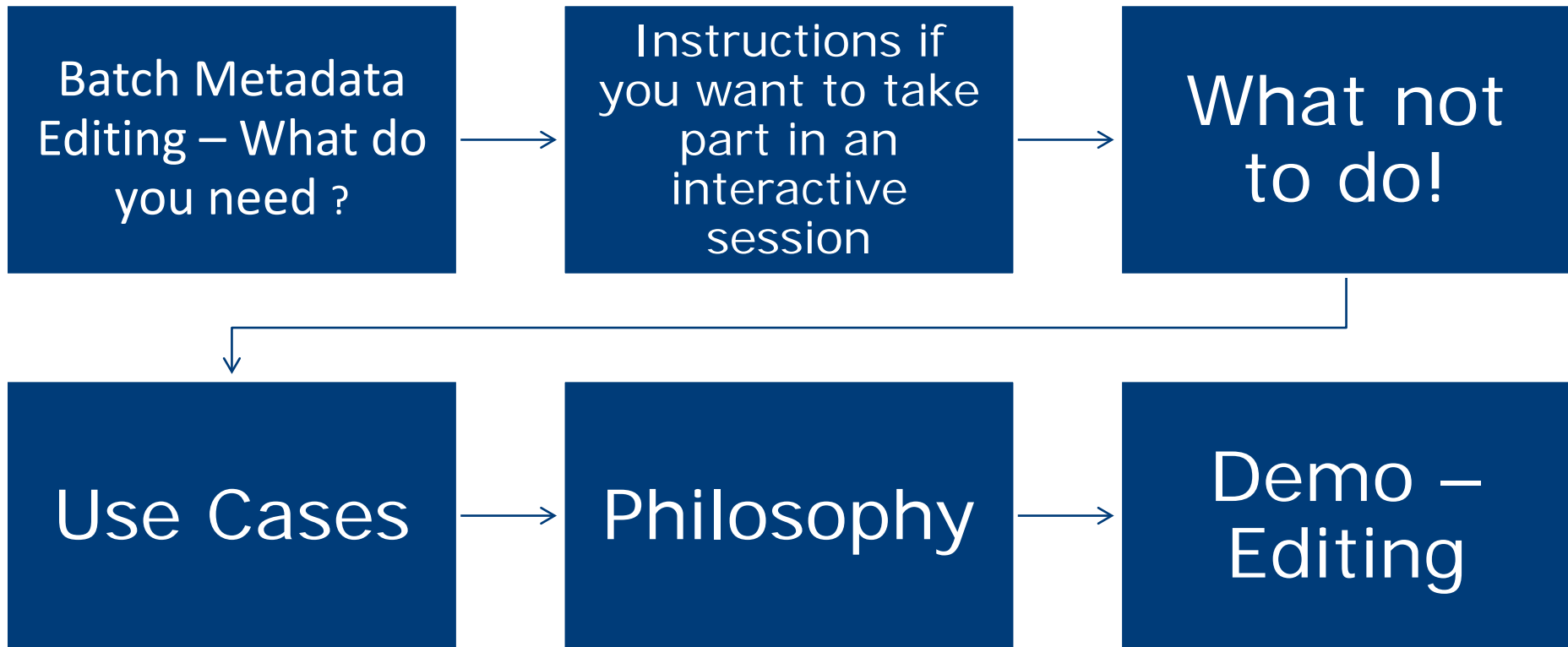


Batch Metadata Editing in Dspace 1.6

<http://researchspace.auckland.ac.nz>

Leonie Hayes, Stuart Lewis, & Vanessa Newton-Wade

**July 2010 – Workshop - Open Repositories 2010
Madrid**



Demo URL: <http://demo.dspace.org/jspui/>

Workshop participation

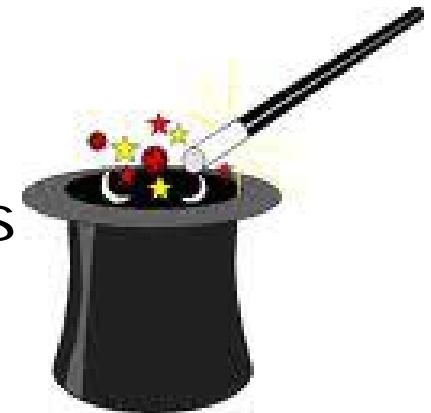
- For a hands on experience
- Login to demo.dspace.org
- Need internet, and spreadsheet software
- 20 demo accounts available
- Partner with a super user
- Feedback loop to us Tweet it

What not to do!

