 027.01.05 Student Events Collection Collection Title: Inclusive Dates: 1950-2006 Bulk Dates: 1950-1970 Extent: [Extent]5 [Unit]Folders Language: Appraisal: Provenance: Taken from 1998.130 Item Status: Available for Research Notes: Summary: Materials related to student events at IIT, mostly flyers and programs. Creator: Multiple creators Creator Role: Bio/Ora History: Flyers and programs for events (concerts, plays, exhibits, etc) held at IIT. Arranged by broad themes, following folders: 1) Theater: Center Stage presents "Look Back in Anger." n.d.; Center Stage presents "No Exit", n.d.; Armour Faculty Club presents "A Moon for the Misbegotten", n.d.; Campus Players Winter Production "Mr. Roberts", 1953; The Women's Club presents "The Late Christopher Bean" March 16, 1957; Rough Castings presenting "The Road to Ruin", May 13-15, 1954; Rough Castings presents "You won't find it here" n.d.; Rough Castings presents "We Never Had It So Good" April 23-25. 1953: Dramatech presents "Miracle on 33rd Street", May 13-14, 2) Music: Music Clubs Spring Concert, n.d.: Music Clubs Concert, n.d.: Music Clubs Christmas Concert, n.d.: Christmas Concert, December 18, 1953; Music Club Spring Concert, 1953; Spring Concert, May 1954; Annual Spring Concert, 1956; IIT Student Recital, 2006, 3) Flyers: "Club Niites" at the Bog; College Bowl, CBS Radio Mini-Week, 1981; Shakespeare in the Center (stationary), The National Shakespeare Company presents Richard II, 1981; The Second City, n.d.; Music Theatre of Chicago presents "Godspell" n.d.; artist reception for Michael James Riddet, 1981; The Music Theater of Chicago presents "Kiss Me Kate", Scope-Content: 1981: The threepenny opera presents "Kurt Weill" 1969: The Berlin Exhibit: concert pianist Julian Bern, 1970: an Exhibition from the People's Republic of China "The Great Bronze Age of China", 1980: exhibit of sculpture by Simon Gordon: 40 paintings by Rufino Silva, 1967; Carl E Schwartz and Kav Hofmann-Schwartz exhibit, 1977; City Lit Theater Company presents "Big Blonde"; list of prices for photographs by Elaine Perlman 4) HUB/Hermann Hall exhibits and programs: Allan Bruce Zee, Photographs, 1974; John Asquith, Serigraphs and Serivacs, 1874; Allan Bruce Zee, Photographs, 1974; Frank Rakoncav and Theodore T. Gall, 1975: Sylvia Birch Halperin, Paintings, Drawings, and Prints, 1975; Zelda W. Werner and Richard L. Hammond, painting and sculpture, 1975; Carl E. Schwartz, Paintings, Drawings, Prints, 1977 Steve Horan, Paintings and Prints, 1978: Jan Miller, Paintings, 1979: Joesph A Kotzman, 1986: St. John Earp, A retrospect of the visionary artist, 1987; B.J. Cohen, Sights, 1987; The National Shakespeare Company presents Julius Caesar, 1987; Brie Taylor, Banner Designs of IIT, 1987; Helen Nelson Reed, 1988; Lawrence W. Oliverson, 1988; The Berlin Exhibit 5) Office of Student Affairs: Graduate Salute, Fall 1999 Organiz/Arrange: Itemized Listing: Assoc Materials: Pers Name:

Itemized Listing: Assoc Materials: Pers Name: Corp Name: Topic Term: Student Life --Musicians Functions/Act.: IIT Buildings: Gener/Form: Geogr Name: Meeting Name:

### You're Never Quite Ready

- ArchivesSpace forced us to make some changes (records without dates, extents)
- Other normalization was easiest with data in csv format

But...

