

Catapult

Newsletter of the LIANZA Cataloguing Special Interest Group

Content

Editorial comment	1
CatSIG News	2
CatSIG looking for new editorial committee	2
Conference input requested	2
NZ cataloguing news	2
RDA (Resource Description and Access) in New Zealand	2
Featured Article	3
Automated WorldCat MARC Record Processing at AUT Library	3
Items in Brief	8
New assignFAST Prototype	8
LC New Bibliographic Framework Initiative	8
Bibliographic Framework as Web of Data	8
MarcEdit's RDA Helper	8
Introducing RDA Workshop by Chris Oliver	9
First issue versus latest issue	9
Linking Wikipedia articles to VIAF	9
Ex Libris promotes Linked Data Techniques	10
Metadata Object Description Schema Conversions Tools	10
Cultural Objects Name Authority (CONA)	10
New WebDewey Tool	10
New Books	11
Directions in Music Cataloging	11
Practical Strategies for Cataloging Departments	11
Special cataloguing	12
Joan Rivers and her card catalogue of jokes	12
What a name!	12

Catapult online

Archived issues available on DescribeNZ wiki: <http://www.describe.org.nz/>

CatSIG Committee

Convenor	Janess Stewart	National Library of NZ
Treasurer	Joanne Rowan	University of Auckland Library
Secretary	Sonya Maclaurin	University of Otago Library
Committee	Jessie Moir	Hastings District Libraries
	Karen Tobin	Central Hawkes Bay District Libraries
	Caroline Sherborne	Auckland Libraries

Contact: catsigcommittee@gmail.com

Editorial comment

Catapult is the quarterly newsletter of CatSIG. The responsibility for editing is shared around libraries and individual CatSIG Committee members.

This issue was prepared by the University of Auckland Library cataloguers: Joanne Rowan, Lai Lam, Linda Thomas, and Ksenija Mincic Obradovic.

I trust you have heard of the cataloguing utility implemented at AUT Library, used to pre-process catalogue records prior to their import into the catalogue. We are bringing you an article about the utility, the automated checking performed, and the workflows that have resulted.

In this Christmas issue, there is also a little treat – a few unusual titles and headings collected over years by UofA cataloguers.

Happy holidays, and all the best for 2013.

Ksenija Mincic Obradovic

Featured Article

Automated WorldCat MARC Record Processing at AUT Library

Sean Thompson

The Auckland University of Technology Library

Abstract

The article covers the rationale behind the use of a cataloguing utility, the automated checking performed, and the workflows that have resulted. Coverage is by necessity highlights rather than comprehensive – the utility is 30,000+ lines of code in length (008 cross checking alone takes more than 2000 lines). The main use for the utility is processing e-resource records - due to the number of records and links provided by vendors, and the extreme variability of quality of these records.

Introduction

AUT University Library has successfully implemented OCLC's WorldCat Cataloguing Partners (WCP) programme (formerly PromptCat), which provides automated delivery of MARC records for monographs invoiced by key vendors. The vast majority of MARC records required for these monographs are supplied through this programme, reflecting the breadth of the WorldCat database. However as the breadth of WorldCat's coverage is largely due to the number and variety of contributing institutions the records provided can be highly variable – both in quality and nature of description. The need to ensure that incoming records are of acceptable quality, and are consistent with similar materials in the catalogue led to the in-house development of a MARC processing utility to pre-process WorldCat records prior to their import into the catalogue.

Prior to the implementation of the WCP programme manual copy cataloguing workflows led to significant backlogs. Several banks of shelves in the workroom were required to temporarily house books awaiting copy cataloguing. Required copy cataloguing processes: importation of records, updating union catalogue holdings, authority control and updating local holdings could only commence once materials were in hand. The WCP programme has enabled these processes to be undertaken in advance. Most books are shelf ready from the point of check in. Shelving for routine copy cataloguing has been removed.

AUT University Library uses the Voyager ILS. Voyager has limited capacity to automate cataloguing processes necessitating offline pre-processing of MARC records. By pre-processing records in bulk, these limitations can be avoided. A consistent, comprehensive approach can be applied to incoming records – that cannot be achieved manually, due to time constraints and the probability of human error arising.

