

Cleaning up legacy metadata for ETDs: Strategies, tools and a look into the future

Xiping Liu, Albert Duran, Anne Washington
USETDA, Columbus, OH 2016

Overview

Background

Tools

Processes

Next Steps







Image: Computer by Edward Boatman from the Noun Project

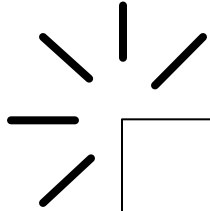
ETDs at UH

- 200 programs
- 1,270 ETDs since 2010
- +150 each Spring/Fall



Image: degree by Wilson Joseph from the Noun Project

Issues



Dictionary of Texas Digital Library Descriptive Metadata for Electronic
Theses and Dissertations, v. 2

September 2015

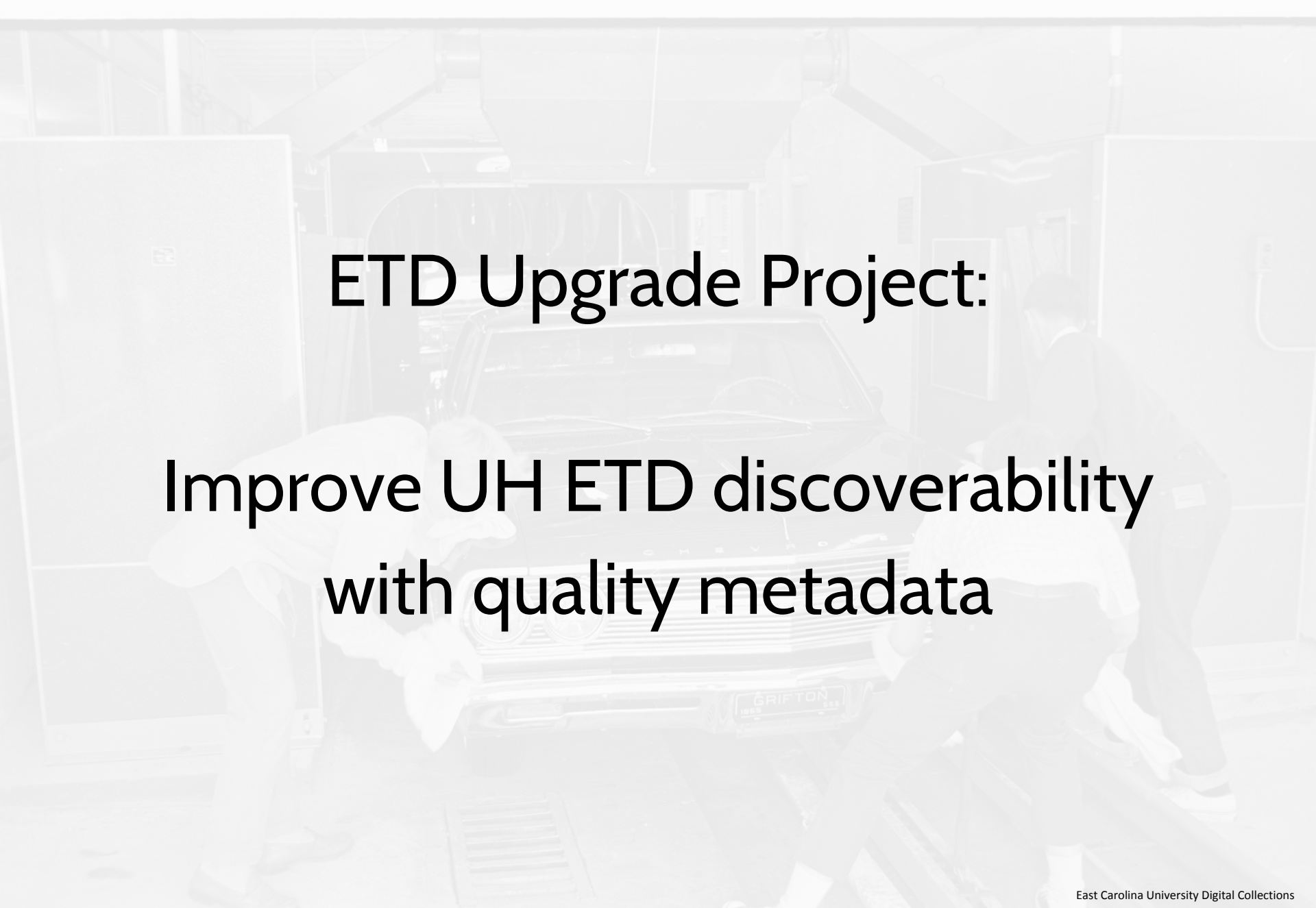
Prepared by the TDL ETD Metadata Working Group

Sarah Potvin, Chair (Texas A&M University)
Kara Long (Baylor University)
Colleen Lyon (University of Texas)
Kristi Park (Texas Digital Library)
Monica Rivero (Rice University)
Santi Thompson (University of Houston)

Note: This dictionary should be read in tandem with the *Report for Texas Digital Library Descriptive Metadata for Electronic Theses and Dissertations, v. 2*.

- Names
- Dates
- URLs





ETD Upgrade Project:

Improve UH ETD discoverability with quality metadata

East Carolina University Digital Collections

Plan

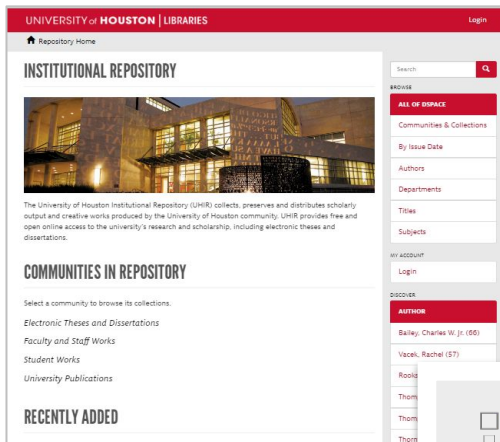
- Fix issues
- Comply with TDL Guidelines
- Update software configurations
- Document processes