The data was never going to be perfect and presentable immediately after ingest

### Find the Right Tools For the Job

- Names and subjects were uncontrolled
- No system-imposed master list existed, and no links between values

Browse Term	Term Type	Browse Ter
College Bowl	Topic term	Hermann U
College of ArchitectureIIT	Corporate name	Hern, Matth
College of Psychology	Topic term	Heron, Reg
College of Psychology-IIT	Corporate name	Heskett, Jo
College of Science and LettersIIT	Corporate name	Heuermann
Collens Scholarship Program	Topic term	Hewitt, Lou
Collens, Lewis (Lew)	Personal name Personal name	Heyd, Kurt
Collens, Marge		Higgins, ???
Collins, John	Personal name	Higgins, Ge
Colonade Park (aka Branch Brook), Newark, New Jersey	Geographic term	Higgins, M High Scho
Commencements	Topic term	
Committee on the Future of IIT (COMFIIT)	Topic term	Hilberseime
Commons (building)	Topic term	Hilberseime
Communication Engineering School	Corporate name	Hill, Doug
Community Relations	Topic term	Hilvers, Rol
Comprehensive Plan for Chicago (1965)	Topic term	Himmelrei
Computer Engineering	Topic term	Hirano, Mar
Computer Science	Topic term	Hiroshima, Hispanics

Browse Term	Term Type		
Hermann Union Building (AKA: HUB) (SEE: Hermann Hall)	Topic term		
Hern, Matthew	Personal name		
Heron, Reginald	Personal name		
Heskett, John	Personal name		
Heuermann, Magda	Personal name		
Hewitt, Louise	Personal name		
Heyd, Kurt	Personal name		
Higgins, ???	Personal name		
Higgins, Geoffrey T.	Personal name		
Higgins, Michael	Personal name		
High Schools-Chicago	Topic term		
Hilberseimer, Ludwig	Personal name		
Hilberseimer, Ludwig K.	Personal name		
Hill, Doug	Personal name		
Hilvers, Robert Joseph	Personal name		
Himmelreich, Dean	Personal name		
Hirano, Marjorie Yasuko	Personal name		
Hiroshima, Japan	Geographic term		
Hispanics	Topic term		

# OpenRefine

- Reconciled name and subject data against Library of Congress authority files
- Uncovered discrepancies
- Automatically added URIs from LC when matches were found

Facet / Filter	Facet / Filter Undo / Redo o		40	10 r	ows					
Refresh	Reset All	Remove All	Show as: rows records			ws records	Show: 5 10 25 50 rows			
× Browse Term change		change				Added By	Added Date	Bio/Hist	Birth Year	Browse Term
400 choices Sort b	y: name count	Cluster								Alumni Memorial Hall)
Animation 1 Annual Reports-Paul V. Galvin Library 1 Antarctic Expedition 1		岔		254.	cbruck	1/13/2010 10:14			Alumni Memorial Hall	
				255	cbruck	2/3/2015 10:00			Alumni-Armour Institute of Technology	
Anthropology 1			-		256.	cbruck	1/14/2010 11:45			Alumnifest
Aquatecture 1					257.	cbruck	12/1/2011 14:56			Alumni-ID
Arcade Building 1					258	cbruck	1/13/2010 16:32			Alumni-IIT
Architecture Class	sesInstitute of					cbruck	8/3/2010 14:22			American Institut
Design 1 Architecture of Mies van der edit include Rohe, The (Michael Blackwood film) 1		25		235.	CHINCK	0.012010 14.22			of Electrical Engineers (AIEE Chicago Chapter	
Architecture prog										
Architecture progr					260.	cbruck	3/3/2014 13:09			American Scene
Architecture students 1 Architecture with Technology: A New Synthesis (symposium) 1 Armour Institute of Technology buildings 1 Armour News, The 1 Armour Research Foundation Personnel 1 Personnel 1 Armour Tech Athletic Association 1 Army ROTC (SEE ALSO: Amy ROTC (SEE ALSO: Reserve Officers Training Corps) 1 At & Sculpture 1		1094.: S		10000					(TV and radio	
		17		261	cbruck	2/25/2015 15:07			series) Animation	
				262	cbruck	12/16/2011 9:35			Annual Reports-	
				202.	COLOCK	12/10/2011 9:55			Paul V. Galvin Library	
				263.	RED	12/4/2009 14:08			Antarctic Expedition	
				264	cbruck	12/3/2010 15:39			Anthropology	
				265.	cbruck	10/22/2010			Aquatecture	
		☆		266.	cbruck	10/27/2011 15:43			Arcade Building	
				267.	cbruck	7/18/2011 15:51			Architecture ClassesInstitute of Design	
art@IIT 1										
Athletics 1			岔		268.	cbruck	6/4/2012 13:53			Architecture of Mies van der
Atomic Bomb 1		*								Rohe, The (Michael Blackwood film)
					269.	cbruck	11/7/2014 15:24			Architecture program-AIT
			Û		270.	RED	12/4/2009 14:08			Architecture program-IIT
						cbruck	1/6/2010 10:38			Architecture students
			3		272	chruck	5/14/2010 9:33			Architecture with