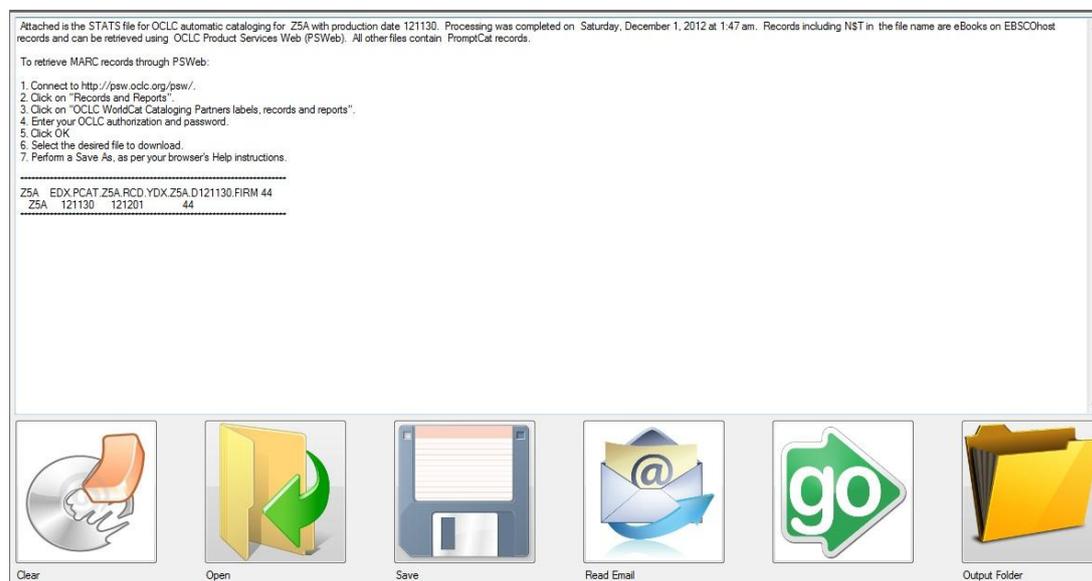
This processing was initially undertaken with the use of a series of scripts. These scripts have subsequently been migrated to a single GUI-based .NET Windows application.

The cataloguing utility performs four core tasks:

- The composite MARC file is downloaded from OCLC;
- Downloaded records are processed;
- Mismatched and subpar records are identified;
- Processed records are presented for evaluation and uploaded to the catalogue.

The cataloguing utility includes an indexing module, which generates the indexes used to assess incoming records. Bibliographic, holding and authority files are exported from the ILS, and are indexed. Authority control indexes are generated through automated parsing of the LCSH (XML) and MESH (MARC) thesauri. Additional index files to scan for spelling variants and specific MARC coding have been written by hand.

FIGURE 1 – WCP Notification Screen



Downloading MARC files

Email notifications are received from OCLC when composite MARC files are ready to be downloaded. These emails are machine read by the utility, which then downloads these files and loads them into memory for processing. Automated downloading saves time and ensures that the correct files are downloaded from the server.

MARC Record Processing

Each MARC record is processed to ensure that it is AACR2 compliant – when RDA is formally adopted by the AUT University Library processing algorithms will be amended to reflect the new cataloguing standards.

As each MARC record is processed unambiguous issues of unneeded fields, incorrect or missing punctuation and the addition of parallel title entries are addressed by discarding, modifying or adding fields to the record. Issues that require assessment by cataloguers: mismatched coding between fields, incomplete data, subject authority relationship, ISBN links to the existing records and classification are triaged by the addition of notes in 011 fields.

Discarding unneeded fields

Raw WorldCat records can contain a significant number of unneeded fields that clutter the field and reduce the readability of the record. Fields including: unneeded standard numbers (015, 016, 019, 029 fields), unsupported classification schema (060, 070) and WorldCat holding details (938) fields are stripped.

The AUT University Library catalogue supports LCSH, MESH, LCGFT and Maori Subject Headings thesauri. All other subject headings are removed.

ISBD punctuation

The utility ensures that each field is correctly punctuated. Terminal marks of punctuation are scanned for, and are appended if missing. Punctuation preceding subfields is corrected to ensure ISBD compliance. Where subfield coding is absent, but the field includes punctuation that unambiguously indicates that a subfield division should exist, subfield codes are added.

