



從電子資源的書目控制 到數位典藏的後設資料

陳亞寧

中央研究院計算中心
April 19, 2005



大綱

- LIS Approach
- Internet Search Engine Approach
- Data mining Approach
- Topic Map Approach
- Metadata Approach
 - Locan Dempsey & Rachel Heery
 - 陳雪華教授
 - Anne Swetland-Gilliand
 - MetaMap
- NDAP Current Status
- NDAP Findings
- Suggestion for NDAP





LIS Approach

- 以分類編目為基礎，進行資源的整理與組織。
- 結合**Spider/Crawler**軟體進行前端的資料蒐集。
- 通常經過篩選，並依據既定的著錄規範進行分編與組織。
- 提供查檢與瀏覽等兩種使用方式。
- 資料新穎性較不具時效。
- 以**W³**網頁方式呈現。
- 成本較高，而且資料正確性不易維護；如**URL**、資源的涵蓋主題與收錄範圍等。



LIS Approach — WebPage



BUBL Information Service

BUBL LINK Catalogue: Selected Internet resources covering all academic subject areas

[Subject Menus](#) | [Countries](#) | [Types](#) | [BUBL UK](#) | [BUBL Archive](#)

[Advanced Search](#)

[A](#) | [B](#) | [C](#) | [D](#) | [E](#) | [F](#) | [G](#) | [H](#) | [I](#) | [J](#) | [K](#) | [L](#) | [M](#) | [N](#) | [O](#) | [P](#) | [Q](#) | [R](#) | [S](#) | [T](#) | [U](#) | [V](#) | [W](#) | [X](#) | [Y](#) | [Z](#)

[000 Generalities](#)

Includes: computing, Internet, libraries, information science

[100 Philosophy and psychology](#)

Includes: ethics, paranormal phenomena

[200 Religion](#)

Includes: bibles, religions of the world

[300 Social sciences](#)

Includes: sociology, politics, economics, law, education

[400 Language](#)

Includes: linguistics, language learning, specific languages

[500 Science and mathematics](#)

Includes: physics, chemistry, earth sciences, biology, zoology

[600 Technology](#)

Includes: medicine, engineering, agriculture, management

[700 The arts](#)

Includes: art, planning, architecture, music, sport

[800 Literature and rhetoric](#)

Includes: literature of specific languages

[900 Geography and history](#)

Includes: travel, genealogy, archaeology

BUBL uses the Dewey Decimal Classification system as the primary organisation structure for its catalogue of Internet resources. The Dewey Decimal Classification is (c) 1996-2005 OCLC Online Computer Library Center. Used with permission.

[E-LIS](#) | [CDLR Projects](#) | [Contacts and Credits](#)

BUBL Information Service, [Centre for Digital Library Research](#), [Strathclyde University](#), Glasgow G1 1XH, Scotland
Tel: 0141 548 4752 Email: bubl@bubl.ac.uk



LIS Approach — WebPAC

← → × 🏠 🔍 🌐 📄 📁 📧 📧 📧

[上一筆 Previous Record](#)
[下一筆 Next Record](#)
[查詢其他 Another Search](#)
[重新查詢 Start Over](#)
[MARC顯示 MARC Display](#)
[轉錄/儲存 Export](#)
[我要預約 Request](#)

(檢索歷史) ▾

書名 ▾ digital library
 [全部館藏](#) ▾

相關資
other resources

書名 Intelligence [electronic resource] : new visions of AI in practice
出版項 New York, NY : Association for Computing Machinery, c1999-

請瀏覽網址 :

[ACM Digital Library full text\(1999-2001\)](#)

叢書名	Digital Library
附註	Quarterly Vol. 10, no. 1 (Spring 1999)- Electronic journal Title from introductory page, viewed June 3, 2002 Subscription required for access to full text Text (electronic journal) in PDF files and HTML format. Abstracts, references and table of contents in HTML format
Summary	Reports on research and development in the field of artificial intelligence
附註	Also available in print System requirements: Web browser and Adobe Acrobat Reader Mode of access: World Wide Web via the Internet A publication of the ACM Special Interest Group on Artificial Intelligence Made available through: ACM Digital Library
標題	Artificial intelligence Artificial intelligence -- Technological innovations
Unif title	Intelligence (New York, N.Y. : 1999 : Online)
其他書名	New visions of AI in practice



