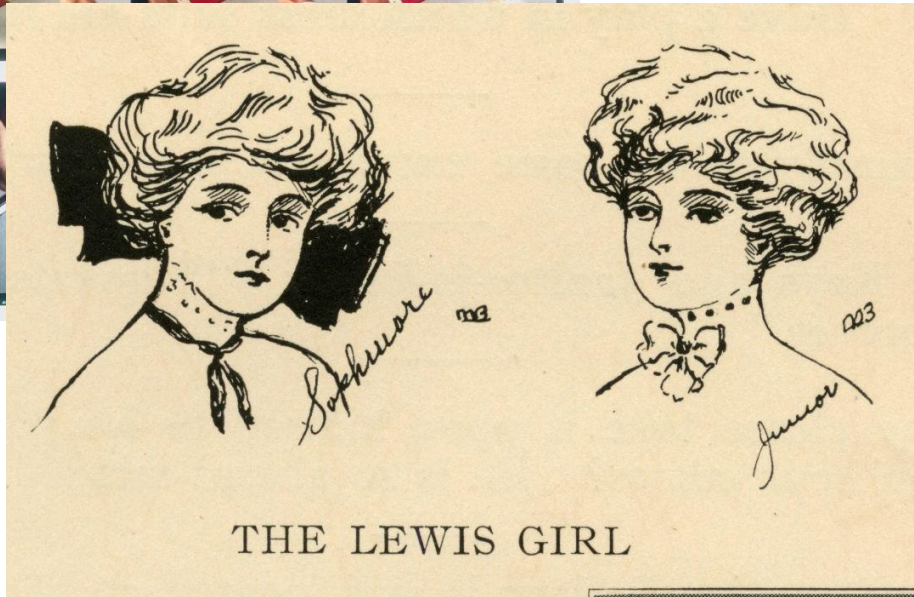


Migrating to ArchivesSpace with OpenRefine

Greer Martin
Paul V. Galvin Library
Illinois Institute of Technology
March 24, 2018

University Archives & Special Collections

- Established 1998
- 1,500 collections, mostly university records
- 4 full time staff
- Use Re:Discovery Proficio for archival records management system
- Planning for migration to ArchivesSpace began in 2015



Re:Discovery public interface

▼ Search

- [HEALD \(IIT Archives Catalog\)](#)
- [This Website](#)
- [Collection Guides](#)
- [Indexes](#)

› About Us

› Exhibits & Collections

› University History

[Home](#) > [Search](#) > [ArchiveSearch](#)

HEALD (IIT Archives Catalog)

[Search](#) | [Browse](#) | [Help](#)

Search IIT Archives Catalog

Collection Nbr: 004.03.01

Collection Title: MMAE Department Collection

Inclusive Dates: 1969- 1977

Bulk Dates:

Extent: [Extent]7 [Unit]Folders

Language:

Appraisal:

Provenance: taken from 1998.130

Item Status: Available for Research

Notes:

Summary:

Creator: Department of Mechanical, Materials and Aerospace Engineering

Creator Role:

Bio/Org History:

1. Metallurgical and Materials Engineering Program Info, 1972- 1977 2. Mechanical Engineering Masters Program, 1969 3. Metallurgical Engineering Graduate Programs, c.1969 4. MMAE Graduate Program Info, 1970s; poster for Metallurgical and Materials Engineering Masters program, 1983 5. X-Ray Courses, 1970 6. Metallurgical and Materials Engineering Program info, n.d. 7. Dept Of Mechanical Materials and Aerospace Engineering, n.d.

Scope-Content:

Organiz/Arrange:

Itemized Listing:

Assoc Materials:

Pers Name:

Corp Name:

Topic Term: Mechanical Engineering --Materials Engineering --Aerospace Engineering (SEE ALSO: Mechanical, Materials, and Aerospace Engineering)

Functions/Act.:

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

Migration

OpenRefine

- Metadata cleanup
- Authorities reconciliation
- JSON creation

ArchivesSpace API

- Import records



Facet / Filter Undo / Redo

400 rows

Extension

Refresh Reset All Remove All

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

< first < previous 251 - 300 next > last

Browse Term change

- 400 choices Sort by: name count Cluster
- Animation 1
 - Annual Reports-Paul V. Galvin Library 1
 - Antarctic Expedition 1
 - Anthropology 1
 - Aquitecture 1
 - Arcade Building 1
 - Architecture Classes--Institute of Design 1
 - Architecture of Mies van der Rohe, The (Michael Blackwood film) 1 edit include
 - Architecture program-AIT 1
 - Architecture program-IIT 1
 - Architecture students 1
 - Architecture with Technology: A New Synthesis (symposium) 1
 - Armour Flats 1
 - Armour Institute of Technology--buildings 1
 - Armour Mission (building) 1
 - Armour News, The 1
 - Armour Research Foundation--Personnel 1
 - Armour Tech Athletic Association 1
 - Army ROTC (SEE ALSO: Reserve Officers Training Corps) 1
 - Art & Sculpture 1
 - art@IIT 1
 - Athletics 1
 - Atomic Bomb 1

