



# 後設資料(metadata) 在圖書資訊學的趨勢研究

陳亞寧 陳淑君  
中央研究院計算中心

第三屆兩岸三院信息技術與應用交流研討會  
21 June 2005

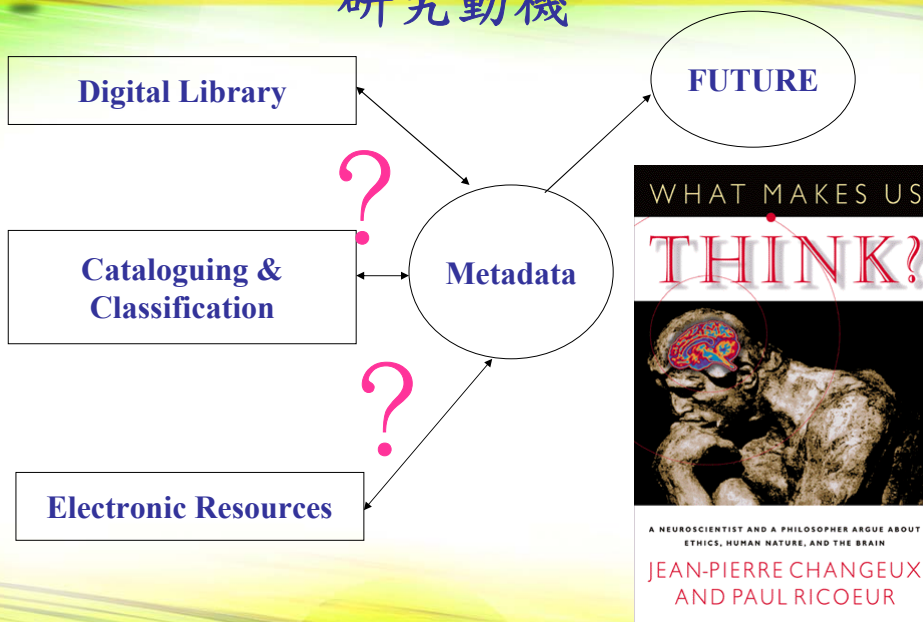


## 大綱

- 研究動機
- 研究設計與限制
- 研究結果與討論
- 研究發現
- 結語與建議



## 研究動機



## 研究設計與限制

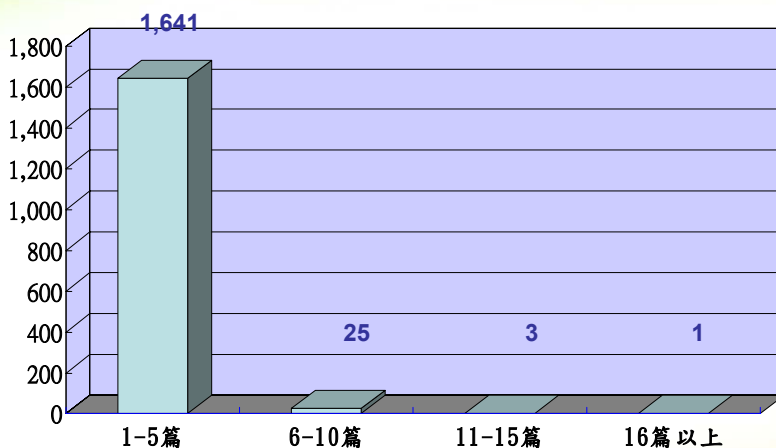
- 方法 — 內容分析法(content analysis)
- LISA資料庫
  - 94年3月21日
  - 1,491篇文獻
- 分析要點
  - 文獻記錄的量化分析
    - 每筆文獻記錄的描述語
    - 每筆文獻的發表作者
    - 每筆文獻所發表的期刊
    - 每筆文獻記錄的描述語與作者數量
    - 每年度所發表文獻的語文
  - 2004年文獻記錄的質性內容分析

