

Reconciling Legacy Archival Metadata

Greer Martin
Paul V. Galvin Library,
Illinois Institute of Technology

Mashcat 2017

ILLINOIS INSTITUTE
OF TECHNOLOGY 


The Project

- Migrate 5,000 records from Re:Discovery Proficio to ArchivesSpace
- Over 1,700 are name records
- Complete in 1 year
- Staff of 1 librarian + 1 volunteer (Alison)
- The catch: No authority control, records are not linked

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
Computers	Topic term
Computers & Computing, History of	Topic term
Concrete Canoe	Topic term

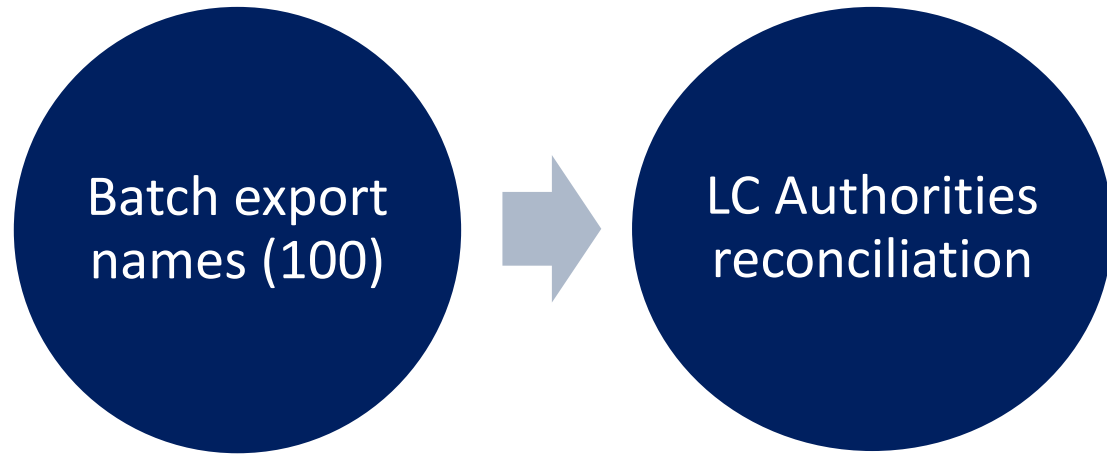
Browse Term	Term Type
Hermann Union Building (AKA: HUB) (SEE: Hermann Hall)	Topic term
Hern, Matthew	Personal name
Heron, Reginald	Personal name
Heskett, John	Personal name
Heuermann, Magda	Personal name
Hewitt, Louise	Personal name
Heyd, Kurt	Personal name
Higgins, ???	Personal name
Higgins, Geoffrey 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
Hock, Daniel	Personal name
Hocker, Thomas	Personal name
Hoffman, Ellis	Personal name
Holabird and Roche	Corporate name