- Don't use this to make large scale changes
- Don't get distracted

Scope

- What can I batch edit ?
 - You can export a whole community/collection or individual records
- What is this useful for?
 - Batch modifying existing records
 - Adding new metadata records
 - Mapping virtual collections
 - Moving items from one collection to another



Scope

- You can only use this for metadata
- This will not replace the existing batch import facility for adding items ie files + metadata
- Command line or JSPUI XMLUI

Use Cases

- Quality control
- Authority control
- Add skeleton records
- Creating virtual collections
- Reporting

Philosophy

- Keep it simple
- Build it on top of the application
- Use the skill set we have – familiar software

Demonstration

Login using a Site
Administrator Account

Navigate your way to
the collection you will
be changing

For those participating
online go to the
assigned community
and select the first



Collection, then export
it



Admin Tools

- Edit...
- Item Mapper
- Edit Submitters
- Export Collection
- Export (migrate) Collection
- Export metadata
- [Admin Help...](#)

Click on the export metadata and
save file in CSV format

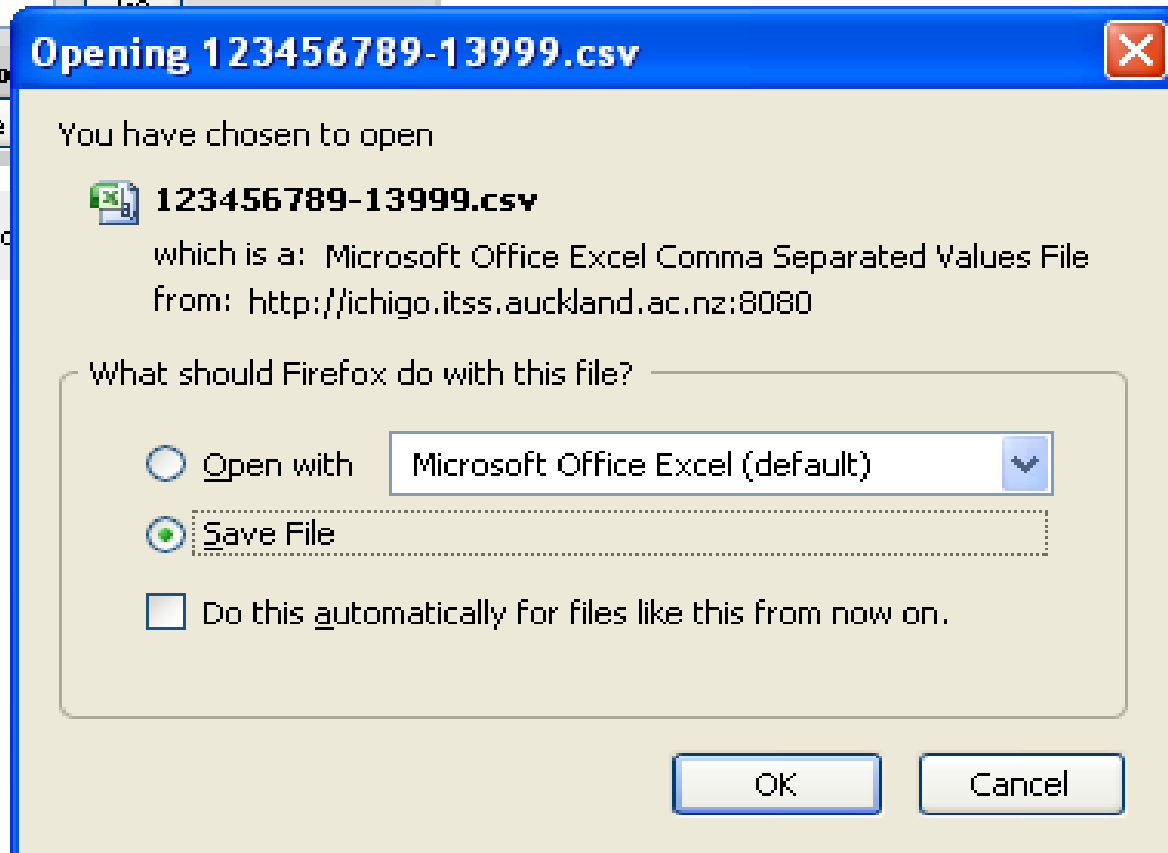
nal Articles 2009 ▼

Go

or bro

ject Title

aily e-mail no



1. With CSV files we have found it better to use Open Office to open the spreadsheet
2. Character set to Unicode (UTF8) important.
3. Select all the columns by clicking on the top left upper corner, change to text.
4. If you open the csv file with Excel you don't get the same options to open the file in Unicode and text column types. Results: dates and other fields like issn are opened as numbers and its very difficult to make them behave like text. Also long fields like abstract maybe truncated.
5. When Using Excel open the application then use the data tab to import from txt

Text Import - [template22july2009.csv]

Import
 Character set: **Unicode (UTF-8)** (circled in red)
 From row: 1

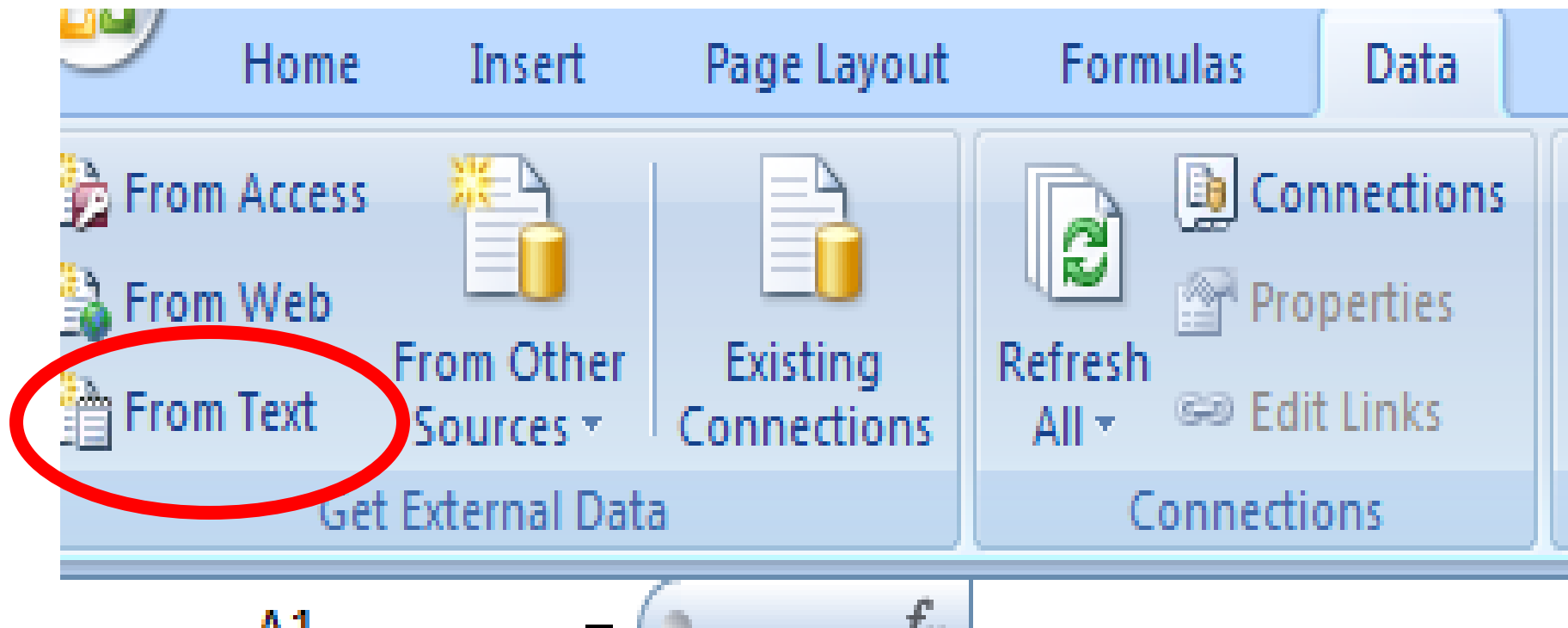
Separator options
 Fixed width
 Separated by
 Tab Comma Other
 Semicolon Space
 Merge delimiters Text delimiter: " "

Fields
 Column type: **Text** (circled in red)