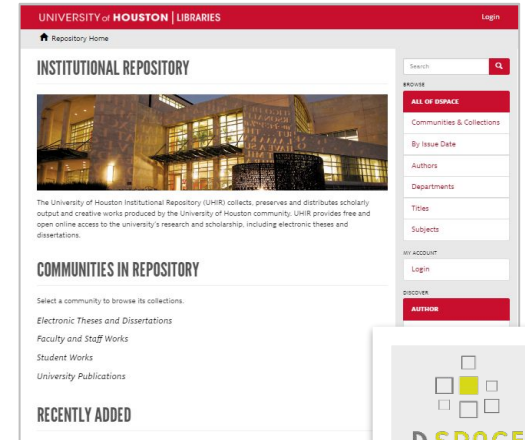
Image: strategy by Ryan Beck from the Noun Project
http://www.hddaviscpas.pro/wp-content/uploads/2014/08/coach_with_clipboard_400_clr_4728.png

Workflow

VIREO

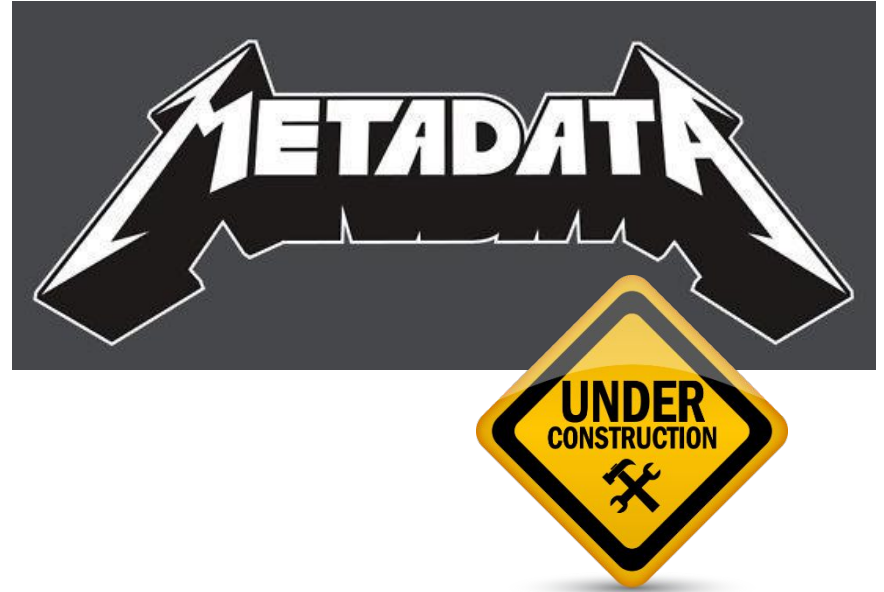


Refine



Strategies

- Communication
- Documentation
- Remediation



<http://www.hipsterpig.com/wp-content/uploads/2014/05/fc550x550asphalt3.jpg>
<http://www.groundtechwales.co.uk/wp-content/uploads/2013/07/underconstruction.png>

Tools

Communication



Basecamp


Documentation



Remediation



Communication

 **Basecamp 2** | Projects | Calendar | Everything | Progress | Everyone | Me

Metadata Unit

[See all to-do lists](#)


ETD Upgrade

- ☐ Add UH Departments List to Vireo [2 comments](#) [albert duran](#)
- ☐ Export DSpace Metadata and Correct with Excel & OpenRefine [1 comment](#) [albert duran](#)
- ☐ Work with TDL to Re-Import Corrected Metadata [Santi Thompson](#)
- ☐ Document ETD Upgrade Workflows [1 comment](#) [Xiping Liu](#)
- ☐ Test Metadata Import [1 comment](#) [Xiping Liu · Fri, Apr 8](#)
- ☐ USETDA Presentation: Albert slides [2 comments](#) [albert duran · Fri, Jul 22](#)
- ☐ USETDA Presentation: Xiping slides [1 comment](#) [Xiping Liu · Fri, Jul 22](#)
- ☐ USETDA Presentation: Anne slides [1 comment](#) [Anne Gaynor · Fri, Jul 22](#)

[Add a to-do](#)

- ☒ USETDA Poster Draft (Completed by Anne (Gaynor) Washington on 4 Aug)
- ☒ Copy ETD LCSH Subjects from Catalog to DSpace (Completed by Santi Thompson on 2 Jun)
- ☒ Clean Up ETD Subjects in DSpace (Completed by Santi Thompson on 2 Jun)
- ☒ Compile List of UH Departments [4 comments](#) (Completed by Santi Thompson on 2 Jun)
- ☒ Identify and Delete Duplicate Records [1 comment](#) (Completed by albert duran on 18 Aug)
- ☒ Review ETD Upgrade Issues Doc [1 comment](#) (Completed by Andrew Weidner on 5 Aug)
- ☒ Review TDL Descriptive Metadata Guidelines [1 comment](#) (Completed by Andrew Weidner on 5 Aug)
- ☒ Confirm DSpace Login Works [1 comment](#) (Completed by albert duran on 22 Jul)
- ☒ Vireo Metadata Check [1 comment](#) (Completed by Andrew Weidner on 1 May)
- ☒ Identify ETD Automation Items [2 comments](#) (Completed by Andrew Weidner on 3 Apr)

Discuss this to-do list

 **Basecamp 2** | Projects | Calendar | Everything | Progress | Everyone | Me

Metadata Unit

ETD Metadata Upgrade Project Meeting Agenda

August 4, 2016

Current progress

Once batches 1-10 are uploaded and the documentation is complete, the upgrade project is over. MDS will need to do regular imports of metadata as new ETDs are archived from Vireo to DSpace. MDS will also consider the list of "List of low-priority tasks that will remain after the upgrade project" and integrate into future work as resources allow. Santi will update MDS on the items in the "Fields to keep in consideration for the future (dependent on Vireo Updates)" list because their implementation depends on new Vireo developments.

The Scholarly Works Team has noticed crazy integration with DSpace and ORCID. They will be moving forward with adding ORCID ids to the dc.identifier.orcid field. Andy should add this field to the DSpace Metadata Registry before Albert can import batches 10 and 11.

Final steps:

- Albert and Xiping will complete documentation for updating ETD metadata. Someone from MDS will also complete batches 11 and 12 once the ORCID field is correctly integrated into DSpace.