LIS Approach – Database



Search Jake - PostgreSQL Server

Search Details		Bib Details for Linking	
Search Title	<input type="text"/>	Truncate: <input checked="" type="checkbox"/>	Volume <input type="text"/>
Search ISSN	<input type="text"/>	Sort (title): <input checked="" type="checkbox"/>	Issue <input type="text"/>
Search jake id	<input type="text"/>	XML: <input type="checkbox"/> (experimental)	Page <input type="text"/>
Search subjects	<input type="text"/>		Year <input type="text"/>
<input type="button" value="Submit"/> <input type="button" value="Clear"/>			

Journal of Documentation [display search \(title\) variations](#)

- o jake id **1941**, type **serial**, issn **0022-0418**
- o indexed by 19 databases
- o fulltext in 1 databases
 - [Documentation](#) (lcsh)
 - [Z1007.J9](#) (lc)
 - [010.5](#) (dewey)

Current Bib Details

Volume

Issue

Page

Year

links	resource	provider	citations	fulltext
	Biosis	Ovid	+	-
	CAB Abstracts	Ovid	+	-
	CINAHL	Ovid	+	-
	Compendex	Ovid	+	-



Internet Search Engine Approach

- Yahoo型

- 前端以Spider/Crawler進行資源蒐集，後端以人工方式進行分編。
- 著錄項目不似LIS Approach多，以註解為主要特色。
- 同時具備查檢與瀏覽功能。

- Google型

- 前端以Spider/Crawler進行資源蒐集，後端以電腦程式進行自動分編。
- 使用界面簡易，較為大眾接受與適應。
- 速度快、查檢率(recall)高為主要特色，但精準率(precision)則是一大隱憂。



