Normalizing Data as Part of an ArchivesSpace Migration

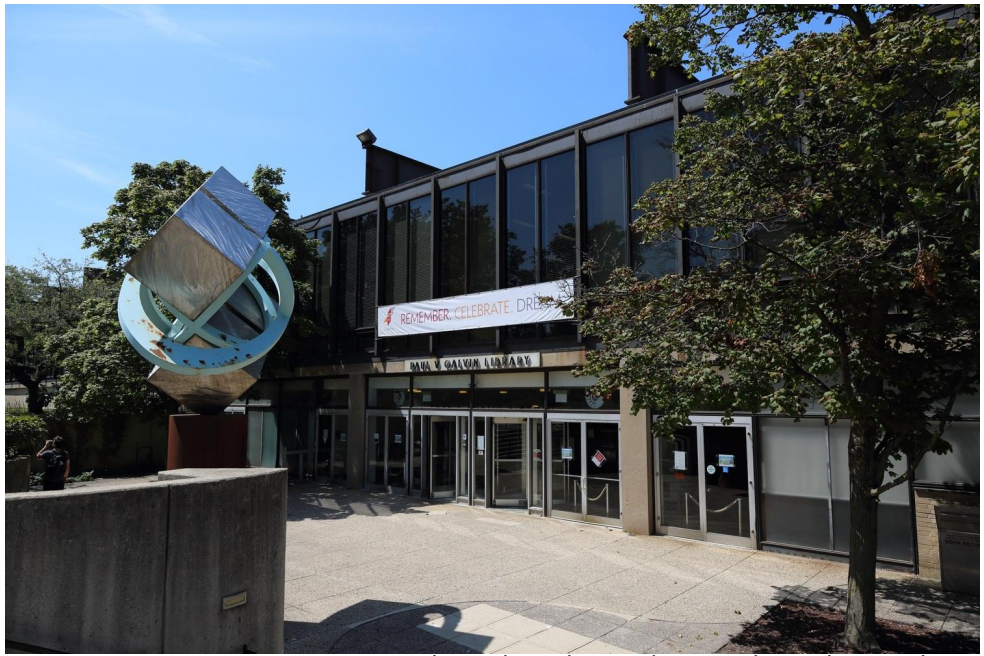
Students in front of Main Building, 1943, photographer unknown

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 SOFTWARE

Proficio

Collection Nbr:	027.01.05
Collection Title:	Student Events Collection
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/Org History:	
Scope-Content:	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", 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 Kay 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 Rakoncay 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:	
Corp Name:	
Topic Term:	Student Life --Musicians
Functions/Act.:	
IIT Buildings:	
Genre/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

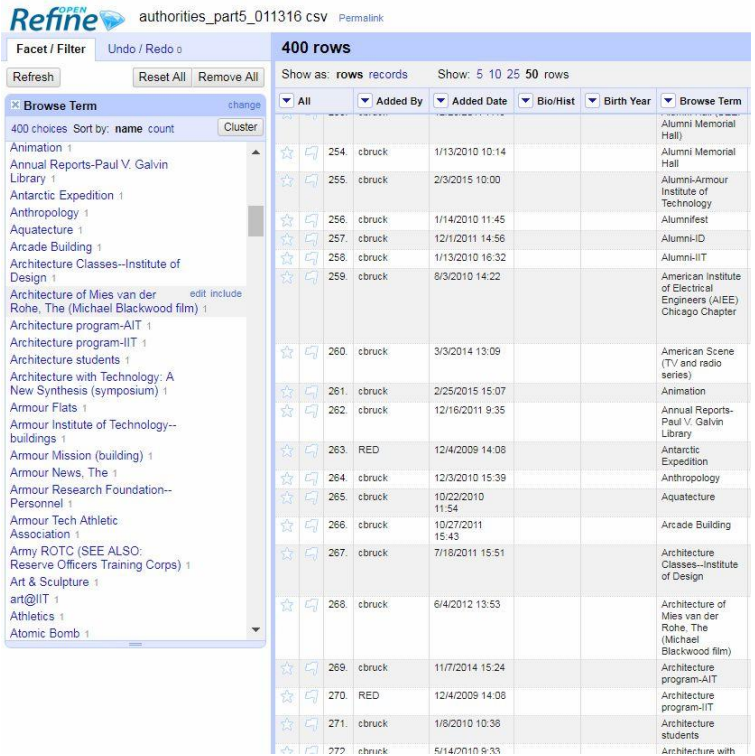
- Names and subjects were uncontrolled
- No system-imposed master list existed, and no links between values

Browse Term	Term Type
College Bowl	Topic term
College of Architecture--IIT	Corporate name
College of Psychology	Topic term
College of Psychology-IIT	Corporate name
College of Science and Letters--IIT	Corporate name
Collens Scholarship Program	Topic term
Collens, Lewis (Lew)	Personal name
Collens, Marge	Personal name
Collins, John	Personal name
Colonade Park (aka Branch Brook), Newark, New Jersey	Geographic term
Commencements	Topic term
Committee on the Future of IIT (COMFIIT)	Topic term
Commons (building)	Topic term
Communication Engineering School	Corporate name
Community Relations	Topic term
Comprehensive Plan for Chicago (1965)	Topic term
Computer Engineering	Topic term
Computer Science	Topic term