	Text	Text	Text
1	id	collection	ac.contributor.author[en_US]
2	14642	123456789/12631	Baker, G.E. Koelmeyer, T.D.
3	+	123456789/13999	Zhang, C Easteal, A.J.
4	+	123456789/13999	Wu, J Deng, C Shao, X Xie, S.Q.
5	+	123456789/13999	Irzik, G Nola, R.
6	+	123456789/13999	Addis,, D.R Sacchetti,, D.C Ally,, B
7	+	123456789/13999	Drede, P Duvier, C Moxner, T Delwa

Buttons: OK, Cancel, Help

Opening in Excel



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O		
1	id	collection	dc.contributor.author	dc.date.accessioned	dc.date.available	dc.date.issued	dc.identifier.cit	dc.identifier.isi	dc.identifier.ot	dc.identifier.uri	dc.description	dc.description	dc.description	dc.description	dc.description	dc.title[en_US]	dc.ty
2	14474	123456789/13999	Gorman, D.F. Brooks, P	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Medical Journ	0025729X	eid=2-s2.0-5811177083X	http://hdl.handle.net/1177083X	Made available Previous issue	Gorman, D.F.,	http://sfx.ack	Medical Journ	On solutions to	Jour	
3	14499	123456789/13999	Plester, B.	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Kotuitui 4 (1),	1177083X	eid=2-s2.0-6511177083X	http://hdl.handle.net/1177083X	Made available Previous issue	Plester, B., De	http://sfx.ack	Kotuitui	Healthy humou	Jour	
4	14544	123456789/13999	Beaglehole, R. Bonita, F	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	New Zealand	11758716	eid=2-s2.0-59111758716	http://hdl.handle.net/11758716	Made available Previous issue	Beaglehole, R.	http://sfx.ack	New Zealand	Heart health in	Jour	
5	14382	123456789/13999	Nichol, S. Deng, Y. Horr	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Journal of the	03036758	eid=2-s2.0-67111758716	http://hdl.handle.net/03036758	Made available Previous issue	Nichol, S., Sch	http://sfx.ack	Journal of the	Preservation o	Jour	
6	14559	123456789/13999	Holmes, M.	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Electronic Cor	1083589X	eid=2-s2.0-61111758716	http://hdl.handle.net/1083589X	Made available Previous issue	Holmes, M., De	http://sfx.ack	Electronic Cor	The scaling lim	Jour	
7	14560	123456789/13999	Lawton, B.A. Rose, S.E	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	BMJ 338 (7686)	09598146	eid=2-s2.0-61111758716	http://hdl.handle.net/09598146	Made available Previous issue	Lawton, B.A.,	http://sfx.ack	BMJ	Exercise on pr	Jour	
8	14577	123456789/13999	Boutros, N. Davison, M.	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Journal of the	00225002	eid=2-s2.0-64111758716	http://hdl.handle.net/00225002	Made available Previous issue	Boutros, N., D	http://sfx.ack	Journal of the	Conditional reir	Jour	
9	14582	123456789/13999	Shooter, D. Brimblecom	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	International Jc	09574352	eid=2-s2.0-57111758716	http://hdl.handle.net/09574352	Made available Previous issue	Shooter, D., S	http://sfx.ack	International Jc	Air quality inde	Jour	
10	14592	123456789/13999	Alatas, V. Cameron, L.	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Southern Ecor	00384038	eid=2-s2.0-59111758716	http://hdl.handle.net/00384038	Made available Previous issue	Alatas, V., Wo	http://sfx.ack	Southern Ecor	Gender, cultur	Jour	
11	14594	123456789/13999	Morgan-Smith, R.K. Ellio	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Journal of For	0895173X	eid=2-s2.0-59111758716	http://hdl.handle.net/0895173X	Made available Previous issue	Morgan-Smith,	http://sfx.ack	Journal of For	Enhancement i	Jour	
