



The CURL Database and Copac

CURL Contributors' Seminar
Thursday, 8th December, 2005

Sarah Davnall
Copac Team
MIMAS

If you want to see more of the notes below, use your mouse to move the bottom of this slide portion higher.

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



The Software

- **Livelihood Discovery Server**
 - Used to be called BRS-Search

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



The Software

- **Why this software**
 - Text retrieval not RDBMS
 - Raw search power
 - Robust updating process

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL

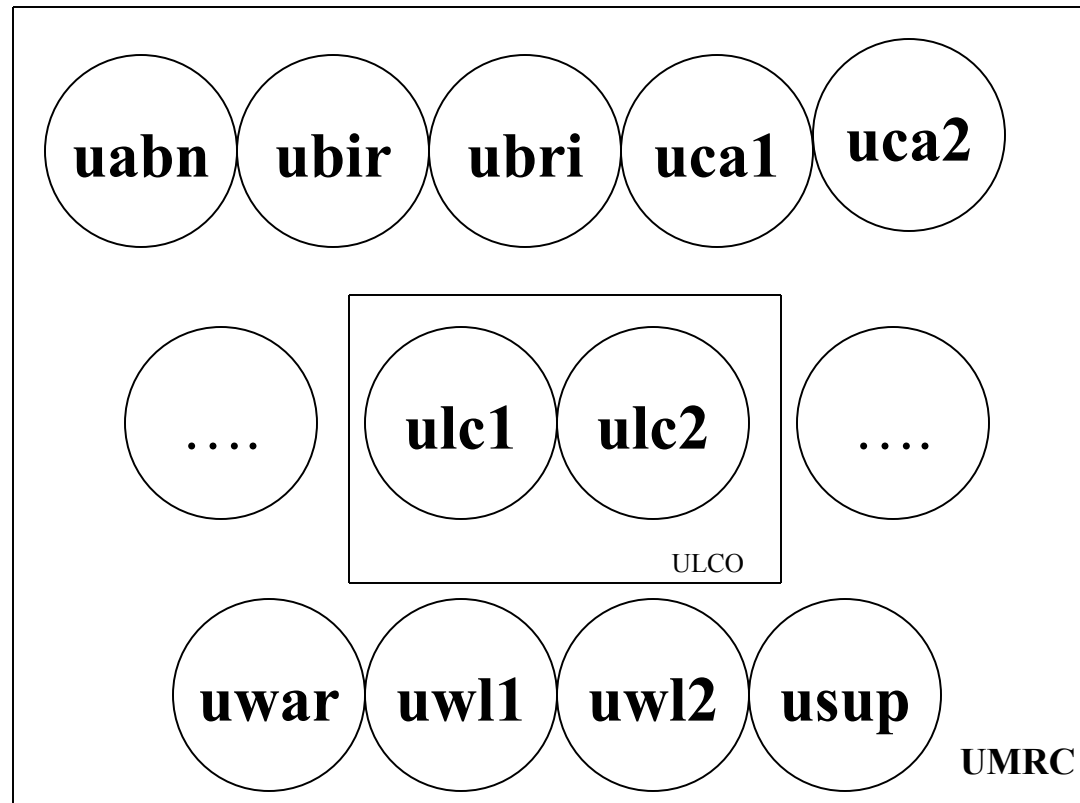


The CURL Database

- **Over-all Structure**
 - Database is in Pieces
 - One (or more) for each library
 - Named Uxxx
 - Concatenated to form the whole