List of low-priority tasks that will remain after the upgrade project:

- Delete word "Abstract" from the abstract field
- Deleting repeated values in the dc.date.accessioned and dc.date.available fields - contingent upon being able to export this data
- Revisiting the discussion on whether or not to change the order of field

Documentation

Metadata & Digitization Operations (MDS) Wiki

- Admin
 - Digital Projects
 - Project Template
 - Bayou City DAMS
- Digitization
 - Documentation
 - Equipment
 - Software
 - Social Media
 - Tools & Forms
 - Handling Guidelines
- Metadata
 - Documentation
 - Resources
 - Upgrade Project
 - Staff Pages
- Preservation
 - Dark Archive Workflow
 - Born Digital Workflow
 - Born Digital Log
- PmWiki
 - WikiSandbox
 - Basic Editing
 - Documentation Index
 - PmWiki FAQ
 - PmWikiPhilosophy

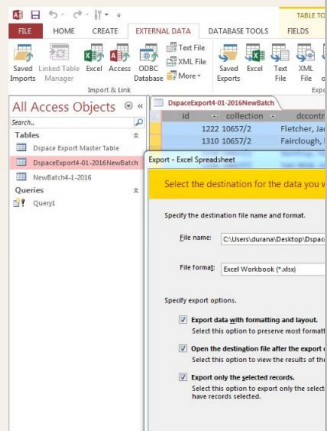
ETDMetadataUpgradeExportingNamesAndThenWorkingInGoogleRefin

Back to Metadata Documentation

Names Cleanup

Exporting from Access

1. First review names and correct any errors such as removing salutation
2. From the Access file, select the id, collection, dc.contributor.advisor
3. Under External Data click Excel and make sure that the file: 1. Op
4. Make sure to save



Metadata & Digitization Operations (MDS) Wiki

- Admin
 - Digital Projects
 - Project Template
 - Bayou City DAMS
- Digitization
 - Documentation
 - Equipment
 - Software
 - Social Media
 - Tools & Forms
 - Handling Guidelines
- Metadata
 - Documentation
 - Resources
 - Upgrade Project
 - Staff Pages
- Preservation
 - Dark Archive Workflow
 - Born Digital Workflow
 - Born Digital Log
- PmWiki
 - WikiSandbox
 - Basic Editing
 - Documentation Index
 - PmWiki FAQ
 - PmWikiPhilosophy

ETDMetadataUpgradeListOfCollegesDepartmentsAndDegreePrograms

Name of College	Name of Department	College URL	Department URL
C. T. Bauer College of Business	Accountancy and Taxation	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/accy/index
C. T. Bauer College of Business	Decision and Information Sciences	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/disc/index
C. T. Bauer College of Business	Finance	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/finance/index
C. T. Bauer College of Business	Management	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/management/index
C. T. Bauer College of Business	Marketing and Entrepreneurship	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/marketing/index
College of Education	Psychological, Health, and Learning Sciences	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/psych/index
College of Education	Curriculum and Instruction	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/ci/index
College of Education	Psychological, Health, and Learning Sciences	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/psych/index
College of Education	Psychological, Health, and Learning Sciences	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/psych/index
College of Education	Psychological, Health, and Learning Sciences	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/psych/index
College of Education	Educational Leadership and Policy Studies	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/edlps/index
College of Education	Educational Leadership and Policy Studies	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/edlps/index
College of Education	Curriculum and Instruction	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/ci/index
College of Education	Curriculum and Instruction	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/ci/index
College of Education	Psychological, Health, and Learning Sciences	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/psych/index
College of Education	Psychological, Health, and Learning Sciences	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/psych/index
College of Education	Psychological, Health, and Learning Sciences	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/psych/index
College of Education	Educational Leadership and Policy Studies	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/edlps/index
College of Education	Educational Leadership and Policy Studies	http://www.bauer.uh.edu	http://www.bauer.uh.edu/departments/edlps/index
Cullen College of Engineering	Mechanical Engineering	https://www.cullen.uh.edu	https://www.cullen.uh.edu/departments/me/index

Metadata & Digitization Operations (MDS) Wiki

- Admin
 - Digital Projects
 - Project Template
 - Bayou City DAMS
- Digitization
 - Documentation
 - Equipment
 - Software
 - Social Media
 - Tools & Forms
 - Handling Guidelines
- Metadata
 - Documentation
 - Resources
 - Upgrade Project
 - Staff Pages
- Preservation
 - Dark Archive Workflow
 - Born Digital Workflow
 - Born Digital Log
- PmWiki
 - WikiSandbox
 - Basic Editing
 - Documentation Index
 - PmWiki FAQ
 - PmWikiPhilosophy

ETDMetadataUpgrade

Back to Metadata Documentation

ETD Upgrade Documentation Home Page

- Updates from departments

#Lists of thesis with a signed front page

1. ID 213 <http://hdl.handle.net/10657/216>

Question	Answer	Need to Follow up (Y/N)	Contact	College/Dep
Did the Engineering Technology Department ever offer a program in "Construction Management"?	Construction Management started as a track under Engineering Technology before branching out to a separate program. This happened around 3 years ago.	Y	Susan and Edrel	Departments and Programs
#1114 Araji, Bahaa has Technology; #581 Nanisetti, Raja Revanth 1988- has Networks and Communications; #463 Bastamka, Ainhoa has NECO; #428 Ritchie, Earl J 1944- has Construction Management; #518 Barber, David has Material Science; #423 Cheng, Wei has System Control Technology; #582 Lu, Yang 1988- has Network Communications; #704 Velusamy, Gandhimathi has Network and Communication as their thesis degree discipline. What should be the correct terms?		Y	Susan and Edrel	Department of Engineering T
#461 Luker, Austin 1985- has Sequence Stratigraphy; #539 Carvalho Brito Campello de Azevedo, Laura 1982- has Oceanography; #302 Kimmler, Jeremy has Sedimentology; #471 Koroglu, Sebnem 1983- has Reservoir Characterization; #634 Unverzagt, Laura 1982- has Structural; #569 Li, Yangyang 1984- has Sedimentology and Stratigraphy; #609 Mead, Kimberly 1988- has Sedimentology; #689 Akhun, Sebnem 1984- has Seismic Interpretation in their thesis degree discipline field. What should be the correct terms?		Y		Department of Earth and At
Was Neuropsychology ever an option or should it be one of the following: 1. Psychology, Clinical 2. Psychology, Developmental 3. Psychology, Industrial and Organizational 4. Psychology, Social	Clinical Psychology	N	Anna	Psychology department
So it sounds like the master degree students have just "Psychology" and the PhD students	correct Master's in PSYC have no concentration and are Psychology. The	N	Anna	Psychology department