12	14596	123456789/13999	Masuda, Y. Mashima, R	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Nippon Serami	18820743	eid=2-s2.0-58111758716	http://hdl.handle.net/18820743	Made available Previous issue	Masuda, Y., D	http://sfx.ack	Nippon Serami	Relationship be	Jour	
13	14057	123456789/13999	Young, P.G. Walanj, R.	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Journal of bac	10985530	eid=2-s2.0-67111758716	http://hdl.handle.net/10985530	Made available Previous issue	Young, P.G., S	http://sfx.ack	Journal of bac	The crystal str	Jour	
14	14071	123456789/13999	Park, J.	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Anthropologic	00035459	eid=2-s2.0-67111758716	http://hdl.handle.net/00035459	Made available Previous issue	Park, J., Unive	http://sfx.ack	Anthropologic	Concepts of hi	Jour	
15	14082	123456789/13999	Beatty, S. Stevens, W.	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	The New Zeal	11758716	eid=2-s2.0-67111758716	http://hdl.handle.net/11758716	Made available Previous issue	Beatty, S., Fac	http://sfx.ack	The New Zeal	Lung cancer p	Jour	
16	14084	123456789/13999	Pasley, T. Poole, P.	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	The New Zeal	11758716	eid=2-s2.0-66111758716	http://hdl.handle.net/11758716	Made available Previous issue	Pasley, T., Fac	http://sfx.ack	The New Zeal	Characteristic	Jour	
17	14085	123456789/13999	Dare, A.J. Petrie, K.J. B	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	The New Zeal	11758716	eid=2-s2.0-66111758716	http://hdl.handle.net/11758716	Made available Previous issue	Dare, A.J., Fac	http://sfx.ack	The New Zeal	Prepared for p	Jour	
18	14086	123456789/13999	Murphy, M. Antia, U. Ch	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	The New Zeal	11758716	eid=2-s2.0-66111758716	http://hdl.handle.net/11758716	Made available Previous issue	Murphy, M., Sc	http://sfx.ack	The New Zeal	Party pills and	Jour	
19	14091	123456789/13999	Andrade, A.D.	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Qualitative Re	10520147	eid=2-s2.0-66111758716	http://hdl.handle.net/10520147	Made available Previous issue	Andrade, A.D.	http://sfx.ack	Qualitative Re	Interpretive re	Jour	
20	14107	123456789/13999	Mittal, A. Phillips, A.R. Iv	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	JOP : Journal c	15908577	eid=2-s2.0-66111758716	http://hdl.handle.net/15908577	Made available Previous issue	Mittal, A., Dep	http://sfx.ack	JOP : Journal c	The proteome	Jour	
21	14109	123456789/13999	Mitchell, E.A.	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Developmental	00121630	eid=2-s2.0-66111758716	http://hdl.handle.net/00121630	Made available Previous issue	Mitchell, E.A., I	http://sfx.ack	Developmental	What is the me	Jour	
22	14118	123456789/13999	Worthley, D.L. Ruszkiev	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Stem Cells 27	10665099	eid=2-s2.0-67111758716	http://hdl.handle.net/10665099	Made available Previous issue	Worthley, D.L.	http://sfx.ack	Stem Cells	Human gastroi	Jour	
23	14632	123456789/13999	Aw, K.C. Tsakadze, Z.	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Scripta Mater	13596462	eid=2-s2.0-53111758716	http://hdl.handle.net/13596462	Made available Previous issue	Aw, K.C., Dep	http://sfx.ack	Scripta Mater	Influence of ra	Jour	
24	14633	123456789/13999	Campbell, D.L.M. Haube	2009-07-17T01:00:00Z	2009-07-17T01:00:00Z	2009	Behavioural Pr	03766357	eid=2-s2.0-57111758716	http://hdl.handle.net/03766357	Made available Previous issue	Campbell, D.L.	http://sfx.ack	Behavioural Pr	Spatial and bel	Jour	
25											Made availabl						