## 研究結果與討論

- 發表文獻的作者分佈
- 發表文獻的期刊分佈
- 發表文獻的年代分佈
- 發表文獻的語文分佈
- 涵蓋研究議題/描述語的分佈



## 發表文獻的作者分佈

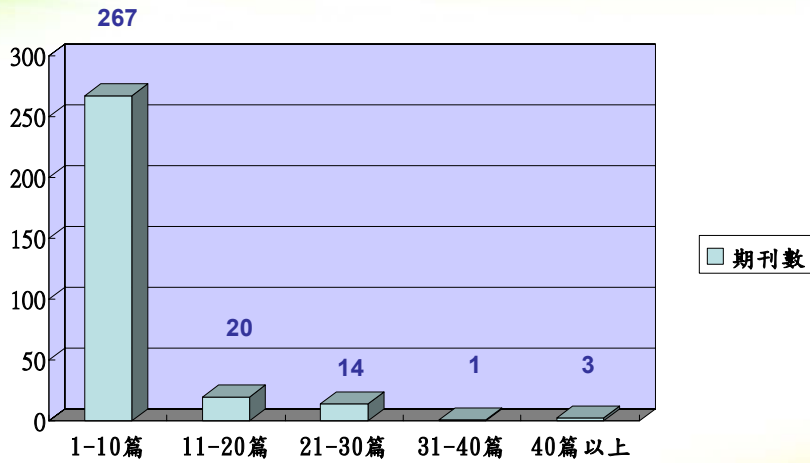


Total: 1,670位作者





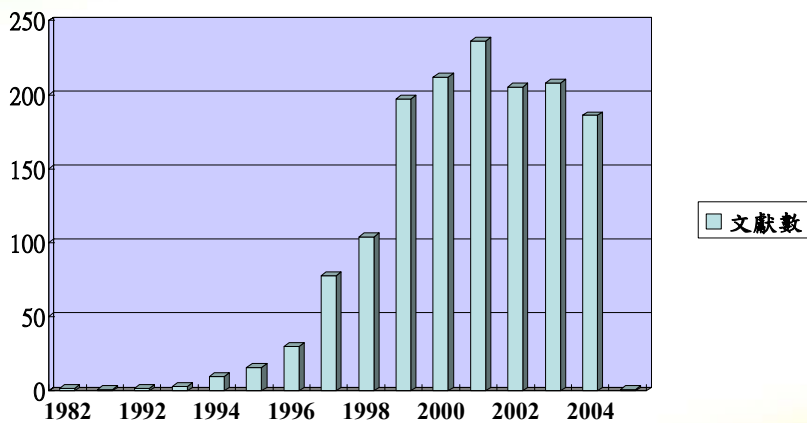
## 發表文獻的期刊分佈



Total: 299種資料來源

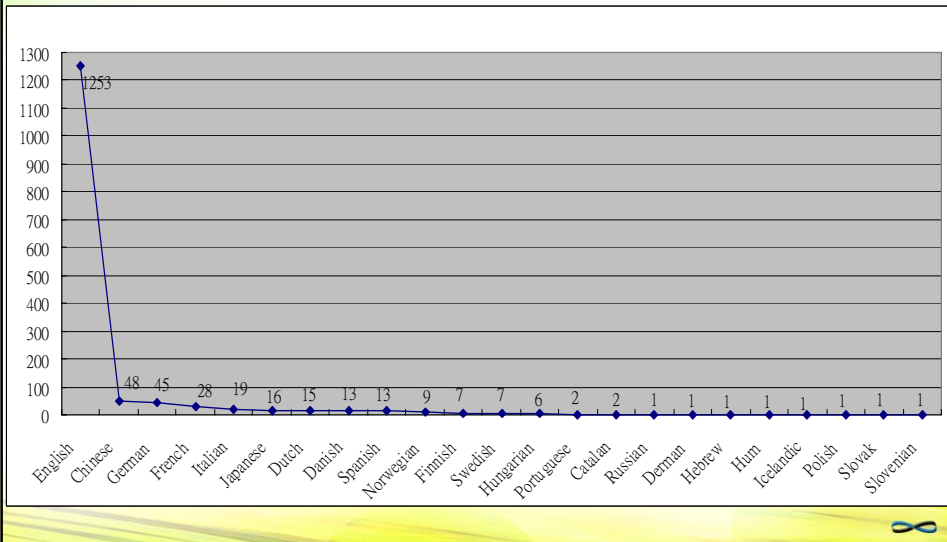


## 發表文獻的年代分佈

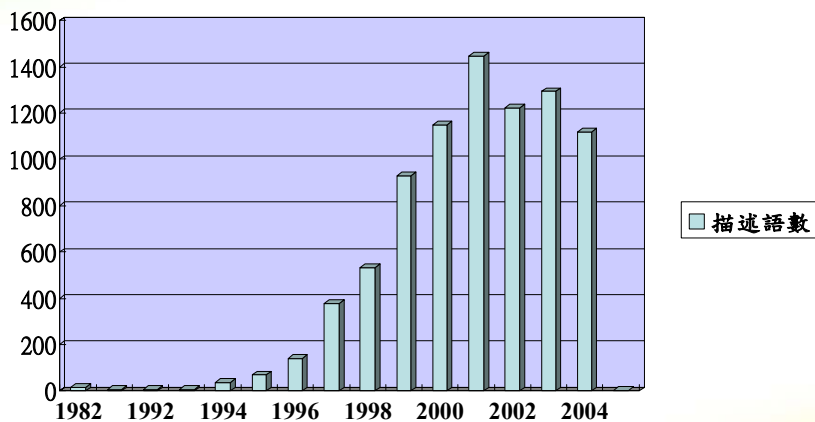




## 發表文獻的語文分佈



## 涵蓋研究議題/描述語的分佈





## 研究發現

- 文獻成長
- 發表文獻的作者
- 期刊與文獻的數量關係
- 蒐錄文獻的期刊學科主題分佈
- LISA記錄的**描述語**分佈
- **Metadata**研究與發展的主題分析



## 文獻成長<sup>1</sup>

年代	文獻數	年代	文獻數	文獻數	文獻數	文獻數	文獻數
1982	2	1994	10	1998	104	2002	205
1990	1	1995	16	1999	197	2003	208
1992	2	1996	30	2000	212	2004	186
1993	3	1997	78	2001	236	不詳	1