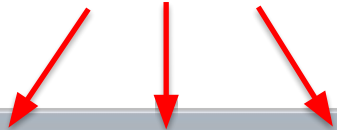
Remediation – MS Access

- Access queries used to:
 - Consolidate values from duplicate columns
 - Remove broken URLs
 - Add LCSH



Consolidate

Duplicate
columns



Update
query

DSpace Export 09-01-15

Query2

```

*
id
collection
dccontributoradvisor
dccontributoradvisor1
dccontributoradvisor2
dccontributorcommitteeMember
dcreator
dcreator1
ddatecreated
ddatecreated1
ddateissued
ddateissued1
ddatesubmitted
ddatesubmitted1
    
```

Field: dccontributoradvisor

Table: DSpace Export 09-01-15

Update To: [dccontributoradvisor1]

Criteria: Is Null

or:

Result

id	collection	dccontributoradvisor	dccontributoradvisor1
10	10657/2		
22	10657/2		
24	10657/2	Lee, Rebecca E.	
157	10657/2	Arbona, Consuelo	Arbona, Consuelo
159	10657/2	Arbona, Consuelo	Arbona, Consuelo
160	10657/2	Coleman, Nicole	Coleman, Nicole
161	10657/2	Chow, Diana	Chow, Diana
162	10657/2	Ryan, Michael	Ryan, Michael
165	10657/2	Yamasaki, Jill	Yamasaki, Jill
167	10657/2	Schwartz, Jonathan	Schwartz, Jonathan
168	10657/2	Mountain, Lee	Mountain, Lee
169	10657/2	Armstrong, Mary	Armstrong, Mary
170	10657/2	Arbona, Consuelo	Arbona, Consuelo
171	10657/2	Olson, Beth	Olson, Beth
172	10657/2	Jowett, Garth	Jowett, Garth
173	10657/2	Ryan, Michael	Ryan, Michael
174	10657/2	Vardeman-Winter, Jennif	Vardeman-Winter, Jennif
175	10657/2	Busch, Steven D.	Busch, Steven D.
176	10657/2	MacNeil, Angus J.	MacNeil, Angus J.
177	10657/2	Craig, Cheryl J.	Craig, Cheryl J.
178	10657/2	White, Cameron	White, Cameron
179	10657/2	Vardeman-Winter, Jennif	Vardeman-Winter, Jennif
180	10657/2	Liberman, Dr. Paul L. Goss	

Remove Broken Links

Broken URLs

dcidentifieruri	
http://hdl.handle.net/10657/417	
http://hdl.handle.net/10657/420	
http://hdl.handle.net/10657/421	
http://hdl.handle.net/10657/422	
http://hdl.handle.net/10657/ETD-UH-2010-05-34	http://hdl.handle.net/10657/423
http://hdl.handle.net/10657/ETD-UH-2010-08-97	http://hdl.handle.net/10657/424
http://hdl.handle.net/10657/ETD-UH-2010-12-98	http://hdl.handle.net/10657/425
http://hdl.handle.net/10657/ETD-UH-2011-05-100	http://hdl.handle.net/10657/426
http://hdl.handle.net/10657/ETD-UH-2011-05-104	http://hdl.handle.net/10657/427
http://hdl.handle.net/10657/ETD-UH-2011-05-106	http://hdl.handle.net/10657/428
http://hdl.handle.net/10657/ETD-UH-2011-05-118	http://hdl.handle.net/10657/429
http://hdl.handle.net/10657/ETD-UH-2011-05-122	http://hdl.handle.net/10657/430
http://hdl.handle.net/10657/ETD-UH-2011-05-126	http://hdl.handle.net/10657/431
http://hdl.handle.net/10657/ETD-UH-2011-05-127	http://hdl.handle.net/10657/432
http://hdl.handle.net/10657/ETD-UH-2011-05-136	http://hdl.handle.net/10657/433
http://hdl.handle.net/10657/ETD-UH-2011-05-143	http://hdl.handle.net/10657/434
http://hdl.handle.net/10657/ETD-UH-2011-05-147	http://hdl.handle.net/10657/437
http://hdl.handle.net/10657/ETD-UH-2011-05-149	http://hdl.handle.net/10657/438
http://hdl.handle.net/10657/ETD-UH-2011-05-154	http://hdl.handle.net/10657/439
http://hdl.handle.net/10657/ETD-UH-2011-05-161	http://hdl.handle.net/10657/444
http://hdl.handle.net/10657/ETD-UH-2011-05-163	http://hdl.handle.net/10657/445

Extract correct URL and update

DSpace Export 09-01-15 Query1

DSpace Export 09-01-15

dcembargoterms
dcembargoterms1
dcformatmimetype
dcformatmimetype1
dcidentifierslug
dcidentifierslug1
dcidentifieruri
dclanguageiso
dclanguageiso1
dclanguageiso2
dclanguageiso3
dclanguageiso4
dcsubject

Field: dcidentifieruri
Table: DSpace Export 09-01-15
Update To: Right([dcidentifieruri],31)
Criteria: Len([dcidentifieruri]) > 46
or:

Result

dcidentifieruri
http://hdl.handle.net/10657/417
http://hdl.handle.net/10657/420
http://hdl.handle.net/10657/421
http://hdl.handle.net/10657/422
http://hdl.handle.net/10657/423
http://hdl.handle.net/10657/424
http://hdl.handle.net/10657/425
http://hdl.handle.net/10657/426
http://hdl.handle.net/10657/427
http://hdl.handle.net/10657/428
http://hdl.handle.net/10657/429
http://hdl.handle.net/10657/430
http://hdl.handle.net/10657/431
http://hdl.handle.net/10657/432
http://hdl.handle.net/10657/433
http://hdl.handle.net/10657/434
http://hdl.handle.net/10657/437
http://hdl.handle.net/10657/438
http://hdl.handle.net/10657/439
http://hdl.handle.net/10657/444
http://hdl.handle.net/10657/445