The Plan

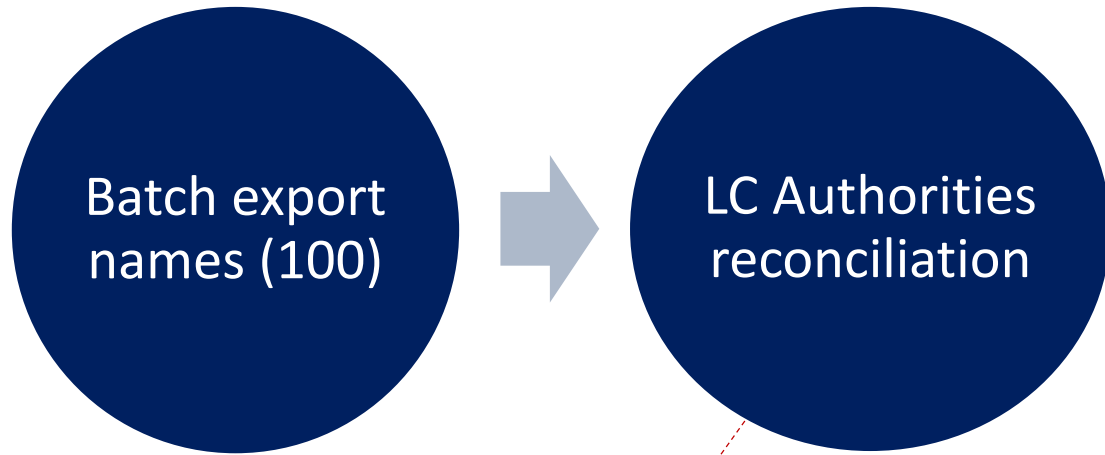


Batch export
names (100)

The Plan



The Plan



Adler and Sullivan

<http://id.loc.gov/authorities/names/nr95023741>

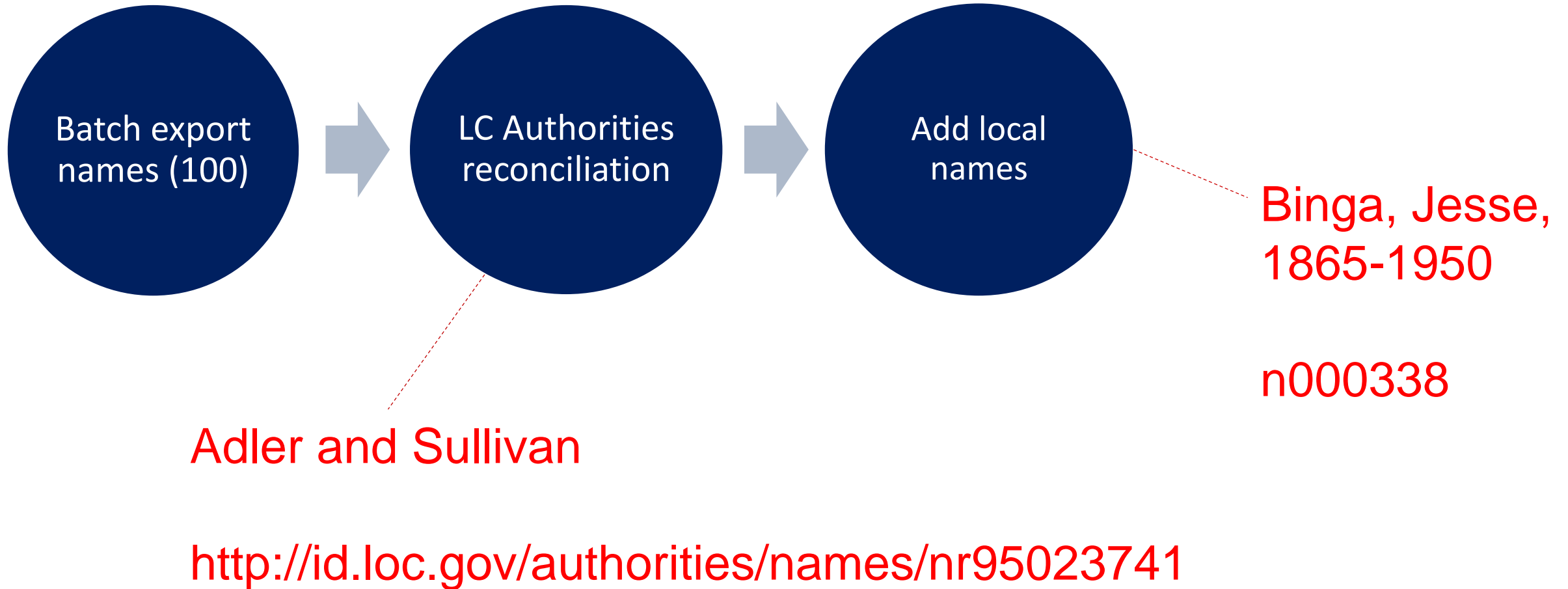
The Plan



Adler and Sullivan

<http://id.loc.gov/authorities/names/nr95023741>

The Plan



Reconcile column "corpname (new)"