Tips

- The multiple metadata fields are delimited with a || double pipe character, but it is possible to choose your own delimiter.
- The default number of items and delimiter can be changed in the DSpace config
- The most important parts are the id and collection

What fields

- Always strip out fields (columns) you don't want to change ie date accessioned and provenance, handle id
- Just leave the fields you want to update
- To ensure that content IS deleted you need to overwrite with a blank cell and leave the column header in place

Editing the csv file

- You may find it easier to save the csv file into Excel format .xls and make your modifications.
- When complete save back into csv for import
- Spreadsheet gurus will be very productive
- *For our demo take out all fields except for id, collection and date issued*
- *Change the date from 2009 to 2010*
- *Save the file with .csv extension*

In the admin menu you will notice an additional item– Import Metadata



- [Communities/ Collections](#)
- [E-people](#)
- [Groups](#)
- [Items](#)
- [Metadata Registry](#)
- [Bitstream Format Registry](#)
- [Workflow](#)
- [Authorization](#)
- [Edit News](#)
- [Edit Default License](#)
- [Supervisors](#)
- [Statistics](#)
- [Import metadata](#)
- [Withdrawn Items](#)
- [Help](#)
- [Log Out](#)

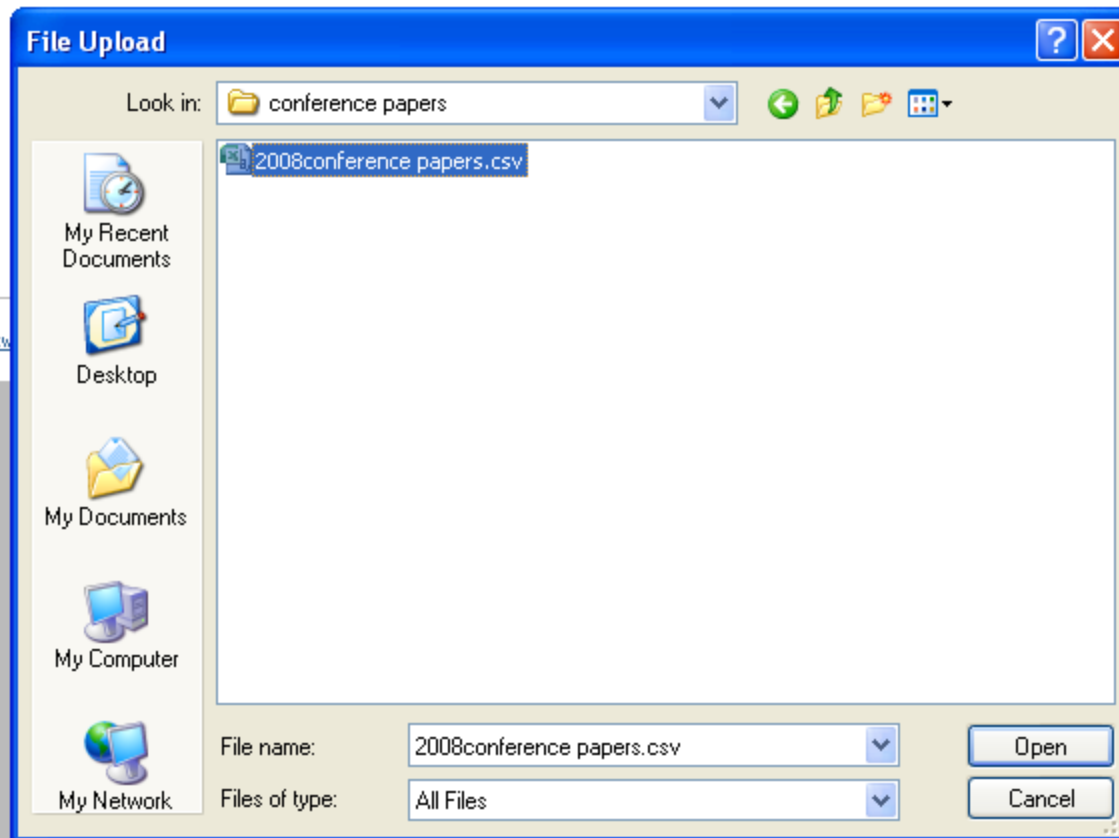
[ResearchSpace@Auckland \(1.5.2\)](#) >

Administration Tools

Please select an operation from the navigation bar on the left.



Browse to locate your files



Changes for items are shown below

[DSpace Demo Repository](#) >
[Administer](#) >

Import metadata

t	Changes for item: 305 (10673/291)	Add (dc.date.issued[])	2010
		Remove (dc.date.issued[])	2009
	Changes for item: 306 (10673/292)	Add (dc.date.issued[])	2010
		Remove (dc.date.issued[])	2009
	Changes for item: 307 (10673/293)	Add (dc.date.issued[])	2010
		Remove (dc.date.issued[])	2009
	Changes for item: 308 (10673/294)	Add (dc.date.issued[])	2010
		Remove (dc.date.issued[])	2009
	Changes for item: 309 (10673/295)	Add (dc.date.issued[])	2010
		Remove (dc.date.issued[])	2009
a	Changes for item: 310 (10673/296)	Add (dc.date.issued[])	2010
		Remove (dc.date.issued[])	2009
	Changes for item: 311 (10673/297)	Add (dc.date.issued[])	2010

Check everything before applying the changes, cancel if you see a mistake



Apply changes



Cancel

Wait till completed.