Consistency of coding between fixed and variable length fields

The utility scans the field to ensure that fixed length coding in 006, 007 and 008 fields is consistent with metadata in variable length fields.

The utility scans the record for fields (e.g. 300 subfield 3, 538, or 856) that can justify the existence of an 006 or 007 field. If found, the text in these fields is assessed for compliance with the precise coding in the 006 or 007. If no supporting fields are found the 006/007 is discarded, and a warning note is added as an 011 field.

The codes in the 008 are compared with the textual values in corresponding variable length fields to ensure consistency. Inconsistencies are appended as warning notes in 011 fields. Comparisons performed include:

- Date values are compared with 260 and 500 fields.
- Illustration codes are compared with the 300a and 300b subfields.
- Target audience codes are compared with contents of 521 fields, and subject juvenile v subfields.
- Nature of contents codes are compared with the contents of 245, 500, 520, and subject v subfields.
- Congress codes are evaluated by checking for meeting authority headings or the v subfield Congresses.
- Bibliographical reference and index codes are compared with the contents of 500 and 504 fields.
- Biographical codes are assessed by comparing the contents of 100, 600 and 700 fields.
- When non-English language coding is present the utility searches for elaboration in 041 or 546 fields.

Language coding

Whenever an 041 or 546 field is included in the record, the utility checks that the corresponding field is also present. Language coding of both 041 and 546 fields are assessed. If the list of language names applied to the 546 field differs from the list of codes in the 041, a warning note will be added as an 011 field.

Parallel Titles

To assist with title searching machine generated parallel titles may be added as 246 fields. These machine generated fields are added without a second indicator (the AUT OPAC doesn't display 246 fields without a second indicator, but still indexes them). 246 fields are added for:

- British spelling equivalents of American English titles;
- Text equivalents of numerals and symbols;
- Numerical and symbolic equivalents of textual values;
- 'Core' titles obtained by stripping author, publisher and series names, or prefixed phrases from the head of the title.

Authority control

The primary subject thesaurus used by AUT University Library is LCSH. MESH headings are utilised provided that these MESH headings are not identical to LCSH headings, nor match 450 references in LCSH authorities. Cross indexing of MESH and LCSH thesauri enables “non-compliant” MESH headings to be machine converted to LCSH headings.

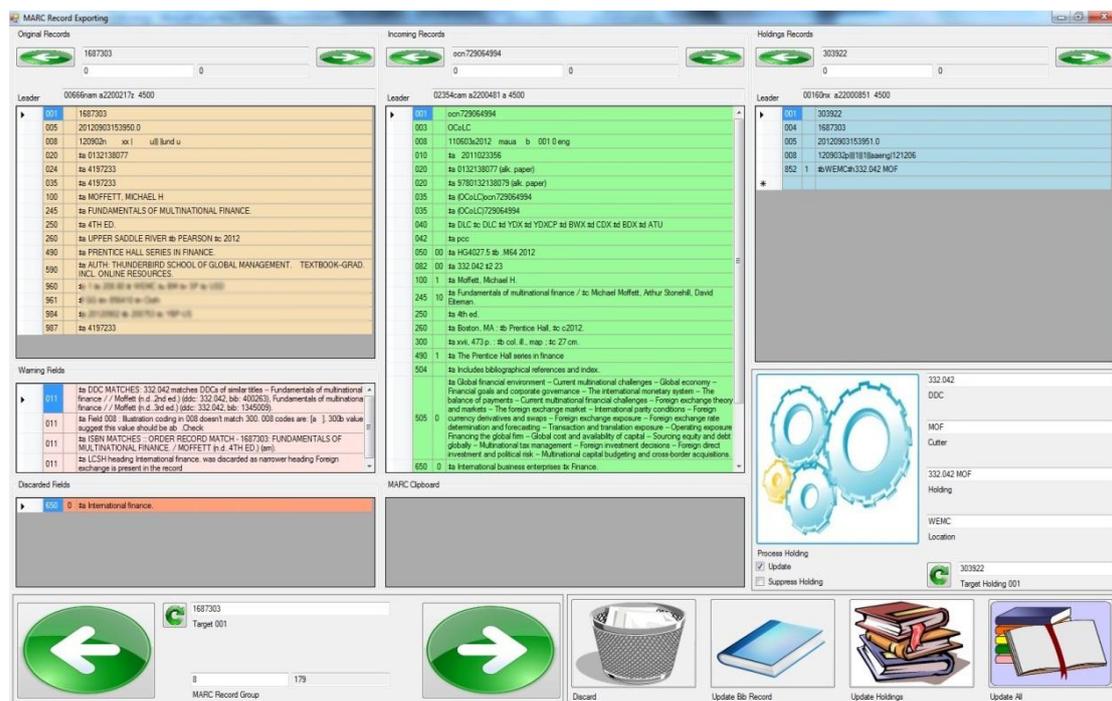
LCSH and MESH headings are assessed to check whether a narrower heading for each heading also exists in the record. If found the broader heading is discarded, with an 011 note added to enable the field to be restored if appropriate.