Add LCSH

CREATOR	TITLE	LCSH	Local Subject Heading	URL
Å"verup, Camilla St	Who you are matters to	Social psychology.		http://hdl.handle.net/10657/ETD-UH-2012-05-
Accrocco, Nicholas J	Examination of upper scl		Professional leadership.	http://hdl.handle.net/10657/1020
Acosta, Jesus B.	The effect of the Texas s		Professional leadership.	http://hdl.handle.net/10657/608
Adejuyigbe, Olawal	Correlation between the	Geophysics.		http://hdl.handle.net/10657/466
Adekunle, Adebow	Spectral decomposition	Geophysics.		http://hdl.handle.net/10657/ETD-UH-2012-05-
Adelman, Chana R.	Factors that influence ag	Autism in children--Diagnosis. Child developm		http://hdl.handle.net/10657/ETD-UH-2010-08-
Afiesimama, Jane T	Teacher-as-researcher : i	Science--Study and teaching (Secondary) Scien		http://hdl.handle.net/10657/587
Agbayani, Kristina A	Patterns of age-related i	Clinical neuropsychology.		http://hdl.handle.net/10657/236
Agbayani, Kristina A	Cognition and depressio	Clinical psychology.		
Aggarwal, Aanchal,	Large deviations approac	Mathematics.		
Ahmed, Farid, 1979-	Environmental applicatio	Environmental engineering.		
Ahmed, Hazem.	A study of frequency dep	Geophysics.		
Ahmed, Owais.	Dataset modification to i	Computer science.		
Ahmed, Shaik M.	Understanding parental	School management and organization.		
Ahn, Jiseon.	The impact of brand equ		Global r	
Aich, Anupam.	A microfluidics approach			
Akbas, Omer.	Microseismic monitoring	Geophysics.		
Akerson, Adam G.	Survey of in-service earl	Early childhood teachers--In-service training. S		
Akhun, Selin.	Evolution of late cenozo			
Ako, Roland A.	Pharmacokinetics	Pharmacokinetics.		
Akujuobi, Christoph	The influences of partisa			
Alhaider, Ibrahim A	The protective effects of	Pharmacology.		
Ali, Quaisar.	Role of Angiotensin II Ty	Pharmacology.		
Ali, Saif.	Exchange rate forecastin		Applied	
Alleyne, Camille W	A comparative case stud		Professi	
Almes, Hardy S.	Stadium games: publicl	Forecasting.		
Almus, Kadir.	The beliefs of principals	Educational tests and measurements--United Sta		

Query to join tables on URL

The screenshot shows a database query interface with two tables: 'Final ETD table' and 'ETD_Subject Headings'. The 'Final ETD table' has columns: dcsujectother, dcdescriptionabstract, dcformatmimetype, dcformatdigitalOrigin, dcidentifierslug, **dcidentifieruri**, dclanguageiso, dcctype, dcctypegenre, dcctypematerial, dcctypedcmi, thesisdegreedepartment, thesisdegreecollege, dcdescriptiondepartment, thesisdegree discipline, thesisdegree major, thesisdegree grantor. The 'ETD_Subject Headings' table has columns: *, CREATOR, TITLE, LCSH, Local Subject Heading, **URL**. A join query is being created, joining the two tables on the 'URL' field. The 'Join Properties' dialog is open, showing the join type as 'Include ALL records from 'Final ETD table' and only those records from 'ETD_Subject Headings' where the joined fields are equal.' The dialog also shows the left table name as 'Final ETD table', the right table name as 'ETD_Subject Headings', the left column name as 'dcidentifieruri', and the right column name as 'URL'. The 'OK' button is highlighted.

Field:	Final ETD table.*	LCSH	Local Subject			
Table:	Final ETD table	ETD_Subject Hea	ETD_Subject Hea			
Sort:						
Show:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Criteria:						

Add LCSH

Query result

thesisdegreegrantor	thesisdegree	thesisdegree	dcrights	LCSH	Local Subject Heading
University of Houston	Doctoral	Doctor of Philo		Chemistry.	
University of Houston	Doctoral	Doctor of Philo		Educational psychology.	
University of Houston	Masters	Master of Educ		Counseling psychology.	
University of Houston	Doctoral	Doctor of Philo		Mathematics.	
University of Houston	Doctoral	Doctor of Philo		Biochemistry.	
University of Houston	Doctoral	Doctor of Philo		Biology.	
University of Houston	Doctoral	Doctor of Educ			Professional leadership
University of Houston	Doctoral	Doctor of Philo		Counseling psychology.	
University of Houston	Masters	Master of Scier		Electrical engineering.	
University of Houston	Masters	Master of Scier		Geology.	

Final table

thesisdegreegrantor	thesisdegree	thesisdegree	dcrights	dcsubjectlcsch	dcsubjectother
University of Houston	Doctoral	Doctor of Educ			Professional leadership
University of Houston	Doctoral	Doctor of Philo		Chemistry	
University of Houston	Doctoral	Doctor of Philo		Educational psychology	
University of Houston	Masters	Master of Educ		Counseling psychology	
University of Houston	Doctoral	Doctor of Philo		Mathematics	
University of Houston	Doctoral	Doctor of Philo		Biochemistry	
University of Houston	Doctoral	Doctor of Philo		Biology	
University of Houston	Doctoral	Doctor of Educ			Professional leadership
University of Houston	Doctoral	Doctor of Philo		Counseling psychology	
University of Houston	Masters	Master of Scier		Electrical engineering	
University of Houston	Masters	Master of Scier		Geology	

Remediation – Open Refine

- Advisor and committee member name clean-up

OPEN
Refine 



Image: University of Southern California Libraries. California Historical Society.