[DSpace Demo Repository](#) >
[Administer](#) >

Import metadata

20 changes completed successfully.

[nat](#)

Changes for item: 305 (10673/291)	Added (dc.date.issued[])	2010
	Removed (dc.date.issued[])	2009
Changes for item: 306 (10673/292)	Added (dc.date.issued[])	2010
	Removed (dc.date.issued[])	2009
Changes for item: 307 (10673/293)	Added (dc.date.issued[])	2010
	Removed (dc.date.issued[])	2009
Changes for item: 308 (10673/294)	Added (dc.date.issued[])	2010
	Removed (dc.date.issued[])	2009
Changes for item: 309 (10673/295)	Added (dc.date.issued[])	2010
	Removed (dc.date.issued[])	2009

**After demo if we have time to do
one more**

Adding items shown below

	A	B
1	id	collection
2	+	10673/229
3	+	10673/229
4	+	10673/229

Creating /Virtual collections using the mapping function

Example of a mapping to more than one collection

	A	B	C	D	E	F	G	H	
1	id	collection	dc.contrib	dc.date.ac	dc.date.av	dc.date.is	dc.identif	dc.identif	dc
2	14113	123456789/14655 123456789/13999	Kelderma	2009-07-1	2009-07-1	2009	Physica D:	1672789	ei
3	14167	123456789/14655 123456789/13999	Garny, A.	2009-07-1	2009-07-1	2009	Philosoph	1364503X	ei
4	14168	123456789/14655 123456789/13999	Christie, G	2009-07-1	2009-07-1	2009	Philosoph	1364503X	ei
5	14172	123456789/14655 123456789/13999	Beard, D.A	2009-07-1	2009-07-1	2009	Philosoph	1364503X	ei
6	14202	123456789/14655 123456789/13999	Wimalara	2009-07-1	2009-07-1	2009	Experime	9580670	ei
7	14237	123456789/14655 123456789/13999	Tawhai, M	2009-07-1	2009-07-1	2009	IEEE Engin	7395175	ei
8	14230	123456789/14655 123456789/13999	Ten, Tusso	2009-07-1	2009-07-1	2009	Experime	9580670	ei
9	14347	123456789/14655 123456789/13999	Cooling, M	2009-07-1	2009-07-1	2009	Biophysic	15420086	ei
10	14399	123456789/14655 123456789/13999	O'Brien, B	2009-07-1	2009-07-1	2009	Applied P	9478396	ei
11	14367	123456789/14655 123456789/13999	Tran, K.	2009-07-1	2009-07-1	2009	Biophysic	15420086	ei
12	14527	123456789/14655 123456789/13999	Kelderma	2009-07-1	2009-07-1	2009	American	3636135	ei
13	14573	123456789/14655 123456789/13999	Cairns, S.F	2009-07-1	2009-07-1	2009	Journal of	87507587	ei
14	14635	123456789/14655 123456789/13999	Ostby, I.	2009-07-1	2009-07-1	2009	PLoS Com	1553734X	ei

Moving items

	A	B
1	id	collection i
2	305	10673/229
3	306	10673/229
4	307	10673/229
5	308	10673/229

	A	B	C
1	id	dc.identifier.slug	dc.identifier.pmid
2	14474		19203316
3	14499		
4	14544		19182845
5	14382		
6	14559		
7	14560		
8	14577		19230511
9	14582		
10	14592		
11	14594		
12	14596		
13	14057		19429619
14	14071		
15	14082		19465945
16	14084		19448774
17	14085		19448772
18	14086		19448787
19	14091		
20	14107		19287105
21	14109		
22	14118		
23	14632		
24	14633		19010401
25	14106		
26	14113		

Removing entries : In this example the field that is empty will be replacing values with null values.

Summary - Advantages

- It give you as a RM the ability to make changes and be in control
- Yea for control freaks
- You can make continuous improvements to your metadata
- Quality control is not as onerous in fact I think it's a piece of cake
- You can manipulate and use the data in lots of different ways.