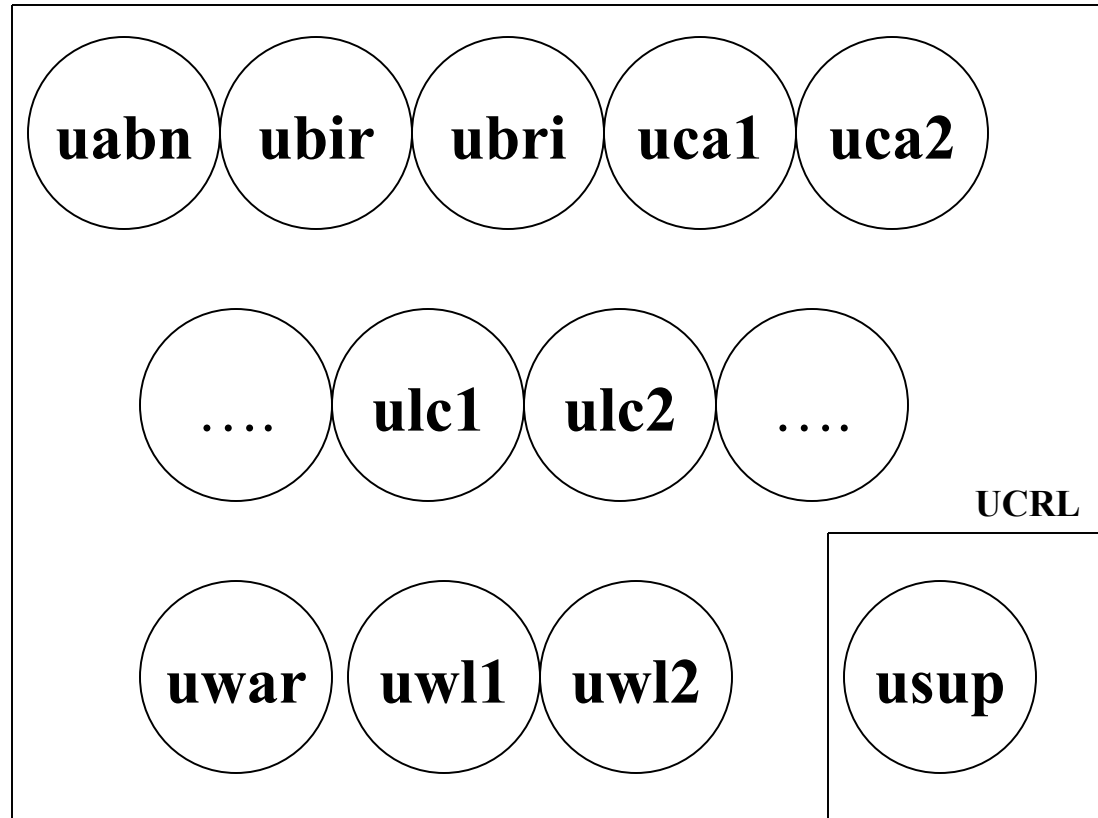
The CURL Database



Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



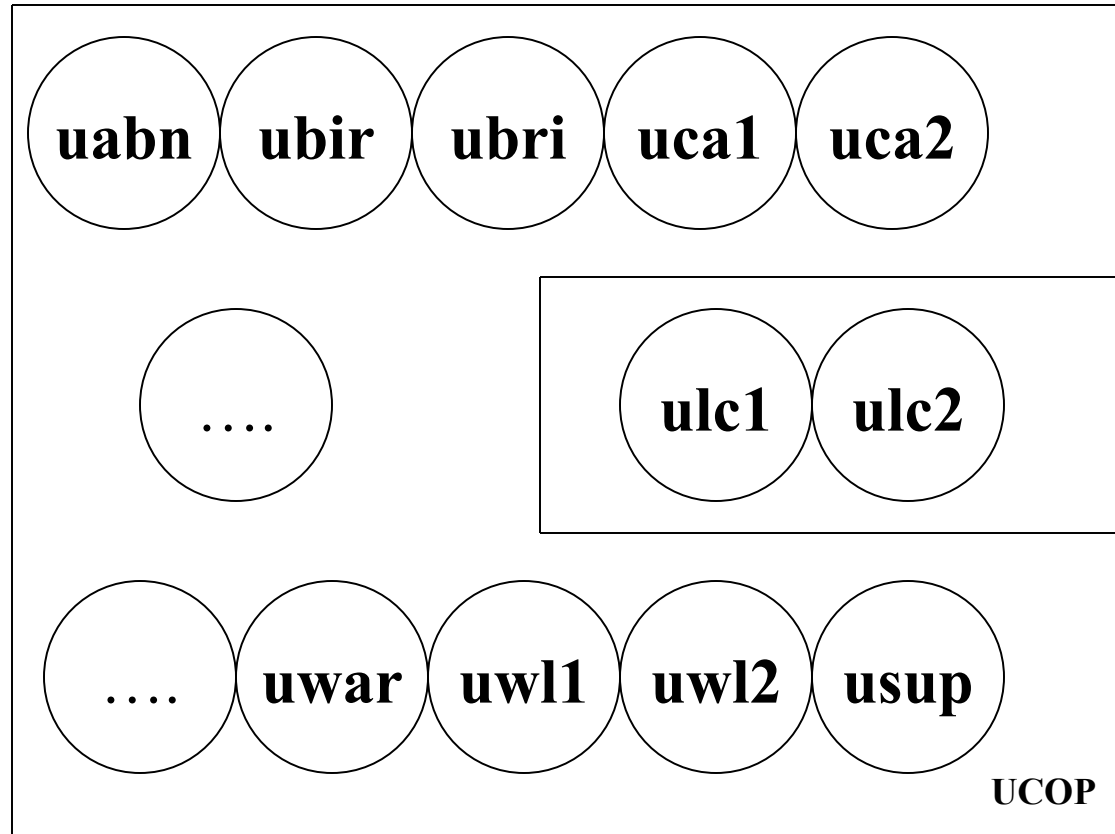
The CURL RR Database



Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



The CURL Copac Database



Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



The CURL Database

- **Piece structure**
 - **Dictionary**
 - Every searchable word in the piece
 - **Inverted index**
 - Every location of every word in the dictionary
 - **Text-index**
 - Pointer to each record in the text
 - **Text**
 - The actual records
 - **Form, Info**
 - Definition of the record structure

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



The CURL Database

■ A typical piece:

Entry	File	Size
DICT	/curlm21/wkly1/UCA1/dict.db	427356160
FORM	/curlm21/wkly1/UCA1/form.db	18341
INFO	/curlm21/wkly1/UCA1/info.db	4006
INV0	/curlm21/wkly1/UCA1/inv0.db	750009800
INV1	/curlm21/wkly1/UCA1/inv1.db	292879800
INV2	/curlm21/wkly1/UCA1/inv2.db	0
STAT	/curlm21/wkly1/UCA1/stat.db	638
TXT0	/curlm21/wkly1/UCA1/txt0.db	1900002193
TXT1	/curlm21/wkly1/UCA1/txt1.db	1802829151
TXIX	/curlm21/wkly1/UCA1/txix.db	27596336

CURL Cambridge MARC 21 Database

Size of Database UCA1 -- 5200696425 Characters

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



The CURL Database

- **Record structure**
 - **Fields**
 - MARC record – not searchable
 - Context-specific fields – searchable
 - All words (including codes) except stop-words
 - **Definition**
 - <http://www.curl.mimas.ac.uk/db-doc/Contents.html>
 - 50-page document

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL

The CURL Database



- **Display fields (1): the MARC record**

MREC 000 00 \$an\$b\$a\$cm\$d040611\$e2\$f1\$gf\$hb\$iu\$j0