Name clean-up

Import

Show: 5 10 25 50 rows

id	dccontributoradvisor	dccontributorcommitteeMember
1	Olson, Beth	Haun, Martha Verheyen, Claremarie
2	Jowett, Garth	Houk, Keith R. Reed, John
3	Ryan, Michael	McCombs, Shawn Liu, Youmei
4	Vardeman-Winter, Jennifer	Heath, Robert Ni, Lan Liu, Youmei
5	Busch, Steven D.	
6	MacNeil, Angus J.	
7	Craig, Cheryl J.	
8	White, Cameron	
9	Vardeman-Winter, Jennifer	

id	dccontributoradvisor	dccontributorcommitteeMember
10	Phillips, Scott	
11	Lee, Rebecca E.	
12	Arbona, Consuelo	
13	Arbona, Consuelo	
14	Coleman, Nicole	
15	Chow, Diana	
16	Ryan, Michael	
17	Yamasaki, Jill	
18	Schwartz, Jonathan	Arbona, Consuelo Wiesner, David
19	Mountain, Lee	Abrahamson, Richard Craig, Cheryl Horn, Catherine
20	Armstrong, Mary	Greer, Gary Gaa, John Dao, Tam Andrews, Nicole
21	Arbona, Consuelo	Watson, Margaret Coleman, Nicole Armstrong, Mary
22	Olson, Beth	Haun, Martha Verheyen, Claremarie
23	Jowett, Garth	Houk, Keith R. Reed, John
24	Ryan, Michael	McCombs, Shawn Liu, Youmei
25	Vardeman-Winter, Jennifer	Heath, Robert Ni, Lan Liu, Youmei

Split multi-valued cells

Result

	id	collection	dccontributoradvisor	dccontributorcommitteeMember
1.	3	10657/2		
2.	10	10657/2		
3.	22	10657/2	Phillips, Scott	
4.	24	10657/2	Lee, Rebecca E.	Layne, Charles S.
5.				O'Connor, Daniel P.
6.				Rifai, Hanadi
7.	157	10657/2	Arbona, Consuelo	Burridge, Andrea
8.				Gaa, John
9.				Backus, Margot
10.	159	10657/2	Arbona, Consuelo	Day, Susan X.
11.				Armstrong, Mary
12.				Olvera, Norma
13.	160	10657/2	Coleman, Nicole	Arbona, Consuelo
14.				Mc Pherson, Robert
15.				Watson, Margaret
16.	161	10657/2	Chow, Diana	Yeung, Jim
17.				Bond, Richard
18.				Liang, Dong
19.				Giovannella, Bappino
20.	162	10657/2	Ryan, Michael	Vardeman, Jennifer
21.				Lewis, Dontee
22.	165	10657/2	Yamasaki, Jill	Haun, Martha

Name clean-up

Transpose cells across columns into rows

The screenshot shows a data table with columns: id, collection, dccontributoradvisor, and dccontributorcommittee. A context menu is open over the 'dccontributorcommittee' column, with the 'Transpose' option highlighted. Below the table is a dialog box titled 'Transpose Cells Across Columns into Rows'.

Transpose Cells Across Columns into Rows

From Column: id, collection, dccontributoradvisor, dccontributorcommittee

To Column: dccontributorcommittee (last column)

Transpose into:

- ☒ Two new columns
 - Key Column: Original Column (containing original columns' names)
 - Value Column: Name (containing original cells' values)
- ☐ One column

☐ prepend the original column's name to each cell followed by : before the cell's value

☒ Ignore blank cells

☐ Fill down in other columns

Buttons: Transpose, Cancel

Result

The resulting table has columns: id, collection, Original Column, and Name. The 'Name' column contains the transposed data. A red box highlights the 'Name' column, and a red arrow points to it.

	id	collection	Original Column	Name
1.	3	10657/2		
2.	10	10657/2		
3.	22	10657/2	dccontributoradvisor	Phillips, Scott
4.	24	10657/2	dccontributoradvisor	Lee, Rebecca E.
5.			dccontributorcommitteeMember	Layne, Charles S.
6.			dccontributorcommitteeMember	O'Connor, Daniel P.
7.			dccontributorcommitteeMember	Rifai, Hanadi
8.	157	10657/2	dccontributoradvisor	Arbona, Consuelo
9.			dccontributorcommitteeMember	Burridge, Andrea
10.			dccontributorcommitteeMember	Gaa, John
11.			dccontributorcommitteeMember	Backus, Margot
12.	159	10657/2	dccontributoradvisor	Arbona, Consuelo
13.			dccontributorcommitteeMember	Day, Susan X.
14.			dccontributorcommitteeMember	Armstrong, Mary
15.			dccontributorcommitteeMember	Olvera, Norma
16.	160	10657/2	dccontributoradvisor	Coleman, Nicole
17.			dccontributorcommitteeMember	Arbona, Consuelo
18.			dccontributorcommitteeMember	Mc Pherson, Robert
19.			dccontributorcommitteeMember	Watson, Margaret
20.	161	10657/2	dccontributoradvisor	Chow, Diana
21.			dccontributorcommitteeMember	Yeung, Jim
22.			dccontributorcommitteeMember	Bond, Richard
23.			dccontributorcommitteeMember	Liang, Dong
24.			dccontributorcommitteeMember	Giovanella, Bappino
25.	162	10657/2	dccontributoradvisor	Ryan, Michael

Name clean-up

Filter on advisor names

The screenshot shows a data table with 990 matching rows (3658 total). The table has columns: All, id, collection, Original Column, and Name. The 'Original Column' is filtered with the value 'advisor'. A context menu is open over the 'Text filter' option, showing options like 'Split into several columns...', 'Add column based on this column...', 'Add column by fetching URLs...', 'Rename this column', 'Remove this column', 'Move column to beginning', 'Move column to end', 'Move column left', and 'Move column right'.