LIS Approach

***Search Engine —
Yahoo***

***Search Engine —
Google***

Rich description

Simple description

More specific

More general



More Structured

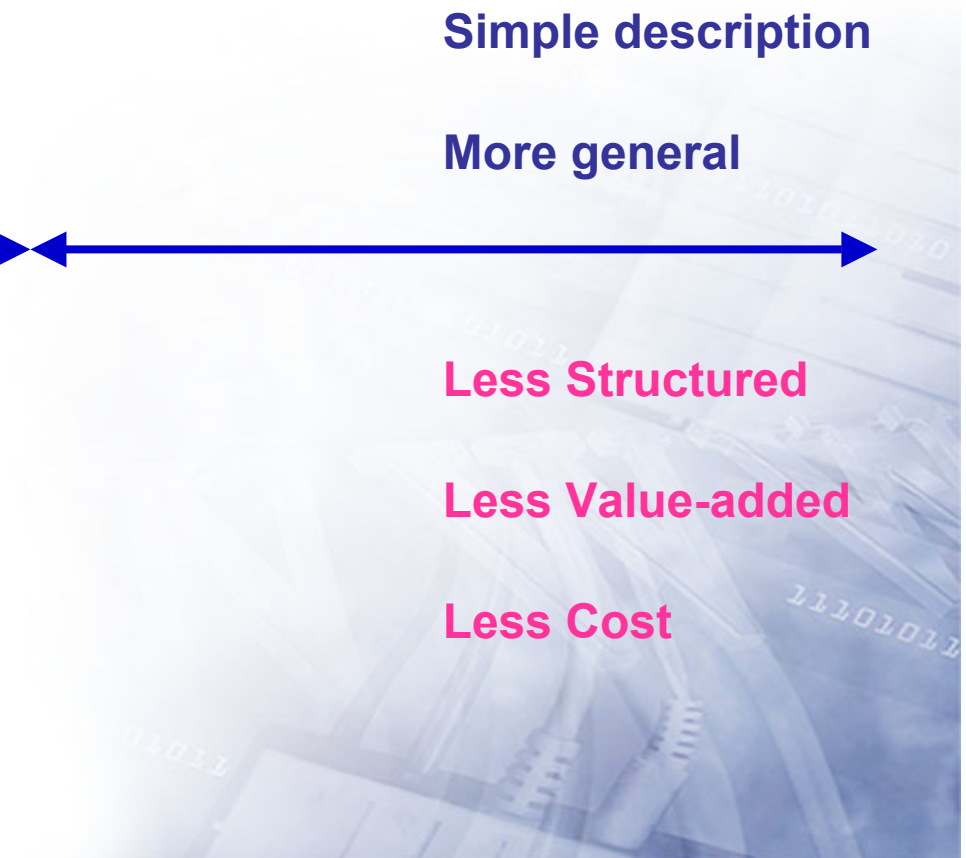
Less Structured

More Value-added

Less Value-added

More Cost

Less Cost





Data Mining Approach

- IBM提出
- 分析軟體類型
 - 統計型(statistical)
 - 機器學習型(machine learning)
 - 神經網路型(neural networks)
- 特質
 - 分析使用者的行為模式
 - 找出使用的特質及使用關聯性
 - 發現最符合使用者需求的服務內容
 - 預測未來的使用行為與需求
- Example 1: Diaper and Beer
- Example 2: Amazon