The utility checks whether headings in each record are authorised in the local authority file. Non matching headings are identified for manual or automated authority extraction via Z3950 or client to client connection.

ISBN

The primary match point for merging/overlaying bibliographic records at AUT is the ISBN. Multiple ISBN matches and no-matches are identified to provide for manual import/overlay of records without unambiguous destination records.

FIGURE 2 - Record editing/loader screen used by associated MARC loading utility



Classification

AUT is a Dewey library. The utility assesses the validity of the assigned DDC by checking whether the coded Dewey edition is current and whether the DDC used is consistent with similarly catalogued materials in the catalogue. The 100a and 245a subfields are concatenated. The DDC in the 082a subfield is then compared against all other DDC values that have been applied to this concatenated string. 011 fields are added detailing records that are consistent or otherwise with the DDC in the incoming record.

Identification of Subpar records

A proportion of supplied WorldCat records are unfit for import into the catalogue. ISBNs are extracted from these records to enable secondary searching of Te Puna or WorldCat for satisfactory equivalents. If no such records exist these books are put aside for original cataloguing.

The quality checking algorithm assesses a number of factors including:

- The encoding level and cataloguing form codes in the leader;
- Presence of specific library codes in the 040 field;
- Whether the capitalisation of the 245 field is in sentence case (rather than upper or title case);
- The completeness of 260 and 300 fields;
- The number of LCSH or MESH headings assigned to the record.

Numerical values are assigned to each of these factors. If the sum of these values exceeds a specified threshold, the record is provisionally accepted for importation into the catalogue.

Incomplete data

If the record exceeds the quality threshold, but data in 260, 300 or subject heading fields is incomplete or missing warning notes will be added as 011 fields.

Evaluation and upload

011 fields

Warning notes are appended to each record through the addition of 011 fields. The 011 field is the lowest numbered un-assigned variable length MARC field, and will display at the top of a MARC record sorted in field order. 011 fields are removed as issues identified are resolved.

Record import

Records are imported individually rather than in bulk, to enable 011 warnings that can be resolved by cataloguers prior to the physical receipt of the item to be resolved. Warnings that cannot be resolved at this stage are retained as 011 fields; any additional instructions for copy cataloguers can also be added as new 011 fields. Holdings records are updated at this point – incomplete holdings records are a red flag for copy cataloguers to pass books to cataloguers for reassessment. For the majority of records processed no further cataloguing work is required from the point that the book's barcode is entered at the point of check in.

Final remarks

At present, the cataloguing utility is hardcoded to work at AUT only. It is currently being rewritten for future proofing to shift hardcoded elements to editable XML. This will allow for updating of elements to occur without needing to recode the program itself, and will enable it to be used with other library systems. This should be completed over summer 2013.

Sean Thompson works as an Assistant Librarian (Cataloguing) at the Auckland University of Technology Library. His focus is cataloguing automation, particularly cataloguing of electronic resources. Sean has been solely responsible for establishing workflows and designing software to implement cataloguing utility. He can be contacted for further information re the utility.

Items in Brief

New assignFAST Prototype

OCLC Research is pleased to announce the availability of [assignFAST](#), a new Web service that automates the manual selection of FAST Subjects (the Authorized and Use For headings) based on autosuggest technology. It takes advantage of the features of FAST (Faceted Application of Subject Terminology). Faceting, along with a fully controlled vocabulary, allows simple selection and authority control to take place that would be hard to accomplish in LCSH.