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

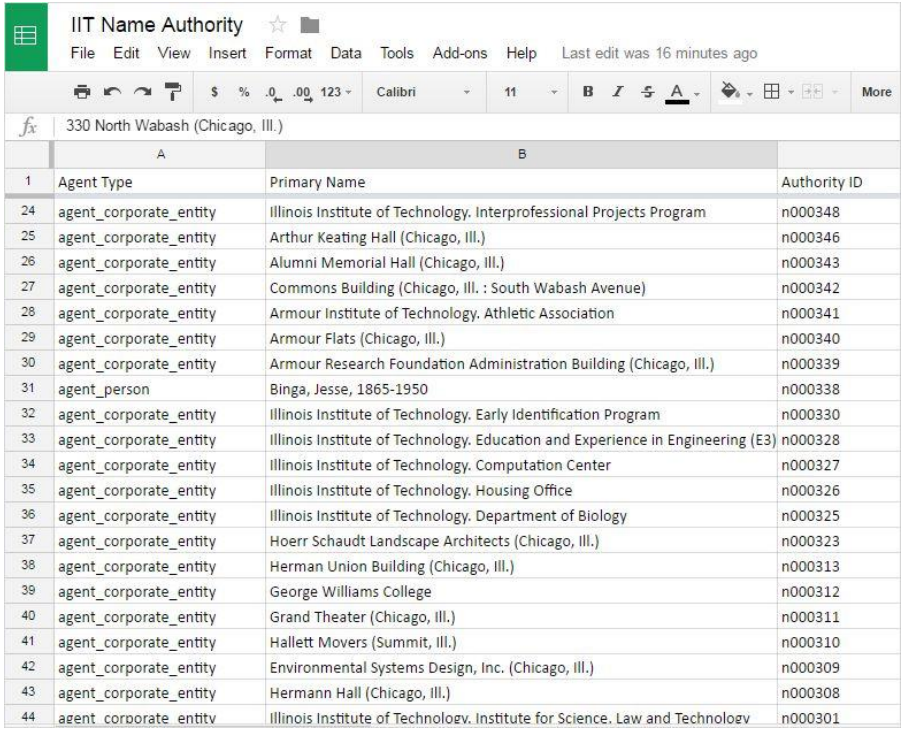
- Reconciled name and subject data against Library of Congress authority files
- Uncovered discrepancies
- Automatically added URIs from LC when matches were found



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

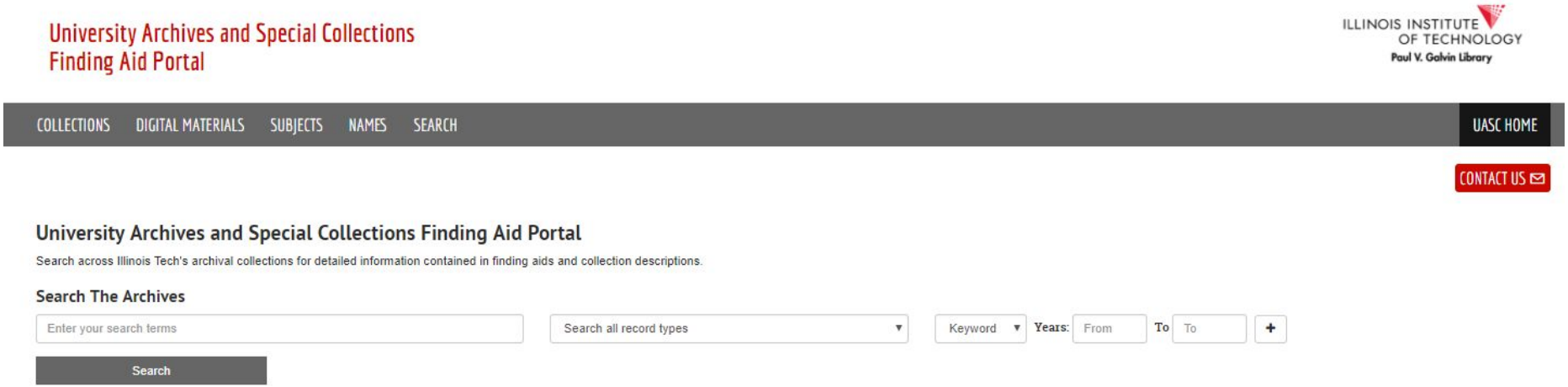
330 North Wabash (Chicago, Ill.)

	A	B	
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	Armour Flats (Chicago, Ill.)	n000340
30	agent_corporate_entity	Armour Research Foundation Administration Building (Chicago, Ill.)	n000339
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, Ill.)	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, Ill.)	n000310
42	agent_corporate_entity	Environmental Systems Design, Inc. (Chicago, Ill.)	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, photographer unknown