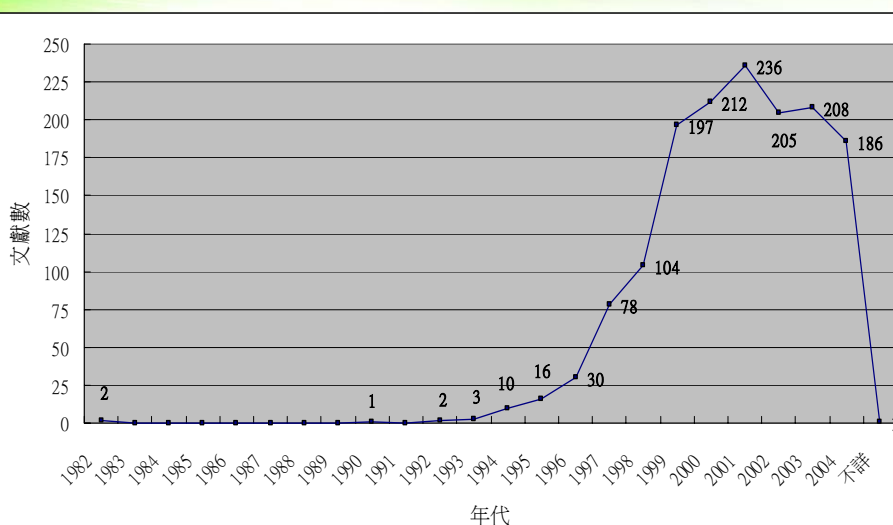


## 文獻成長<sup>2</sup>

- 年代區間(以10年為基本單位)
  - 1982-1991 – 幾乎呈現零的現象
  - 1992-2001 – 文獻是呈現成長的現象，尤其是2001年是最高峯
  - 2002-2004 – 每一年有近200篇文獻，但已有下滑的現象
  - Metadata可以算是一個新興的次研究領域或研究主題，是否會有第二次的高峯，則有待觀察
  - 幾近指數為3的倍數成長 – 1990(1 - 1)、1993(3 - 3)、1994(9 - 10)、1996(27 - 30)、1997(81 - 78)、2001(243 - 236)



## 文獻成長<sup>3</sup>



## 文獻成長<sup>4</sup>

- 議題
  - 是否因某種原因或趨勢的影響下，在某一段時間內，**metadata**形成一種熱門或新興的研究議題？
  - 自2001年後，有關**metadata**議題的文獻開始減少，可能原因是**metadata**的研究與應用已趨於平穩現象，或是**metadata**此一詞彙已轉換為其他名詞？



## 發表文獻的作者<sup>1</sup>

作者	文獻數	作者	文獻數	作者	文獻數	作者	文獻數
N. Medeiros	18	S. Proberts	10	T.C. Craven	8	M. Day	7
S. Weibel	14	J. Ahronheim	9	J. Greenberg	8	L. Dempsey	7
C. Oppenheim	11	D. Campbell	9	R. Heery	8	W. Duff	7
A. Powell	11	D. Bearman	8	W. Jones	8	<b>C-J. Wu</b>	7
<b>Y.-N. Chen</b>	10	P. Caplan	8	E Miller	8		
E. Gadd	10	<b>S. Chen</b>	8	R. Tennant	8		





## 發表文獻的作者<sup>2</sup>

- 發表文獻最多的作者 – N. Medeiros (USA)
- 排名前22名的作者與洛卡定律
  - 共發表157篇文獻，約1,491篇文獻的10.5%
  - 十分之一的1,670位作者(171)的總發表文獻量(548)少於746篇，未達洛卡定律所定義的現象
- 科學生產力根號定律的檢定
  - 1,710開根號求出值為42，亦即42位作者
  - 42位作者共發表242篇文獻
  - 242篇文獻未達1,491篇文獻的一半，因而也未符合科學生產力根號定律所定義的現象