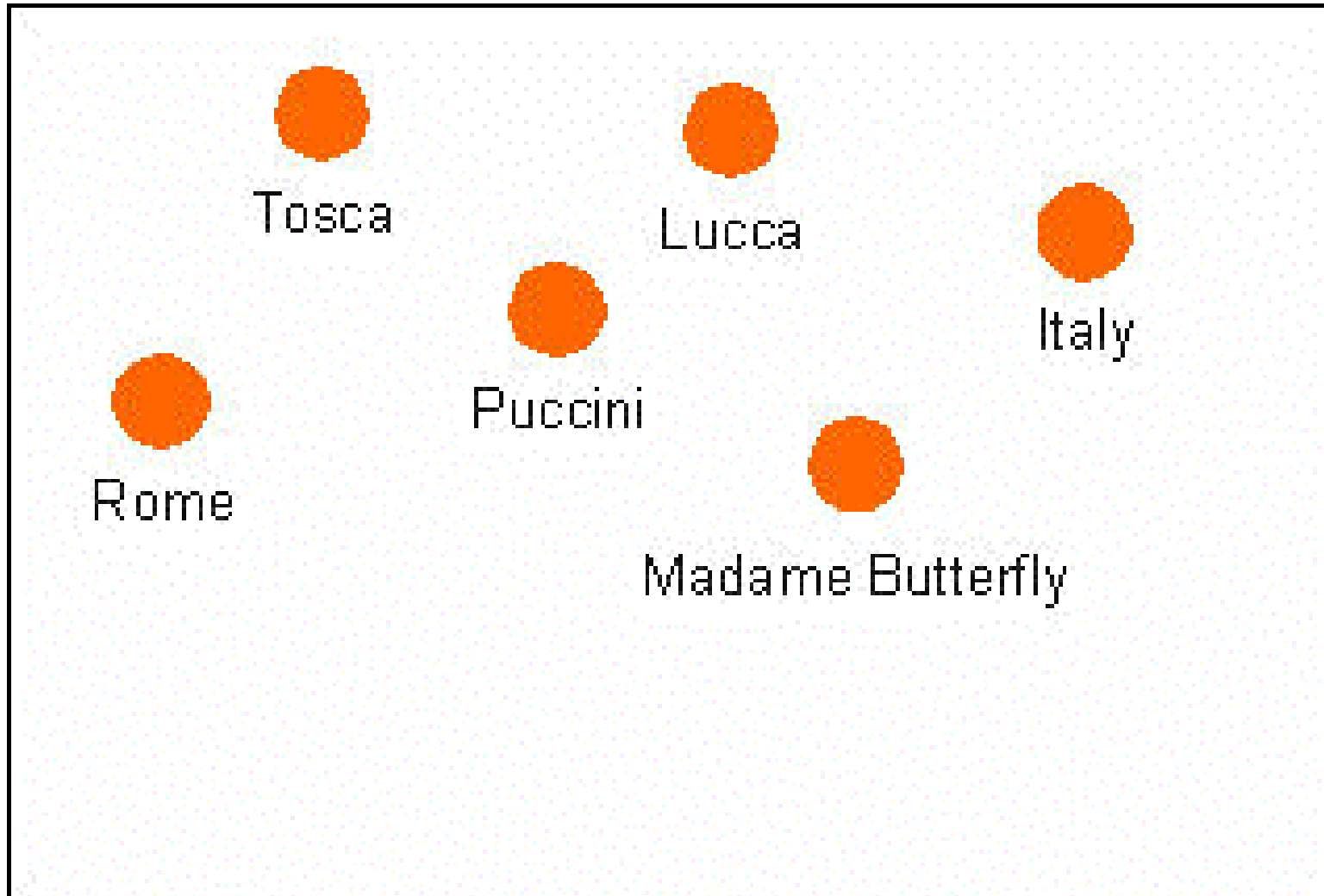
Topic Map Approach¹

- 2003年已成為ISO標準
- 概念
 - 主題及其類型(topic and topic type)
 - 關係實例(occurences)
 - 關聯性(association)