MREC 001 03b14850527

MREC 003 UkLCURL

MREC 008 950317s1973 xxua 000 uaeng

MREC 010 \$a 73085910

MREC 020 \$a0874510910

MREC 035 \$a(StGIU)b14850527

MREC 038 \$aUkLCURL

MREC 049 \$jCU\$k03b14850527\$l\$m2

MREC 050 4 \$aZ720.S833 A3

MREC 090 \$aGla\$bBibliog A1:5 1973-S

MREC 100 1 \$aStillwell, Margaret Bingham,\$d1887-

MREC 245 10 \$aLibrarians are human :\$bmemories in and out of the rare-book world, 1907-1970.

MREC 260 \$aBoston :\$b[The Colonial Society of Massachusetts],\$c1973.

MREC 300 \$axiv, 401 p :\$billus ;\$c24 cm.

MREC 600 14 \$aStillwell, Margaret Bingham,\$d1887-

MREC 650 0 \$aLibrarians\$zUnited States\$xBiography.

MREC 650 0 \$aRare books\$xBibliography\$xMethodology.

MREC 650 0 \$aRare book libraries.



The CURL Database

- Database control fields:
- **CRN, library code, provenance, tags**

CRN	03b14850527
LIB	Gla
MSTD	2
PROV	1
TAGS	000 001 003 008 010 020 035 038 049 050 090 100 245 260 300 600 650
TCNT	17
SCNT	4
SFLG	y
TTOT	19
LDAT	20040413
RCL	DLC

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



The CURL Database

- Fields from Ldr, 006, 007, 008:
- **Record type, bib level, date, country, language**

RTYP	a
BIBL	m
ENCL	f
DCF	0
CTYP	0
MTYP	bk
DTYP	s
SDAT	1973
COP	mau
LANG	eng

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



The CURL Database

- **Control and Classification fields:**
- **ISBN, ISSN, LCN, BNB**
- **Dewey, LCCL, local classification no**
- **Opus no., Publisher no., statement of Scale**

CTRL	StGIU-b14850527
LCCN	73085910 73-85910
ISBN	0874510910
LCCL	Z720 S833
LOCL	Bibliog A1:5 1973-S

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



The CURL Database

- Bibliographic fields:
- author, title, series, subjects
- acronym searching, keyword searching