Freebase Query-based
Reconciliation

LoC Reconciliation Service

LC (by way of VIAF)

Pick a Service or Extension on Left

<https://github.com/cmh2166/lc-reconcile>

Add Standard Reconciliation Service

Enter the service's URL:

<https://lc-reconcile.herokuapp.com/>

Add Service

Cancel

Reconcile column "corpname (new)"

» [Access Service API](#)

Freebase Query-based Reconciliation

LoC Reconciliation Service

LC (by way of VIAF)

Reconcile each cell to an entity of one of these types: Also use relevant details from other columns:

- Names
- Subjects
- LoC

Column	Include?	As Property
unitid	<input type="checkbox"/>	
unittitle	<input type="checkbox"/>	
corpname	<input type="checkbox"/>	
corpnameid	<input type="checkbox"/>	
persname	<input type="checkbox"/>	
persnameid	<input type="checkbox"/>	
creatorrole	<input type="checkbox"/>	
rules	<input type="checkbox"/>	
source	<input type="checkbox"/>	
bulk	<input type="checkbox"/>	
inclusive	<input type="checkbox"/>	
single	<input type="checkbox"/>	

Reconcile against type:

Reconcile against no particular type

Auto-match candidates with high confidence

Add Standard Service

Add Namespaced Service

Start Reconciling

Cancel

Post-Reconciliation

Google refine authorities_part4_011316.xlsx Permalink

Facet / Filter Undo / Redo 1

Refresh Reset All Remove All

Browse Term (new): judgment change

2 choices Sort by: name count

matched 23
none 77

Facet by choice counts

Browse Term (new): best candidate's score change reset

44.00 — 101.00

Numeric 45 Non-numeric 0 Blank 0 Error 55

100 rows

Show as: rows records Show: 5 10 25 50 rows

w/Hist	Birth Year	Browse Term	Browse Term (new)
		Moholy, Lucia edit	Moholy, Lucia Choose new match
		Moholy-Nagy, Hattula	Moholy-Nagy, Hattula Choose new match
		Moholy-Nagy, Laszlo	Moholy-Nagy, Laszlo <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Moholy-Nagy, László, 1895-1946 (73) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Moholy-Nagy, László, 1895-1946 (73) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Moholy-Nagy, László, 1895-1946. Works. Selections (52) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic
		Moholy-Nagy, Sibyl	Moholy-Nagy, Sibyl <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Moholy-Nagy, Sibyl, 1903-1971 (77) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Moholy-Nagy, Sibyl, 1903-1971 (77) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic
		Moldauer, John	Moldauer, John <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic

Matches

Suggested matches

Unmatched

Adding URIs

141 matching rows (1291 total)


Add column based on column persname

New column name

On error set to blank store error copy value from original column

Expression Language

No syntax error.



Preview History Starred Help

row	value	cell.recon.match.id
18.	Summers, Gene R.	http://id.loc.gov/authorities/names/nr94033253
22.	Illinois Institute of Technology. College of Architecture	http://id.loc.gov/authorities/names/no2014014836
845.	Greene-Mercier, Marie Zoe, 1911-2001	http://id.loc.gov/authorities/names/no2006027282
846.	Dearstyne, Howard	http://id.loc.gov/authorities/names/n85135755
851.	Aperture, Inc	http://id.loc.gov/authorities/names/n83176342
852.	Kessler, William H. (William Henry), 1924-2002	http://id.loc.gov/authorities/names/nr2006032643

OK Cancel

Adding URIs

Google refine collections_noRecordGroups_URIs_101to1300 csv Permalink

Facet / Filter Undo / Redo 16

Refresh Reset All Remove All

141 matching rows (1291 total)

Show as: rows records Show: 5 10 25 50 rows

Facet: Name: judgment change invert reset
3 choices Sort by: name count
(blank) 1040
matched 141 exclude
none 110
Facet by choice counts