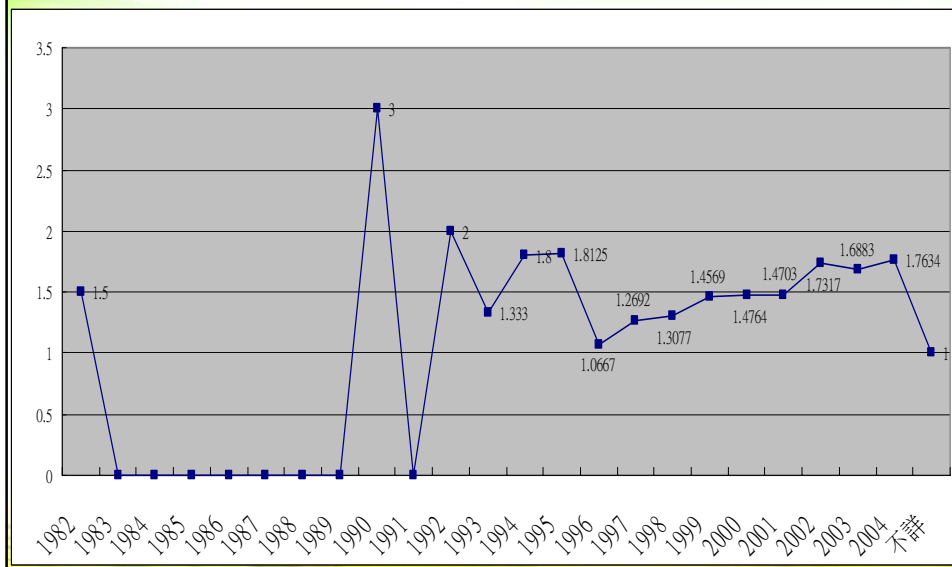
## 發表文獻的作者<sup>3</sup>

- 文獻與作者關係
  - 1982-2001 – 作者數與文獻數呈正比關係，亦即文獻數增加時，作者數也隨之增加
  - 2002以後 – 文獻數逐年下降，但作者數在2002年卻是最高的頂峯點
- 平均每篇文獻的作者數
  - Min. – Max. – 0...14
  - Mean = 1.6048
  - 以單一作者為主，但逐年有趨近2位共同作者的現象





## 發表文獻的作者<sup>4</sup>



## 期刊與文獻的數量關係<sup>1</sup>

- D-Lib Magazine (69)
- Journal of Internet Cataloguing (59)
- Library Hi Tech (44)
- Cataloging and Classification (39)
- OCLC Newsletter (30)
- Advanced Technology Libraries, Ariadne, Computer Networks, OCLC System and Services (28)
- Vine, Zeitschrift fur Bibliothekswesen und Bibliographie (24)
- International Cataloguing and Bibliographic Control (23)
- Journal of the American Society for Information Science、Journal of the American Society for Information Science Technology (15與7篇，合計22篇)



## 期刊與文獻的數量關係<sup>2</sup>

- Journal of China Society for Scientific and Technical Information (21)
- Library Collections, Acquisitions, and Technical Services (19)
- Online (18)
- Electronic Library (17)
- Biblioteche Oggi、Serials Librarian (16)
- Serials Review (15)
- Computer Networks and ISDN Systems、Information Today、Technicalities (14)
- DF Revy、Journal of Information Science、Managing Information (13)
- Canadian Journal of Information and Library Science、Informatie Professional (12)
- Bulletin des Bibliothèques de France, Bulletin of Library and Information Science, Cataloguing Australia、Journal of Library Administration、Learned Publishing (11)
- Archivaria、First Monday、Nodinfo Nytt、Online Information Review (10)



## 期刊與文獻的數量關係<sup>3</sup>

- 排名前16種期刊包括504篇文獻，佔總文獻的三分之一(33.8%)
- 排名前35種期刊包括748篇文獻，佔總文獻的二分之一(50.17%)
- 上述期刊是多數作者列為主要的投稿對象，也可被視為研究metadata議題的主要來源；其中又以英文期刊居多，也進一步解釋了metadata議題主要文獻源自於英文文獻