Read the [news announcement](#) for more information, see the [assignFAST activity page](#) for complete details, then try the assignFAST [demo](#) or use the assignFAST [Web service](#).

Source: OCLCRLP Announcements to Partners, 26 Oct. 2012

LC New Bibliographic Framework Initiative

The new bibliographic framework transition initiative will be focused on the Web environment, Linked Data principles and mechanisms, and the Resource Description Framework (RDF) as a basic data model. Zepheira, led by Eric Miller, has been active in the development of Semantic Web and library standards as well as open source tools to support linked data technologies and library applications. Zepheira will provide both a model (or models) to serve as a starting point for discussion and an analysis of related initiatives. Miller spoke to Library of Congress staff with an overview of the project and a report of progress to date.

http://www.loc.gov/today/cyberlc/feature_wdesc.php?rec=5605

Source: AUTOCAT, 10 Oct. 2012

Bibliographic Framework as Web of Data

The Library of Congress is pleased to report that we have reached two important milestones with respect to our Bibliographic Framework Initiative: the introduction of a draft data model for web-based bibliographic description and a first meeting of a small group of early experimenters currently exploring the feasibility of the proposed model. The new model is simply called BIBFRAME, short for Bibliographic Framework.

<http://loc.gov/marc/transition/news/bibframe-112312.html>

Source: ADA-L 24 Nov. 2012

MarcEdit's RDA Helper

A new feature of MarcEdit is the RDA Helper, a development tool that provides the ability to RDAize AACR2 records. The function allows users to select from a broad range of RDA field options. RDA Helper includes automatic GMD generation.

<http://people.oregonstate.edu/~reaset/blog/archives/1080>

Source: AUTOCAT, 30 Oct. 2012

Introducing RDA Workshop by Chris Oliver

A 90-Minute Workshop on January 9, 2013 at 2:30pm Eastern.

RDA: Resource Description and Access is the new global cataloging standard. The March 2013 full implementation by Library of Congress and other national libraries is coming up fast, and libraries are working intensively to familiarize staff, whether they're going to be simply interacting with records created using RDA or doing original cataloging themselves. In this workshop, Chris Oliver, author of the best-selling *Introducing RDA*, will give an overview of RDA and look at the key aspects that make RDA different from AACR2. The workshop is designed to provide background for cataloguers before they begin RDA training and will also serve as a refresher for those wanting to brush up on their understanding of RDA.



Topics include:

- Organization of RDA and its key attributes
- Similarities and differences between RDA and AACR2
- Basic vocabulary and reasons for the changes
- Improving resource description and resource discovery

First issue versus latest issue

Recent discussions on RDA-L have high-lighted that in RDA the description of resources issued in more than one part is based on the earliest issued or lowest numbered part.(2.1.2.3)

JSC acknowledges that this is not a FRBR requirement and creates a barrier to the adoption of RDA. JSC notes that different approaches have different strengths and weaknesses: the earliest entry approach is considered to be more economical; but the latest entry approach emphasises convenience of the user.

JSC takes the view that data created under RDA should be sufficiently flexible to support any approach, without compromising the capability to control and link descriptions of serial resources.

JSC invites Deutsche Nationalbibliothek (DNB) to prepare a discussion paper for development of RDA to support a more flexible approach, while sustaining the process of harmonization with ISSN Network and ISBD.

<http://www.rda-jsc.org/firstissue.html>

Source: RDA-L, 15 Nov. 2012

Linking Wikipedia articles to VIAF

If you are interested in authority control and identifiers, or linking information between systems, including those outside the library domain, please check the summary of the results of the VIAFbot on "[VIAFbot Debriefing](#)".

Max Klein, OCLC's Wikipedian-in-Residence, finished linking Wikipedia biography articles to VIAF.org.

"We can now know how much linked authority data is on Wikipedia, its composition, and the similarities between languages."

Source: Hangingtogether blog: <http://hangingtogether.org/?p=2306>, 28 Nov.2012

Ex Libris promotes Linked Data Techniques

Ex Libris is pleased to announce that it has become a member of the W3C Schema Bib Extend Community Group. This community group brings together diverse library organizations and individuals who are creating proposals for extending schemas to improve the representation of bibliographic information on the Internet. The group operates under the umbrella of the schema.org initiative.