Facet: Name: best candidate's score change reset
52.00 — 101.00
 Numeric 141 Non-numeric 0 Blank 0 Error 0

All	unitid	unittitle	Name	Name ID
☆	18.	006.01.03	Tutorial Unit Collection Summers, Gene R. Choose new match	http://id.loc.gov/authorities/names/nr94033253
☆	22.	006.01.02	Grade Reports Collection Illinois Institute of Technology. College of Architecture Choose new match	http://id.loc.gov/authorities/names/no2014014836
☆	845.	2005.022	Marie Zoe Greene-Mercier Papers Greene-Mercier, Marie Zoe, 1911-2001 Choose new match	http://id.loc.gov/authorities/names/no2006027282
☆	846.	1998.275	Howard Dearstyne Papers Dearstyne, Howard Choose new match	http://id.loc.gov/authorities/names/n85135755
☆	851.	2010.027	Institute of Design Collection Aperture, Inc. Choose new match	http://id.loc.gov/authorities/names/n83176342
☆	852.	2003.036	William Kessler Papers Kessler, William H. (William Henry), 1924-2002 Choose new match	http://id.loc.gov/authorities/names/nr2006032643

The Plan: local names



Local names

IIT Name Authority ☆

File Edit View Insert Format Data Tools Add-ons Help Last edit was 16 minutes ago

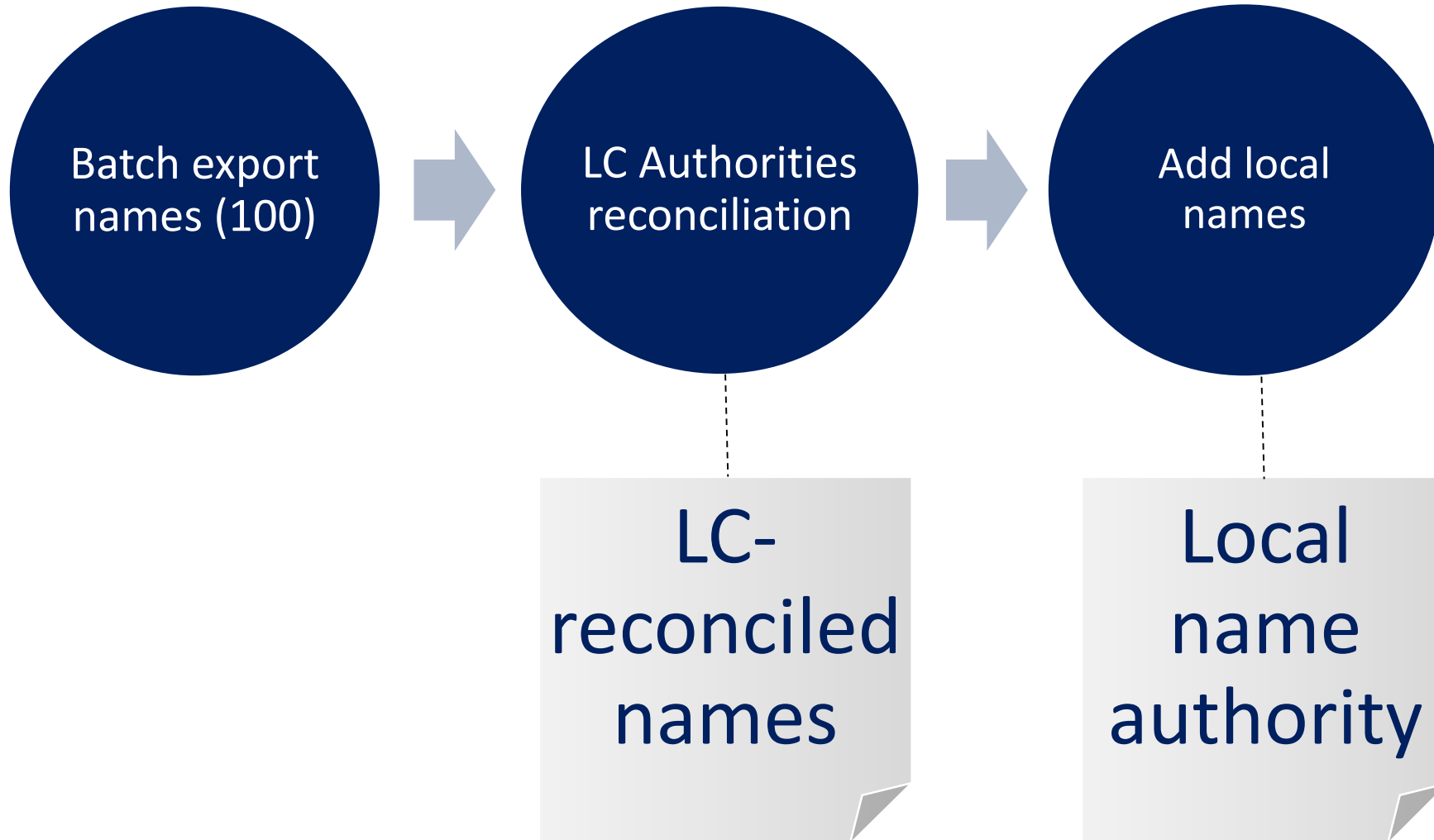
330 North Wabash (Chicago, Ill.)