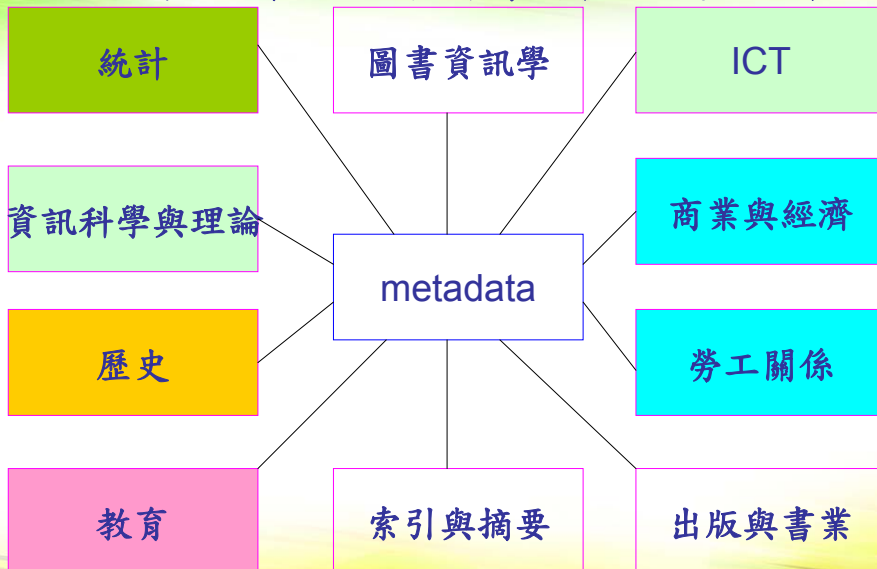


## 期刊與文獻的數量關係<sup>4</sup>

- 中文文獻是研究metadata議題的第二大主要來源(僅次於英文文獻)，總數為48篇文獻
- 蒐錄情形
  - 情報學報 (21)
  - 圖書與資訊學刊 (9)
  - 圖書館學與資訊科學 (7)
  - 教育資料與圖書館學 (6)
  - 中國圖書館學會會報 (3)
  - 資訊傳播與圖書館學 (2)



## 蒐錄文獻的期刊學科主題分佈





## LISA記錄的描述語分佈<sup>1</sup>

- 1,491篇文獻，共包括1,601個描述語，累計總次數為8,317次。
- 有12個描述語出現的累計次數超過100次以上
  - Metadata(696, 8.37%), Electronic media (370, 4.45%), World Wide Web (369, 4.44%), Internet (281, 3.38%), Formats (167, 2.01%), Digital Libraries (149, 1.79%), Bibliographic Description (126, 1.51%), Dublin Core Format (119, 1.43%), Searching (116, 1.39%), Standards (114, 1.37%), Online Cataloguing (105, 1.26%), Online Information Retrieval (105, 1.26%)



## LISA記錄的描述語分佈<sup>1</sup>

- 初步發現包括 – 1,491篇文獻是
  - 以探討Metadata議題相關的文獻為主，包括 – metadata格式(format, Dublin Core Format)與標準(Standards)
  - 由於媒體的電子化、數位化與網路化等多重趨勢衝擊下，引發Metadata此一議題的研究與探討；描述語包括 – Electronic Media, World Wide Web, Internet, Digital Libraries, Searching與Online Information Retrieval
  - 在圖書資訊界的分類編目中，也有許多研究者以書目描述 (bibliographic description) 與線上編目(Online Cataloguing)的觀點切入，探討有關Metadata此一議題