ATK	STIL-LIBR
TKEY	LIB-AR-HU-
QAU	Stillwell-MB
QSUB	Stillwell-MB
AU	Stillwell, Margaret Bingham, 1887-
TI	Librarians are human; memories in and out of the rare-book world, 1907-1970.
LCSH	Stillwell, Margaret Bingham, 1887-
LCSH	Librarians United States Biography.
LCSH	Rare books Bibliography Methodology.
LCSH	Rare book libraries.
NAME	Stillwell, Margaret Bingham, 1887-

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



The CURL Database

- Other fields
 - Place, publisher, pagination

POP	Boston
PUB	[The Colonial Society of Massachusetts]
PAGE	xiv 401 p

- Fields in alternate script
 - author, title, series, place, publisher

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



The CURL Database

- **Display fields (2):**
 - **local holdings, local fields**
 - **CURL brief display**

MHLD 859 01 \$!852 \$bgul11\$hBibliog A1:5 1973-S

MLOC 049 \$jCU\$k980073085910\$I\$m+

MLOC 059 \$aBibliog\$eA1:5

MLOC 907 \$a.b14850527\$b25-06-03\$c19-07-95

MLOC 998 \$agul\$b17-10-95\$cb\$d-\$e-\$feng\$gxxu\$h0\$i1

SDIS

Librarians are human :\$bmemories in an(0874510910) xxu 1973 Gla 1: : : 04



Data Loading

- Update procedure
- Initial full loads or reloads similar
 - But more data checking and discussion
 - New software written

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



Data Loading (1)

- Exchange-format update file
 - Weekly or monthly
 - Name is lib code plus sequence number
 - New records, updated records, deletions
 - Status identified through Leader cp5 or equivalent local field
 - Character set is marc-8 or UTF-8

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



Data Loading (2)

- SPLIT the exchange format into records
- CONVERT to an internal format
 - Generate CURL Record Number
 - Check for serious errors: reject
 - No 245
 - No holdings
 - Others agreed with the library
 - Check for other errors: warn
 - Invalid characters
 - Incorrect record type
 - Incorrect field and sub-field format
 - Others agreed with the library

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



Data Loading (3)

- **OUTPUT to LDS load format**
 - Format the MARC record for display
 - Separate the bib and local fields
 - Create the searchable fields
 - Separate out the deletion records
 - Separate out the suppressed records

- **VERIFY for LDS**
 - Check the record structure and size
 - Main and suppressed records

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



Data Loading (4)

- Check the reports
 - Make the reports available
- Additional data tasks
 - OCLC Worldcat file
 - Oxford LDLSCP file

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



Data Loading (5)

- Match CRNs against database
 - Database update is by delete and add
 - This identifies records already on dbase
 - CRNs are from deletion and update records
 - These will be deleted
 - File created of LDS deletion commands
 - Use the library's database piece for this
 - The only or latest piece

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



Data Loading (6)

- LDS deletion
 - Using the LDS deletion commands file
 - Deletions and updated records
 - And the library's only or latest database piece
- LDS load
 - Using the LDS load format file
 - New and updated records
 - And the library's only or latest database piece
 - Stores the record in the Text portion
 - Adds word addresses to the index chains
 - Adds new words to the dictionary

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



Data Loading (7)

- LDS reorganization
 - Using the library's only or latest dbase piece
 - Tidies up the inverted index chains
 - LDS utility
- LDS deletion
 - For the library's earlier database pieces
 - Using the CRNs file
 - Match against the database piece
 - Create LDS deletion commands file
 - Run the LDS deletion commands file
 - Deletes records which were in these pieces
 - Don't need to add: updates are in latest piece



Data Loading (8)

- **The suppressed records**
 - Match CRNs against database
 - Using all the CRNs
 - Because some records become unsuppressed
 - LDS deletion
 - LDS load
 - LDS reorganize

Copac is a MIMAS service
Funded by JISC and using records supplied by CURL



Copac vs CURL database

- Completely separate database
- Not an exact replica of the CURL database
- Non-MARC record format
- Pieces are not per library



Copac priorities

- Completeness and currency
 - a replica of the local catalogue
- No more results than necessary
 - de-duplication and consolidation
- Simplicity for the user
 - keep complications behind the scenes