	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

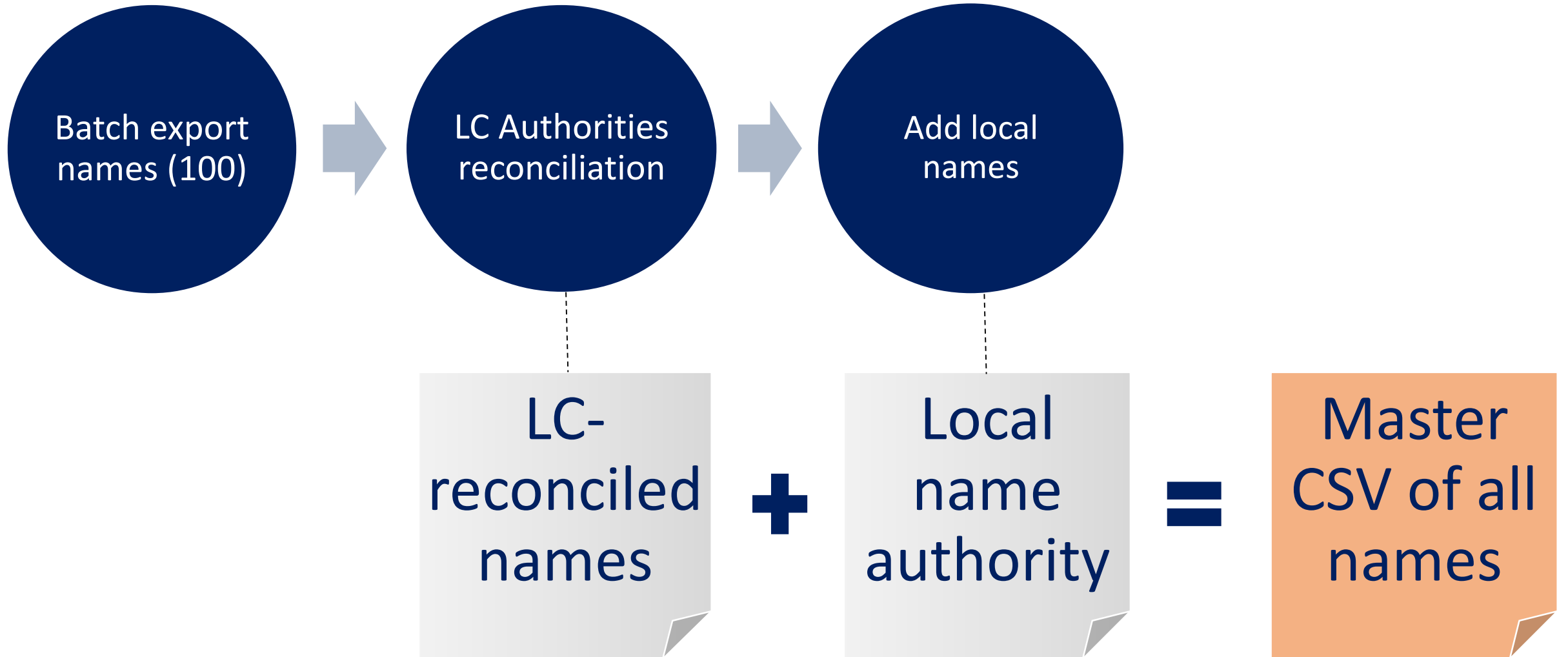
reconcile-CSV

- reconcile-csv matches one dataset against another
- Found at: <http://okfnlabs.org/reconcile-csv/>
- Need CSV file + java + OpenRefine

The Plan



The Plan



Master CSV

IIT Name Authority ☆

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

fx

	A	B	C
1	Agent Type	Primary Name	Authority ID
283	agent_person	Callahan, Harry M.	http://id.loc.gov/authorities/names/n50032177
284	agent_person	Moholy-Nagy, Sibyl, 1903-1971	http://id.loc.gov/authorities/names/n50001336
285	agent_corporate_entity	Illinois Central Railroad Company	http://id.loc.gov/authorities/names/n50000516
286	agent_person	Evon, Darcy L.	http://id.loc.gov/authorities/names/n2014076796
287	agent_person	Grubiak, Margaret M.	http://id.loc.gov/authorities/names/n2013065890
288	agent_person	Gordon, Robert Philip	http://id.loc.gov/authorities/names/n2009025113
289	agent_person	Barrett, Paul F., 1943-2004	http://id.loc.gov/authorities/names/n2006008278
290	agent_person	Dabbert, James R.	http://id.loc.gov/authorities/names/n2005087261
291	agent_corporate_entity	2x4 (Design Studio)	http://id.loc.gov/authorities/names/n2005016695
292	agent_person	Boder, David P. (David Pablo), 1886-1961	http://id.loc.gov/authorities/names/n00114984