As a member of the group, Ex Libris will work toward making the data managed by Ex Libris solutions more accessible to others in the scholarly community and, similarly, making data from other systems available to libraries that use Ex Libris solutions. Membership in the community group is part of the Company's ongoing initiative to take advantage of linked data for the benefit of its customers and the wider library community. Linked data is a method of Internet publishing that enables data to be linked to automatically, rendering the data more accessible and hence more valuable.

Source: News from Ex Libris. 5 Dec. 2012

Metadata Object Description Schema Conversions Tools

A new set of MODS 3.4 XSLT 2.0 stylesheets has been added to our MODS Conversions page <http://www.loc.gov/standards/mods/mods-conversions.html> for community testing and review. These new XSLT 2.0 stylesheets are based on the mappings made available by the Library of Congress on the same page.

The conversions include:

- DC to MODS 3.4
- MARCXML to MODS 3.4
- MODS 3.4 to DC
- MODS 3.4 to HTML
- MODS 3.4 to MARCXML

Source: Library of Congress, 27 Nov. 2012

Cultural Objects Name Authority (CONA)

The Getty Institute is compiling the Cultural Objects Name Authority (CONA), a structured vocabulary that can be used to improve access to information about art, architecture, and material culture.

<http://www.getty.edu/research/tools/vocabularies/cona/about.html>

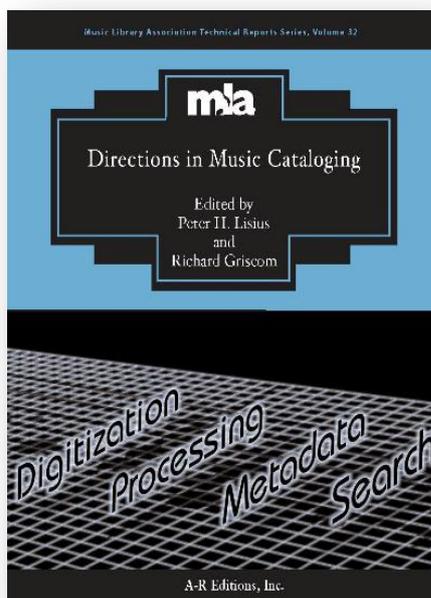
Source: Catalogablog, 16 Nov. 2012

New WebDewey Tool

A nice-looking beta number building, user contribution tool in WebDewey was recently made available. "You'll notice a new box, Create built number, associated with the record for any valid schedule or table entry."

Source: Catalogablog, 12 Nov. 2012

New Books



Directions in Music Cataloging

*Edited by Peter H. Lisius and Richard Griscom
173 p., A-R Editions Inc., 2012*

In *Directions in Music Cataloging*, ten of the field's top theoreticians and practitioners address the issues that are affecting the discovery and use of music in libraries today. Anyone who uses music in a library—be it a teacher, researcher, student, or casual amateur—relies on the work of music catalogers, and because these catalogers work with printed and recorded materials in a wide variety of formats, they have driven many innovations in providing access to library materials. As technology continues to transform the discovery and use of music, they are exploring ways to describe and provide access to music resources in a digital age. It is a time of flux in the field of music cataloging, and never has so much change come so quickly.

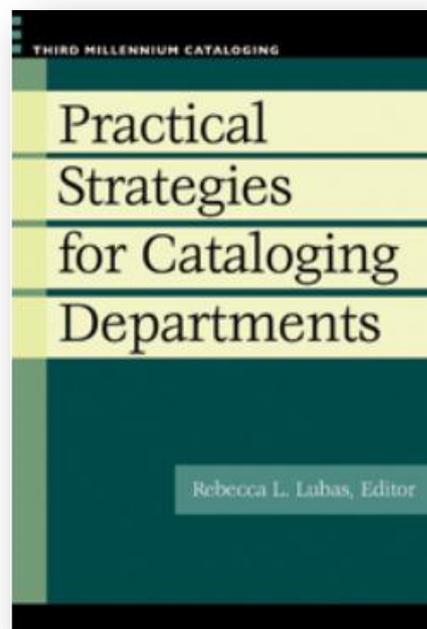
Source: <https://www.areditions.com/books/TR32.htm>

Practical Strategies for Cataloging Departments

*By Rebecca L. Lubas
117 p., Libraries Unlimited, 2011*

Demand for instant access, 24/7. Outsourcing issues. Constantly evolving standards. How can catalogers create a flexible, Web 2.0-compliant, flexible, multi-metalingual cataloging department? A daunting challenge, definitely; but with the right guidance, any cataloger can get up to speed and handle these common and confusing problems.