Facet / Filter Undo / Redo 3

Refresh Reset All Remove All

990 matching rows (3658 total)

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

Original Column

advisor

☐ case sensitive ☐ regular expression

990 matching rows (3658 total)

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

All	id	collection	Original Column	Name
3.	22	10657/2	Facet	Phillips, Scott
4.	24	10657/2	Text filter	Lee, Rebecca E.
8.	157	10657/2	Arbona, Consuelo	Arbona, Consuelo
12.	159	10657/2	Arbona, Consuelo	Arbona, Consuelo
16.	160	10657/2	Coleman, Nicole	Coleman, Nicole
20.	161	10657/2	Chow, Diana	Chow, Diana
25.	162	10657/2	Ryan, Michael	Ryan, Michael
28.	165	10657/2	Yamasaki, Jill	Yamasaki, Jill
31.	167	10657/2	Schwartz, Jonathan	Schwartz, Jonathan
35.	168	10657/2	Mountain, Lee	Mountain, Lee
39.	169	10657/2	Armsworth, Mary	Armsworth, Mary
44.	170	10657/2	Arbona, Consuelo	Arbona, Consuelo
48.	171	10657/2	Olson, Beth	Olson, Beth
51.	172	10657/2	Jowett, Garth	Jowett, Garth
54.	173	10657/2	Ryan, Michael	Ryan, Michael
57.	174	10657/2	Vardeman-Winter, Jennifer	Vardeman-Winter, Jennifer
61.	175	10657/2	Busch, Steven D.	Busch, Steven D.
65.	176	10657/2	MacNeil, Angus J.	MacNeil, Angus J.
69.	177	10657/2	Craig, Cheryl J.	Craig, Cheryl J.
73.	178	10657/2	White, Cameron	White, Cameron
77.	179	10657/2	Vardeman-Winter, Jennifer	Vardeman-Winter, Jennifer

Result

The screenshot shows the resulting data table after the 'Text filter' operation. A new column, 'dccontributoradvisor', has been added to the table, containing the filtered values. The table has 990 matching rows (3658 total).

990 matching rows (3658 total)

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

All	id	collection	Original Column	Name	dccontributoradvisor
3.	22	10657/2	dccontributoradvisor	Phillips, Scott	Phillips, Scott
4.	24	10657/2	dccontributoradvisor	Lee, Rebecca E.	Lee, Rebecca E.
8.	157	10657/2	dccontributoradvisor	Arbona, Consuelo	Arbona, Consuelo
12.	159	10657/2	dccontributoradvisor	Arbona, Consuelo	Arbona, Consuelo
16.	160	10657/2	dccontributoradvisor	Coleman, Nicole	Coleman, Nicole
20.	161	10657/2	dccontributoradvisor	Chow, Diana	Chow, Diana
25.	162	10657/2	dccontributoradvisor	Ryan, Michael	Ryan, Michael
28.	165	10657/2	dccontributoradvisor	Yamasaki, Jill	Yamasaki, Jill
31.	167	10657/2	dccontributoradvisor	Schwartz, Jonathan	Schwartz, Jonathan
35.	168	10657/2	dccontributoradvisor	Mountain, Lee	Mountain, Lee
39.	169	10657/2	dccontributoradvisor	Armsworth, Mary	Armsworth, Mary
44.	170	10657/2	dccontributoradvisor	Arbona, Consuelo	Arbona, Consuelo
48.	171	10657/2	dccontributoradvisor	Olson, Beth	Olson, Beth
51.	172	10657/2	dccontributoradvisor	Jowett, Garth	Jowett, Garth
54.	173	10657/2	dccontributoradvisor	Ryan, Michael	Ryan, Michael
57.	174	10657/2	dccontributoradvisor	Vardeman-Winter, Jennifer	Vardeman-Winter, Jennifer
61.	175	10657/2	dccontributoradvisor	Busch, Steven D.	Busch, Steven D.
65.	176	10657/2	dccontributoradvisor	MacNeil, Angus J.	MacNeil, Angus J.
69.	177	10657/2	dccontributoradvisor	Craig, Cheryl J.	Craig, Cheryl J.
73.	178	10657/2	dccontributoradvisor	White, Cameron	White, Cameron
77.	179	10657/2	dccontributoradvisor	Vardeman-Winter, Jennifer	Vardeman-Winter, Jennifer
81.	180	10657/2	dccontributoradvisor	Liberman, Dr. Dov	Liberman, Dr. Dov
83.			dccontributoradvisor	Gore-Laird, Dr. Helen	Gore-Laird, Dr. Helen

Create new column

Name clean-up

New column for
committee members

3658 rows

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

All	id	collection	Original Column	Name	dccontributorcom	dccontributorad
1.	3	10657/2				
2.	10	10657/2				
3.	22	10657/2	dccontributoradviser	Phillips, Scott		Phillips, Scott
4.	24	10657/2	dccontributoradviser	Lee, Rebecca E.		Lee, Rebecca E.
5.			dccontributorcommitteeMember	Layne, Charles S.	Layne, Charles S.	
6.			dccontributorcommitteeMember	O'Connor, Daniel P.	O'Connor, Daniel P.	
7.			dccontributorcommitteeMember	Rifai, Hanadi	Rifai, Hanadi	
8.	157	10657/2	dccontributoradviser	Arbona, Consuelo		Arbona, Consuelo
9.			dccontributorcommitteeMember	Burridge, Andrea	Burridge, Andrea	
10.			dccontributorcommitteeMember	Gaa, John	Gaa, John	
11.			dccontributorcommitteeMember	Backus, Margot	Backus, Margot	

3658 rows

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

All	id	collection	dccontributorcommitteeMember	dccontributorad
1.	3	10657/2		
2.	10	10657/2		
3.	22	10657/2		Phillips, Scott
4.	24	10657/2	Layne, Charles S. O'Connor, Daniel P. Rifai, Hanadi	Lee, Rebecca E.
5.				
6.				
7.				
8.	157	10657/2	Burridge, Andrea Gaa, John Backus, Margot	Arbona, Consuelo
9.				
10.				
11.				
12.	159	10657/2	Day, Susan X. Armsworth, Mary Olvera, Norma	Arbona, Consuelo
13.				