293	agent_corporate_entity	Illinois Institute of Technology. Manufacturing/Industrial Programs	n000035
294	agent_corporate_entity	Illinois Institute of Technology. Mies van der Rohe Society	n000036
295	agent_corporate_entity	Illinois Institute of Technology. National Commission for IIT	n000358
296	agent_corporate_entity	Illinois Institute of Technology. Naval Reserve Officers Training Corps	n000037
297	agent_corporate_entity	Illinois Institute of Technology. Naval Reserves Officers Training Corps	n000331
298	agent_corporate_entity	Illinois Institute of Technology. Office of Alumni Relations	n000038
299	agent_corporate_entity	Illinois Institute of Technology. Office of Auxiliary Services	n000124
300	agent_corporate_entity	Illinois Institute of Technology. Office of Building and Grounds	n000332
301	agent_corporate_entity	Illinois Institute of Technology. Office of Campus and Conference Centers	n000039
302	agent_corporate_entity	Illinois Institute of Technology. Office of Campus Life	n000125
303	agent_corporate_entity	Illinois Institute of Technology. Office of Co-operative Education	n000040

reconcile-csv Results

- reconcile-csv matches one dataset against another
- Out of 100 names: 56 matches, 44 suggested matches- 100%
- 15 minutes per 100 records
- Can also pull in IDs (or any other column) from CSV