All	Added By	Added Date	Bio/Hist	Birth Year	Browse Term	Corporate Name	Corporate Name	Corporate Name	Corporate Name
					Alumni Memorial Hall)				
☆	cbruck	1/13/2010 10:14			Alumni Memorial Hall				
☆	cbruck	2/3/2015 10:00			Alumni-Armour Institute of Technology				
☆	cbruck	1/14/2010 11:45			Alumnifest				
☆	cbruck	12/1/2011 14:56			Alumni-ID				
☆	cbruck	1/13/2010 16:32			Alumni-IIT				
☆	cbruck	8/3/2010 14:22			American Institute of Electrical Engineers (AIEE) Chicago Chapter				
☆	cbruck	3/3/2014 13:09			American Scene (TV and radio series)				
☆	cbruck	2/25/2015 15:07			Animation				
☆	cbruck	12/16/2011 9:35			Annual Reports-Paul V. Galvin Library				
☆	RED	12/4/2009 14:08			Antarctic Expedition				
☆	cbruck	12/3/2010 15:39			Anthropology				
☆	cbruck	10/22/2010 11:54			Aquitecture				
☆	cbruck	10/27/2011 15:43			Arcade Building				
☆	cbruck	7/18/2011 15:51			Architecture Classes--Institute of Design				
☆	cbruck	6/4/2012 13:53			Architecture of Mies van der Rohe, The (Michael Blackwood film)				
☆	cbruck	11/7/2014 15:24			Architecture program-AIT				
☆	RED	12/4/2009 14:08			Architecture program-IIT				
☆	cbruck	1/8/2010 10:38			Architecture students				
☆	cbruck	5/14/2010 9:33			Architecture with				

Post-reconciliation

Google refine authorities_part4_011316.xlsx [Permalink](#)

Facet / Filter Undo / Redo 1 **100 rows**

Refresh Reset All Remove All Show as: rows records Show: 5 10 25 50 rows

v/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

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

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.

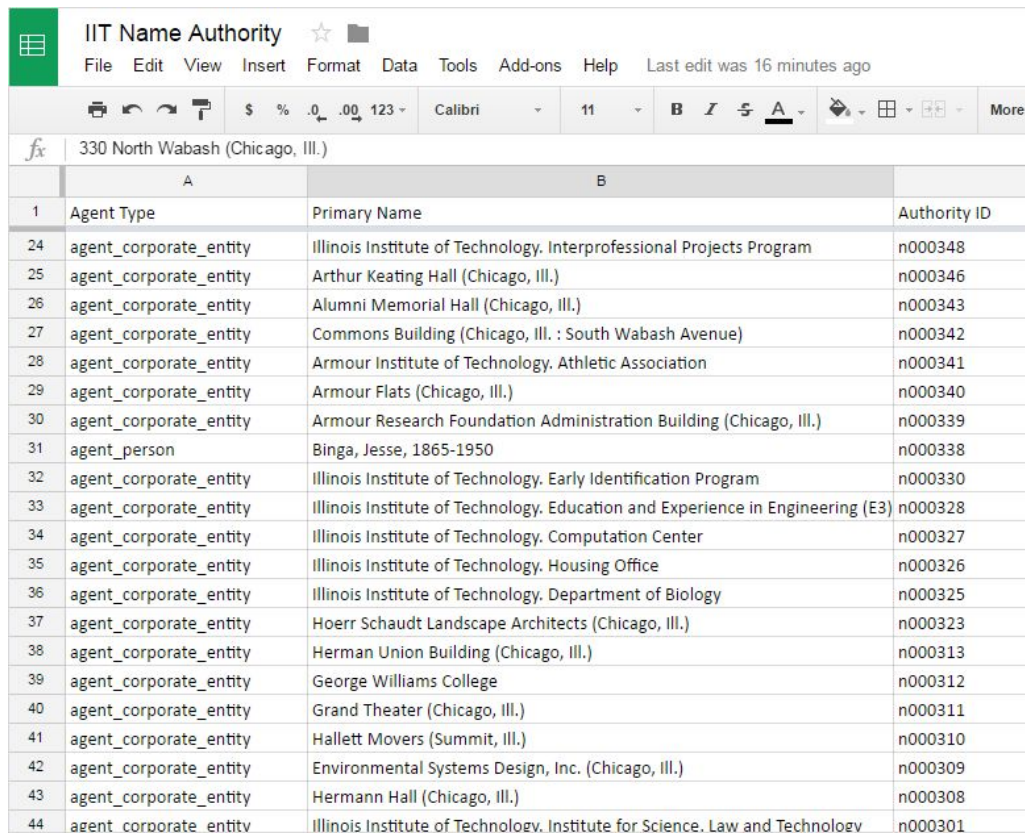
Preview History Starred Help

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

Reconcile-csv

- OpenRefine reconciliation service that matches one dataset against another
- Found at: <http://okfnlabs.org/reconcile-csv/>
- Good for adding local metadata



The screenshot shows the OpenRefine interface with a reconciliation table. The table has columns for Agent Type, Primary Name, and Authority ID. The data rows list various entities from the Illinois Institute of Technology, such as 'Illinois Institute of Technology. Interprofessional Projects Program' and 'Arthur Keating Hall (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

ArchivesSpace API

What is it?