The Copac record

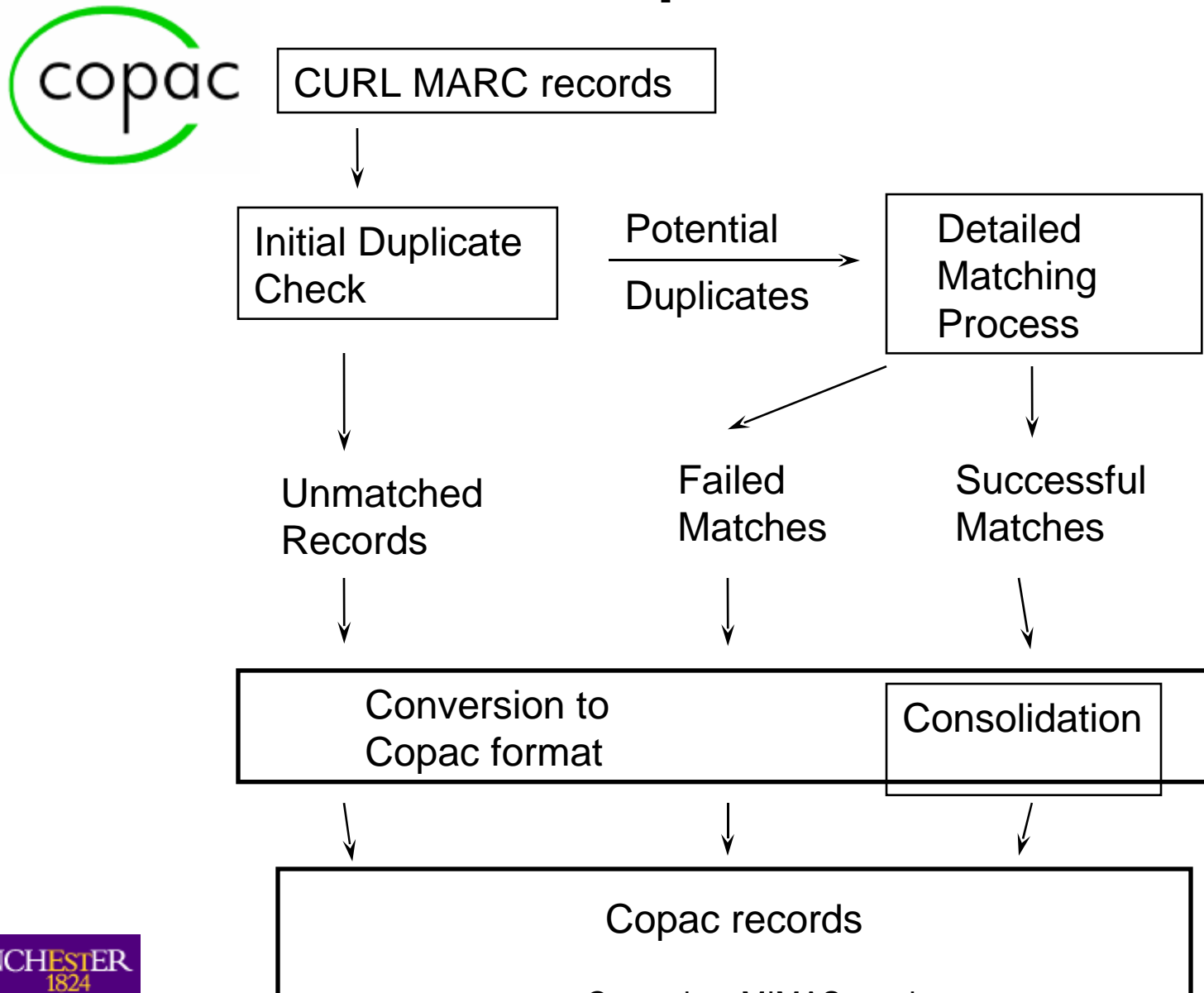
- Even longer than the CURL record
 - definition not a public document
- Similar fields to CURL:
 - author, title, publication, subjects
 - ISBN etc, classification codes
- Additions:
 - note fields
 - indexes for browse lists and sorting
 - CRN(s) and local control no(s)



Copac Updating

- A CURL update creates a Copac update
 - deletion CRNs
 - addition CRNs
- More complicated process:
 - consolidation of several CURL records
 - original record may be anywhere in Copac
- Still basically delete and add

Overview: Copac Consolidation



Copac is a MIMAS service



Copac updating: Delete phase

- Find Copac recs containing CRNs
 - save any other CRNs there
 - create Copac record deletion commands
 - LDS deletion run
- Find CURL records for saved CRNs
 - consolidate
 - build replacement Copac records
 - LDS verify and load runs



Copac updating: Add phase

- Find CURL recs matching CRNs
 - pass to Initial matching stage
- Find CURL recs for potential dups
 - detailed matching and consolidation stage
 - create delete commands for old Copac recs
- Add new Copac records
 - LDS verify and load runs
- Delete old Copac records



CURL Contributors' Seminar

Thursday, 8th December, 2005

End of Slides