#### reconcile-csv

- OpenRefine Reconciliation service that matches one dataset against another
- Good for adding local metadata

http://okfnlabs.org/reconcile-csv/

	IIT Name Authority File Edit View Insert	Format Data Tools Add-ons Help Last edit was 16 minutes ago			
	erat s %	.0, .00, 123 - Calibri - 11 - <b>B Z</b> - <b>5</b> <u>A</u> - <b>3</b>	H + H + More		
fx	330 North Wabash (Chicago,	Ш.)			
	A	в	1		
1	Agent Type	Primary Name	Authority ID		
24	agent corporate entity	Illinois Institute of Technology. Interprofessional Projects Program	n000348		
25	agent_corporate_entity	Arthur Keating Hall (Chicago, Ill.)	n000346		
26	agent_corporate_entity	Alumni Memorial Hall (Chicago, Ill.)	n000343		
27	agent_corporate_entity	Commons Building (Chicago, Ill. : South Wabash Avenue)	n000342		
28	agent_corporate_entity	Armour Institute of Technology. Athletic Association	n000341		
29	agent_corporate_entity	oorate_entity Armour Flats (Chicago, Ill.)			
30	agent_corporate_entity	ent_corporate_entity Armour Research Foundation Administration Building (Chicago, Ill.)			
31	agent_person	Binga, Jesse, 1865-1950	n000338		
32	agent_corporate_entity	Illinois Institute of Technology. Early Identification Program	n000330		
33	agent_corporate_entity	Illinois Institute of Technology. Education and Experience in Engineering (E3	) n000328		
34	agent_corporate_entity	Illinois Institute of Technology. Computation Center	n000327		
35	agent_corporate_entity	Illinois Institute of Technology. Housing Office	n000326		
36	agent_corporate_entity	Illinois Institute of Technology. Department of Biology	n000325		
37	agent_corporate_entity	Hoerr Schaudt Landscape Architects (Chicago, Ill.)	n000323		
38	agent_corporate_entity	Herman Union Building (Chicago, III.)	n000313		
39	agent_corporate_entity	George Williams College	n000312		
40	agent_corporate_entity	Grand Theater (Chicago, Ill.)	n000311		
41	agent_corporate_entity	Hallett Movers (Summit, III.)	n000310		
42	agent_corporate_entity	Environmental Systems Design, Inc. (Chicago, III.)	n000309		
43	agent_corporate_entity	Hermann Hall (Chicago, Ill.)	n000308		
44	agent corporate entity	Illinois Institute of Technology. Institute for Science. Law and Technology	n000301		

#### Sometimes You Have to Get Your Hands Dirty

- Batch edits couldn't fix everything manual editing was required
- 4 staff members edited ca. 1,500 records by hand
- Made titles more consistent, added subjects & names, formatted dates, cleaned up Scope and Contents notes

- We're not done (and never will be)
- No normalization process is ever normal
- It can't all happen at once
- Clean-up is possible, and you're not alone



Thanks! astrohm@iit.edu findingaids.archives.iit.edu

SPRING THING 1970

Spring Thing 1970, photographer unknown