- Set of instructions for importing and retrieving data without using the interface

Why use it?

- Data not in the importing tool's required format
- Batch importing

XML

```
<person>
  <firstName>John</firstName>
  <lastName>Smith</lastName>
  <age>25</age>
  <address>
    <streetAddress>21 2nd Street</streetAddress>
    <city>New York</city>
    <state>NY</state>
    <postalCode>10021</postalCode>
  </address>
  <phoneNumber>
    <type>home</type>
    <number>212 555-1234</number>
  </phoneNumber>
  <phoneNumber>
    <type>fax</type>
    <number>646 555-4567</number>
  </phoneNumber>
  <gender>
    <type>male</type>
  </gender>
</person>
```

JSON

```
{
  "firstName": "John",
  "lastName": "Smith",
  "age": 25,
  "address": {
    "streetAddress": "21 2nd Street",
    "city": "New York",
    "state": "NY",
    "postalCode": "10021"
  },
  "phoneNumber": [
    {
      "type": "home",
      "number": "212 555-1234"
    },
    {
      "type": "fax",
      "number": "646 555-4567"
    }
  ],
  "gender": {
    "type": "male"
  }
}
```

Schemas

- Map name data to ArchivesSpace Agent schema
- Different schemas for person and corporate agents
- JSON-ize records using OpenRefine

JSONModel(:agent_corporate_entity)

```
uri          string
title        string
is_linked_to_published_record boolean
agent_type   enum string agent_person agent_corporate_entity
              agent_software agent_family user
agent_contacts array JSONModel(:agent_contact) object
linked_agent_roles array string
external_documents array JSONModel(:external_document) object
rights_statements array JSONModel(:rights_statement) object
system_generated boolean
notes        array [object Object]
dates_of_existence array JSONModel(:date) object
publish      boolean
lock_version integer,string
jsonmodel_type string
created_by   string
```

Templating Export

Prefix

Row Template

```
{
  "jsonmodel_type": "agent_corporate_entity",
  "agent_contacts": [
    {
      "jsonmodel_type": "agent_contact",
      "telephones": [
        {
          "jsonmodel_type": "telephone",
          "number_type": "cell",
          "number": "{{jsonize(cells["agent_contact_tel",
          "ext": "{{jsonize(cells["agent_contact_teleph",
        }
      ]
      "name": "{{jsonize(cells["agent_contact_name"].va",
      "address_1": "{{jsonize(cells["agent_contact_addr",
      "address_2": "{{jsonize(cells["agent_contact_addr",
```

Row Separator

Suffix

Reset Template

```
],
  "notes": [
  ],
  "names": [
    {
      "authority_id": "n000024",
      "jsonmodel_type": "name_corporate_entity",
      "qualifier": null,
      "source": "Local sources",
      "rules": null,
      "authorized": false,
      "is_display_name": false,
      "sort_name_auto_generate": true,
      "primary_name": "Illinois Institute of Technol",
      "subordinate_name_1": null,
      "subordinate_name_2": null,
      "number": null,
      "sort_name": null,
      "description_note": null,
      "description_citation": null
    }
  ],
  "agent_type": "agent_corporate_entity",
}
{
  "jsonmodel_type": "agent_corporate_entity",
  "agent_contacts": [
```

Export

Cancel

POST Agents



- Consult API for POST Corporate Agents command
- Post one at a time, or build a script that will post them all



```
curl -H "X-ArchivesSpace-Session: $SESSION"
-d {
  "jsonmodel_type": "agent_corporate_entity",
  "agent_contacts": [
    {
      "jsonmodel_type": "agent_contact",
      "telephones": [
        {
          "jsonmodel_type": "telephone",
          "number_type": "cell",
          "number": "753 25202 1654 07054",
          "ext": "MHDPA"
        }
      ],
      "name": "Name Number 5",
      "address_1": "4H735534V",
      "country": "82691V643281",
    },
    "agent_type": "agent_corporate_entity"
  }
  'http://localhost:8089/agents/corporate_entities'
```

Conclusion

- Structured data = migration is possible!
- OpenRefine for cleaning, reconciliation, JSON export
- API for importing without using interface



The screenshot shows the 'Agents' interface in OpenRefine. At the top, it says 'Showing 11 - 20 of 854 Results' and 'Sort by: Agent Name Ascending'. Below this is a table with columns: Record Type, Agent Name, Authority ID, Source, Rules, and a column for creation/modification dates. Each row has a checkbox on the left and 'View' and 'Edit' buttons on the right.

Record Type	Agent Name	Authority ID	Source	Rules	
<input type="checkbox"/> Corporate Entity	Allon 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
<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 17:56:38 -0600
<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
<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
<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
<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

Thank you!

gmart15@iit.edu

Resources:

http://bit.ly/mac2018_martin