Join values

Result

941 rows

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

All	id	collection	dccontributorcommitteeMember	dccontributoradviser
1.	3	10657/2		
2.	10	10657/2		
3.	22	10657/2		Phillips, Scott
4.	24	10657/2	Layne, Charles S. O'Connor, Daniel P. Rifai, Hanadi	Lee, Rebecca E.
5.	157	10657/2	Burridge, Andrea Gaa, John Backus, Margot	Arbona, Consuelo
6.	159	10657/2	Day, Susan X. Armsworth, Mary Olvera, Norma	Arbona, Consuelo
7.	160	10657/2	Arbona, Consuelo Mc Pherson, Robert Watson, Margaret	Coleman, Nicole
8.	161	10657/2	Yeung, Jim Bond, Richard Liang, Dong Giovannella, Bappino	Chow, Diana
9.	162	10657/2	Vardeman, Jennifer Lewis, Dontee	Ryan, Michael
10.	165	10657/2	Haun, Martha Addison Otey, Penny	Yamasaki, Jill
11.	167	10657/2	Arbona, Consuelo Wiesner, Margit Olvera, Norma	Schwartz, Jonathan
12.	168	10657/2	Abrahamson, Richard Craig, Cheryl Horn, Catherine	Mountain, Lee
13.	169	10657/2	Greer, Gary Gaa, John Dao, Tam Andrews, Nicole	Armsworth, Mary
14.	170	10657/2	Watson, Margaret Coleman, Nicole Armsworth, Mary	Arbona, Consuelo
15.	171	10657/2	Haun, Martha Verheyen, Claremarie	Olson, Beth
16.	172	10657/2	Houk, Keith R. Reed, John	Jowett, Garth
17.	173	10657/2	McCombs, Shawn Liu, Youmei	Ryan, Michael
18.	174	10657/2	Heath, Robert Ni, Lan Liu, Youmei	Vardeman-Winter, Jennifer
19.	175	10657/2	MacNeil, Angus J. Emerson, Michael W. Prater, Doris L.	Busch, Steven D.
20.	176	10657/2	Lieberman, David Emerson, Michael W. Waxman, Hersh	MacNeil, Angus J.

Import into DSpace

UNIVERSITY of HOUSTON | LIBRARIES

Repository Home / Import Metadata

IMPORT METADATA

Choose File Batch10_with_abstract.csv

Upload CSV

Notice
Upload successful

IMPORT METADATA

Pending changes are listed below for review

Changes pending for item 1131 (10657/1117)

Add: (dc.contributor.advisor)	Massman, Paul J.
Add: (dc.contributor.committeeMember)	Tian, T. Siva
Add: (dc.contributor.committeeMember)	Dulay, Mario F., Jr.
Add: (dc.date.created)	May 2015
Add: (dc.date.issued)	2015-05
Add: (dc.description.department)	Psychology
Add: (dc.format.digitalorigin)	born digital
Add: (dc.language.iso)	eng
Add: (dc.subject.lcsh)	Clinical psychology
Add: (dc.type.dcmi)	Text
Add: (dc.type.genre)	Thesis
Remove: (dc.contributor.advisor)	Massman, Paul J.
Remove: (dc.contributor.committeeMember)	Tian, Siva
Remove: (dc.contributor.committeeMember)	Dulay, Mario

Remove: (dc.date.created)	2015-05
Remove: (dc.date.issued)	2015-05-12
Remove: (dc.date.submitted)	May 2015
Remove: (dc.type)	Thesis
Remove: (dc.type.material)	text
Remove: (dc.type.material[*])	text

Changes pending for item 1132 (10657/1118)

Add: (dc.contributor.advisor)	Zvolensky, Michael J.
Add: (dc.contributor.committeeMember)	Knee, C. Raymond
Add: (dc.date.created)	December 2014
Add: (dc.date.issued)	2014-12
Add: (dc.description.department)	Psychology
Add: (dc.format.digitalorigin)	born digital
Add: (dc.language.iso)	eng
Add: (dc.subject)	Emotional Non-Acceptance
Add: (dc.subject)	Anxiety Sensitivity
Add: (dc.subject)	Post-traumatic Stress
Add: (dc.subject)	Smoking
Add: (dc.subject.lcsh)	Clinical psychology
Add: (dc.type.dcmi)	Text
Add: (dc.type.genre)	Thesis
Add: (thesis.degree.discipline)	Psychology

Where are we now?

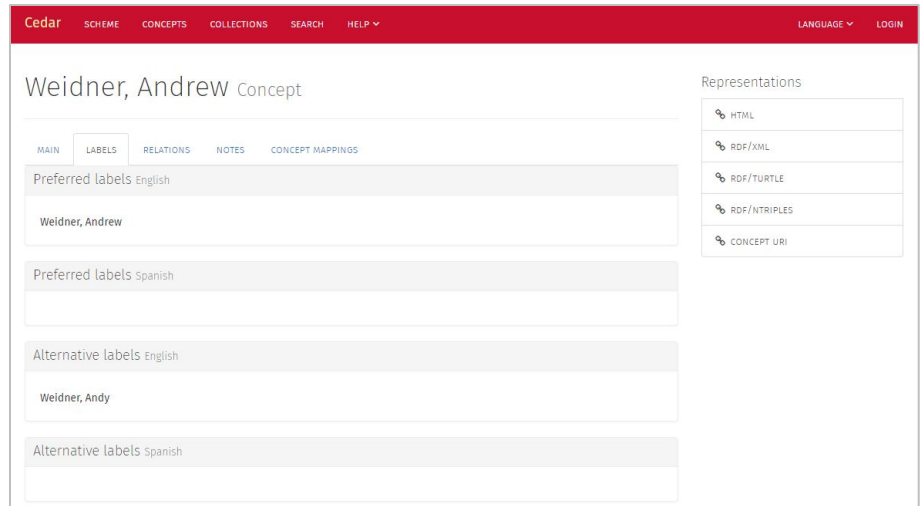
- ✓ 941 corrected records
- ✓ Workflow for new batches
- ✓ Corrections to Vireo



<http://i16.picdn.net/shutterstock/videos/6597086/thumb/10.jpg>

Next steps

- Local thesaurus
 - <https://vocab.lib.uh.edu/en.html>
- Vireo improvements
- Workflow efficiencies
- Data sharing



Acknowledgements

- Colleen Lyon at University of Texas-Austin
- Monica Rivero at Rice University
- Texas Digital Library staff
- DSpace Google Group
- OpenRefine Google Group

Thank you

- Xiping Liu xliu54@central.uh.edu
- Albert Duran aduran8@central.uh.edu
- Anne Washington awashington@uh.edu