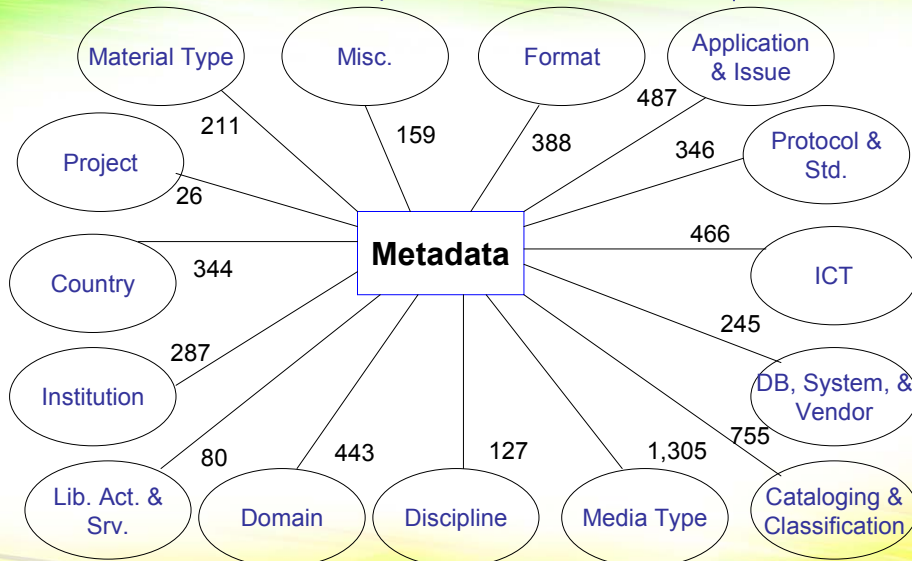


## LISA記錄的描述語分佈<sup>2</sup>

- 排除出現1次與2次的描述語，共計1,266個描述語。此外, Metadata(696次)此一描述語亦排除在外
- 以出現3次以上的描述語(335個)為分析對象，占描述語總出現次數的80%
- 發現
  - Metadata與ICT有極為密切的關係
  - Metadata與圖書資訊界的分類編目有高度的相關性
  - Metadata的興起與媒體電子化與網路化確實存有極大的關聯性
  - 數位圖書館的興起與與蓬勃發展，也引起對Metadata此一議題的廣泛研究與應用



## LISA記錄的描述語分佈<sup>3</sup>



## Metadata研究與發展的主題分析

- 以2004年186篇的LISA文獻記錄內容為分析對象
- 重點
  - Metadata應用的領域範疇
  - Metadata在特定國家的發展
  - Metadata應用的資料類型
  - Metadata相關格式標準與技術
  - Metadata的管理



## Metadata應用的領域範疇

主類目	次類目	細類目
社群導向	實體	檔案館·圖書館·博物館·政府機構·出版社·大學
	虛擬	網站·主題 (Gateway)·數位學習·知識管理·學術傳播系統·內容管理系統·語意網·學術論文預刊本 (ePrint)·數位圖書館·eScience·聯合目錄
學科導向		圖書館學·統計學·企業管理·科學·領域本體·書目計量
功能導向		智財權·數位保存·聯合查詢 (Federated Search)·文本探勘 (Text Mining)·搜尋引擎





## Metadata在特定國家的發展

- 多數文獻內容討論重點以特定機構或計畫的個案研討為主
- 以整體國家為研究範圍的文獻只有2篇
  - 中國大陸
  - 澳洲



## Metadata應用的資料類型

- 影像與圖像
- 音樂（譜）
- 書目
- 多媒體
- 電子期刊
- 電子資源
- 電子論文







## Metadata相關格式標準與技術

Metadata 格式標準類型		內容
資料結構標準	一般性	DC
	特定社群	MARC, EAD, LOM, MODS, MPEG
	整合性	METS
資料內容標準		CCO
概念模型與架構		FRBR, RDF
標誌語言		XML
互通協訂		DOI, OAI, Z39.50



## Metadata的管理

其他 metadata 議題的類型	內容
基本問題	<ul style="list-style-type: none"><li>◆ Metadata 目的、必要性與價值</li><li>◆ Metadata 教育</li><li>◆ Metadata 專業人員</li></ul>
作業方法論	<ul style="list-style-type: none"><li>◆ Metadata 的作業原則與工作流程</li><li>◆ Metadata 標準的在地化之方法論</li><li>◆ Metadata 品質</li></ul>
自動化機制	<ul style="list-style-type: none"><li>◆ 自動化 metadata</li><li>◆ Metadata Registry</li><li>◆ 管理系統與軟體的發展</li></ul>



## 結語與建議

- 完成其他年份的質性內容分析
- 分析文獻共同作者數與跨學科領域的相互關係
- 採用布萊福定律（**Bradford's Law**）來檢驗 **metadata** 此一議題的文獻分佈情形
- 採取其他方法（如：共字分析），以建構 **metadata** 的知識圖
- 未來擴大選取的研究樣本（如：**Web of Science** 或 **Scopus** 資料庫），以更全面性分析不同學科間有關 **metadata** 議題的研究發展與實務應用



## 參考書目

- 蔡明月. (民92). 資訊計量學與文獻特性. 台北市: 國立編譯館.



## 誌謝

- 中央研究院計算中心  
王智豐、張琬人與龔家珍小姐
- 淡江大學資訊與圖書館學系  
紀孫傑同學



謝謝， 並歡迎指教！