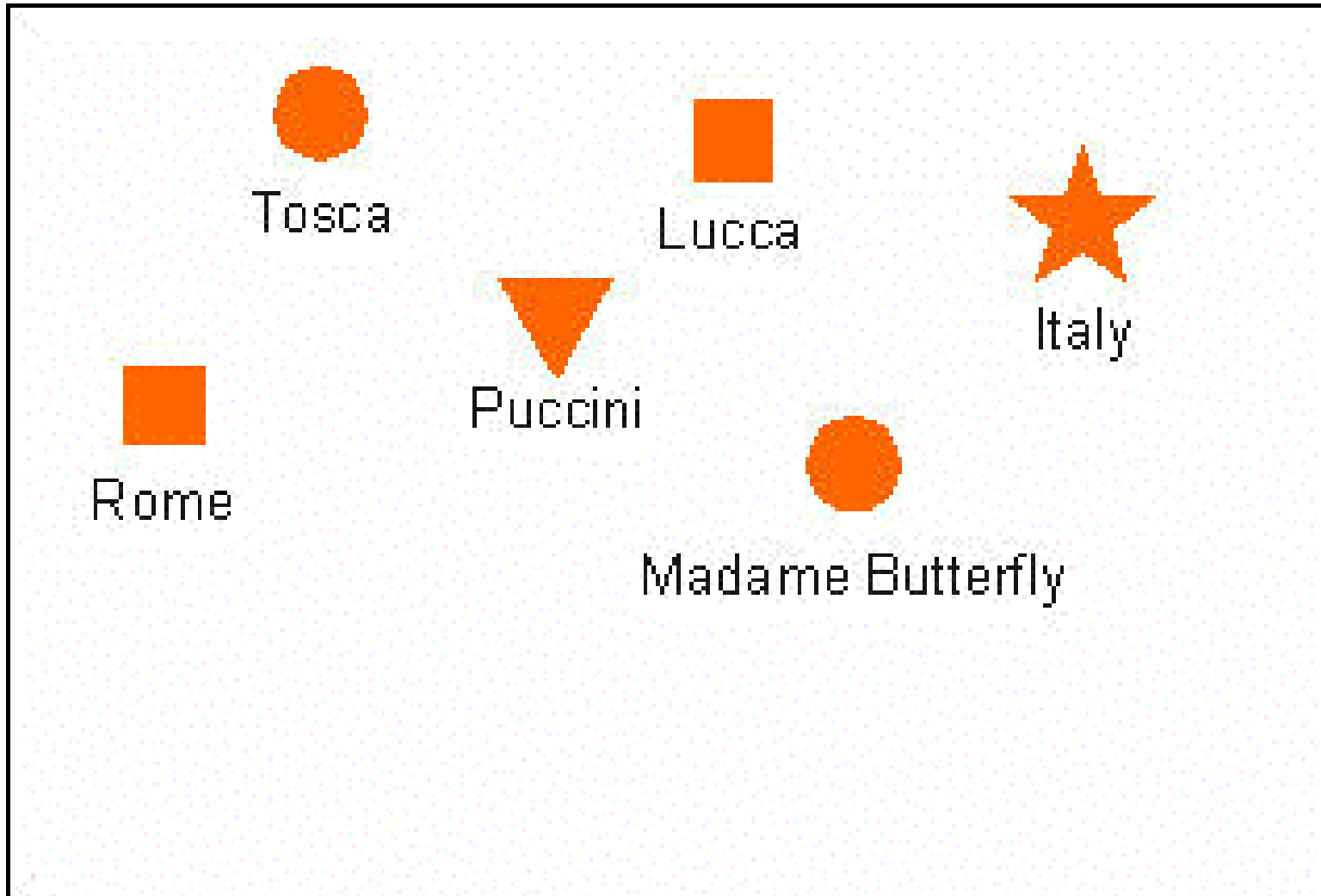


Topic Map Approach²



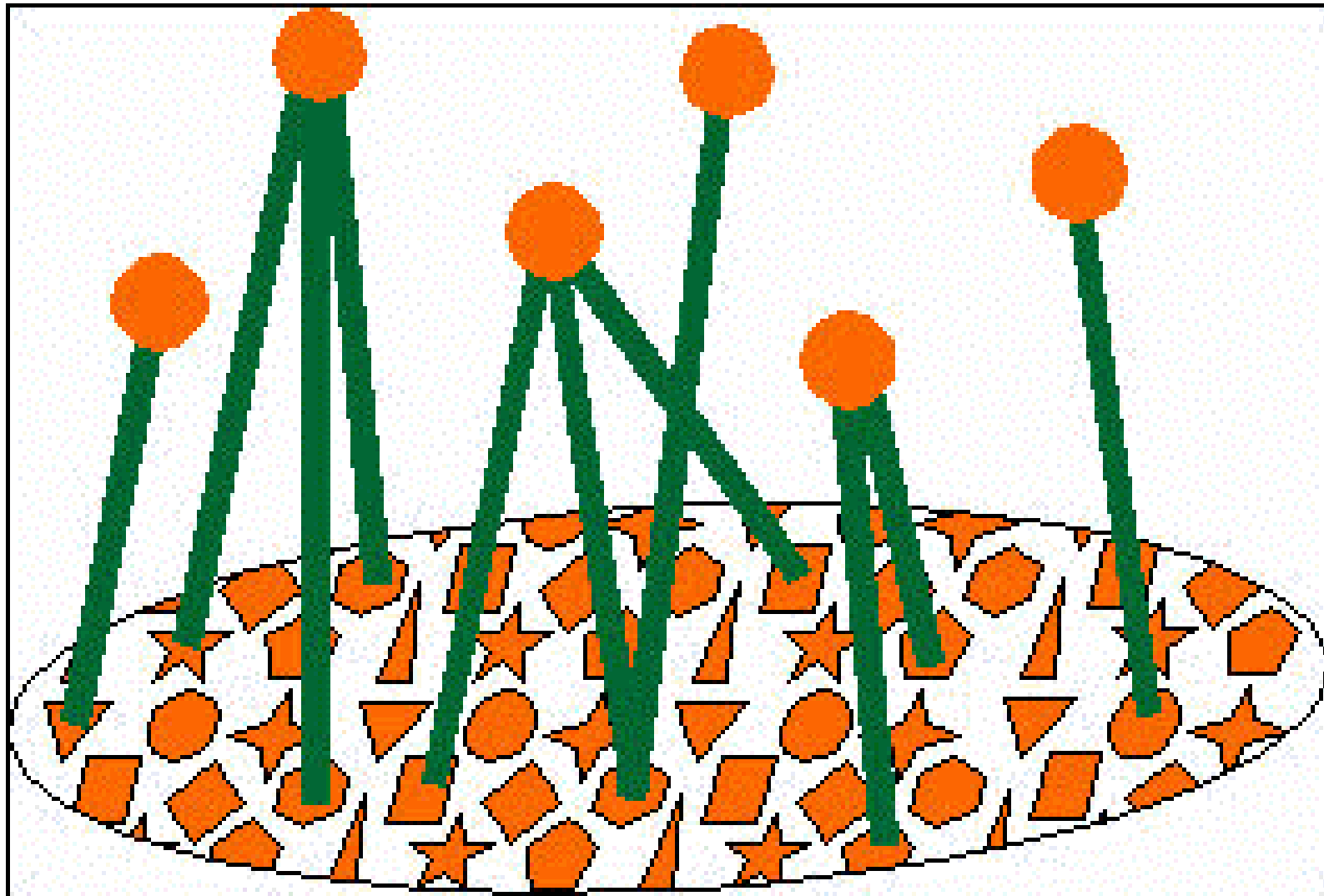