Subject CSV

IIT Name Authority ☆

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

fx <http://id.loc.gov/authorities/subjects/sh85028363>

	A	B
1	Term	Authority ID
2	Academic libraries	http://id.loc.gov/authorities/subjects/sh85076584
3	African Americans	http://id.loc.gov/authorities/subjects/sh85001932
4	African Americans college students	http://id.loc.gov/authorities/subjects/sh85001821
5	Amateur radio stations	http://id.loc.gov/authorities/subjects/sh85004123
6	Animated films	http://id.loc.gov/authorities/subjects/sh85088032
7	Anniversaries	http://id.loc.gov/authorities/subjects/sh85005357
8	Architecture	http://id.loc.gov/authorities/subjects/sh85006611
9	Architecture students	http://id.loc.gov/authorities/subjects/sh2007006992
10	Architecture--Designs and plans	http://id.loc.gov/authorities/subjects/sh85006624
11	Art--Exhibitions	http://id.loc.gov/authorities/subjects/sh85007479
12	Atomic bomb	http://id.loc.gov/authorities/subjects/sh85009310
13	Awards	http://id.loc.gov/authorities/subjects/sh00007725
14	Balls (Parties)	http://id.loc.gov/authorities/subjects/sh85011340
15	Baptists	http://id.loc.gov/authorities/subjects/sh85011739
16	Baseball	http://id.loc.gov/authorities/subjects/sh85012026
17	Basketball	http://id.loc.gov/authorities/subjects/sh85012111
18	Battery charging stations (Electric vehicles)	http://id.loc.gov/authorities/subjects/sh2012003590
19	Branding (Marketing)	http://id.loc.gov/authorities/subjects/sh2007006470
20	Campus planning	http://id.loc.gov/authorities/subjects/sh85019262
21	Chemical engineering	http://id.loc.gov/authorities/subjects/sh85022900
22	Chemistry	http://id.loc.gov/authorities/subjects/sh85022986
23	Christmas	http://id.loc.gov/authorities/subjects/sh85025303

Conclusion

- Took the most time = Evaluating LC matches + creating local names
- Took the least time = CSV reconciliation
- Next time: concentrate volunteer time on authorities reconciliation, not CSV matching.
- Great for linked data!

Agents

Showing 11 - 20 of 854 Results Sort by: Agent Name Ascending and: Select

Record Type	Agent Name	Authority ID	Source	Rules		
<input type="checkbox"/> Corporate Entity	Alion Science and Technology, Inc.	http://id.loc.gov/authorities/names/no2010153047	NACO Authority File		Created by greer 2016-04-05 13:41:49 -0500 Last Modified by greer 2017-01-14 17:43:50 -0600	View Edit
<input type="checkbox"/> Corporate Entity	Alpha Phi Omega	http://id.loc.gov/authorities/names/n85306914	NACO Authority File		Created by greer 2016-09-13 12:32:39 -0500 Last Modified by greer 2017-01-14 17:55:38 -0600	View Edit
<input type="checkbox"/> Person	Alschuler, Alfred S. (Alfred Samuel), 1876-1940	http://id.loc.gov/authorities/names/nr90000030	NACO Authority File		Created by greer 2016-09-13 12:04:02 -0500 Last Modified by greer 2017-01-14 17:56:20 -0600	View Edit
<input type="checkbox"/> Person	Altschuler, Dori	n000001	Local sources	Resource Description and Access	Created by greer 2016-02-23 14:04:16 -0600 Last Modified by sebastian 2016-05-05 13:43:36 -0500	View Edit
<input type="checkbox"/> Corporate Entity	Alumni Memorial Hall (Chicago, Ill.)	n000343	Local sources		Created by greer 2016-09-13 12:32:41 -0500 Last Modified by greer 2016-11-01 16:36:04 -0500	View Edit
<input type="checkbox"/> Corporate Entity	American Association of Retired Persons	http://id.loc.gov/authorities/names/n79113955	NACO Authority File		Created by greer 2016-04-05 15:34:48 -0500 Last Modified by greer 2017-01-14 17:57:06 -0600	View Edit

Thank you!

Greer Martin

gmarti15@iit.edu