Each chapter of *Practical Strategies for Cataloging Departments* is authored or coauthored by a leader in cataloging, metadata practice, or education in these specialties. This book offers practical advice—based on direct experience—for facing the challenges of organizing information today. Topics include training, collaborating across the library, coping with changes in standards, making strategic selections of vendor cataloging products,



developing cooperative organizations, and more. The specific techniques that will help catalogers meet the needs of 21st century patrons are emphasized.

Source: <http://www.abc-clio.com/product.aspx?id=2147491932>

Special cataloguing

Joan Rivers and her card catalogue of jokes

<http://www.youtube.com/watch?v=87yztkvEslk>

For the past thirty-some years, Rivers has been filing each and every joke she's written (at this point she's amassed over a million) in a library-esque card cabinet housed in her Upper East Side apartment. The jokes—most typed up on three-by-five cards—are meticulously arranged by subject, which Rivers admits is the hardest part of organizing: "Does this one go under ugly or does it go under dumb?"



Source: Youtube – Clip from Joan Rivers: A Piece of Work

What a name!

Margo Knightbridge and Roslyn Smith from Auckland University Library Cataloguing Department collect unusual and amusing personal names, interesting book titles and subject headings. Here is a further selection from their files:

Personal Names

- Ariane Grizzelle
- Brenda Wineapple
- Brian Spickelmire
- Bryan Carrott
- Burkhard Springstubbe
- Debasis Bandyopadhyay
- DeMond Shondell Miller
- Dusadee Swangviboonpong
- Egamberdieva Dilfuza
- Erik Bongcam-Rudloff
- Esmeralda Rupp-Spangle
- Gerard Scronx
- Harald K.M. Bugmann
- Jan Manyfeathers
- Jeff Corntassel
- Kiribathgodage Indrapriya Akithalal Kularatne
- Levity Beet
- Marie-Laure Ngo Mpeck-Nyemeck
- Marvelene C. Moore
- Milky Wimpshake (punk rock band)
- Mouldy Sioud
- Nancy Pynch-Worthylake
- Patricia A. Holoday Worret
- Pixey Anne Mosley
- Quinny Sacks
- Rita Pinto Eliseu Mendes Godinho Veloso
- Robyn Barnacle
- Rong Gong
- Roy Doi
- Sheramy D. Bundrick
- Timberlake Wertenbaker
- Tuna Tasan-Kok
- Twylla Martinique Garrett-Bryant



Book Titles

- How to study law and take law exams in a nutshell
 - Time series modelling using support vector regression with chunking
 - Atlas of bacterial flagellation
 - The atlas of Messier objects
 - The adoration of the Magi in whalebone
 - The growing lawlessness of trees
 - The scope and aims of the Imperial Bureau of Agricultural Parasitology
 - Happiness and fulfilment in dentistry
-



Author/Title Combination

- Effective legal writing / Gertrude Block
 - Electrical appliances / Penny Sparke
 - The case against punishment / Deidre Golash
 - A dictionary of weather / Storm Dunlop
 - Using money laundering control to fight corruption / Heba Shams
 - Contemporary art and the home / edited by Colin Painter
 - Book by Ms Rossignol [French for nightingale] / translated by Ms Parrott
-



Corporate Bodies

- British Columbia. Ministry of Forests. Research Branch
 - Midwest Friends of the Pleistocene
 - Friends of the Algae
 - Baldy Centre for Law and Social Policy
 - Dropsie University
 - British Forging Industry Association
 - International Conference on the Use of Fly Ash, Silica Fume, Slag, and Natural Pozzolans in Concrete
-



Subject Headings

- Messes in art
- Frequent flyer programs in art
- Hazardous waste storage in art
- Dog scootering
- Horse ballet
- Confused flour beetle
- Congresses and conventions – Congresses
- Leg – Personal narratives
- Multi-user dungeons – Social aspects
- Plank cookery
- Spouted bed processes
- Volcanoes – Vocational guidance