Topic Map Approach³



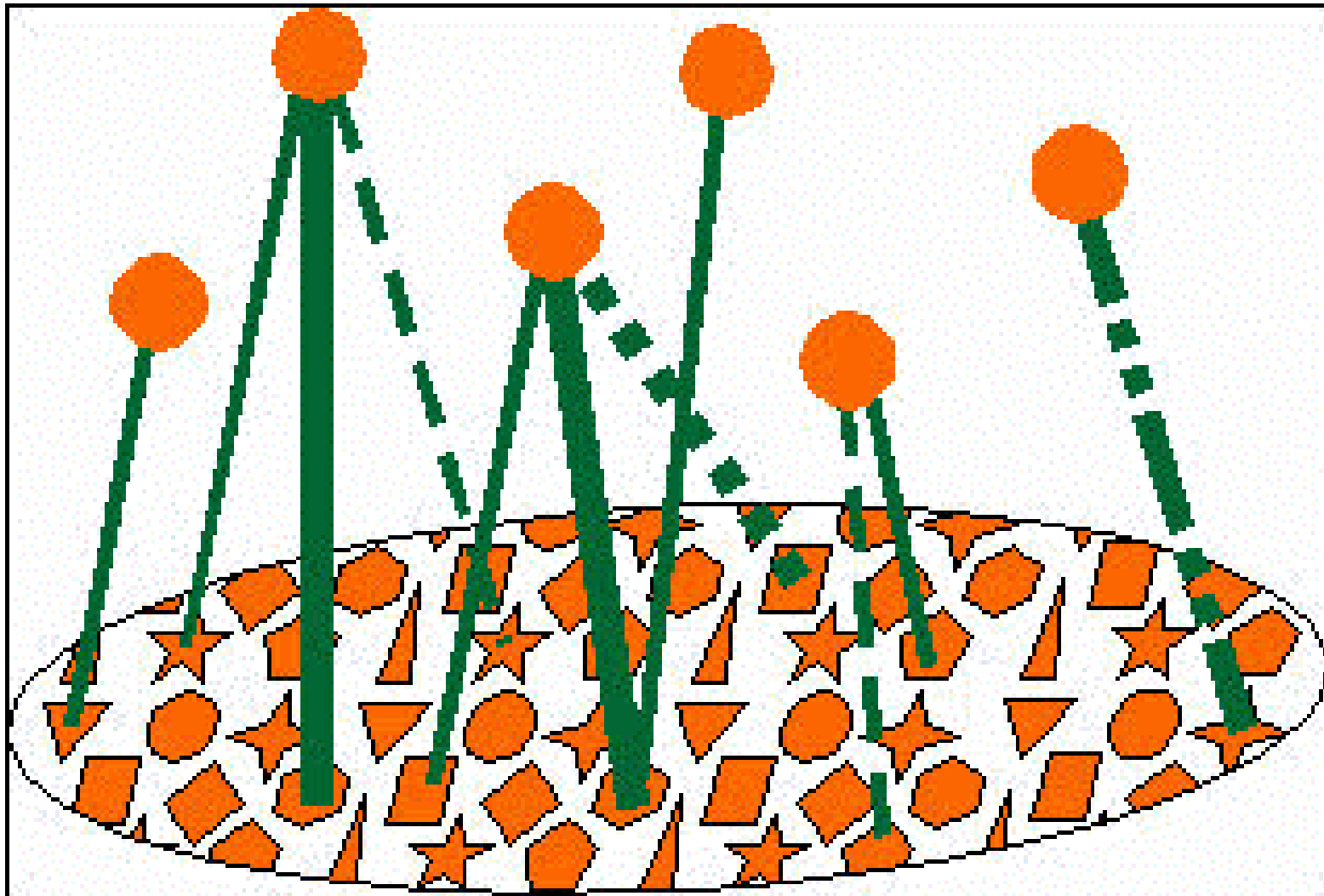


Topic Map Approach⁴



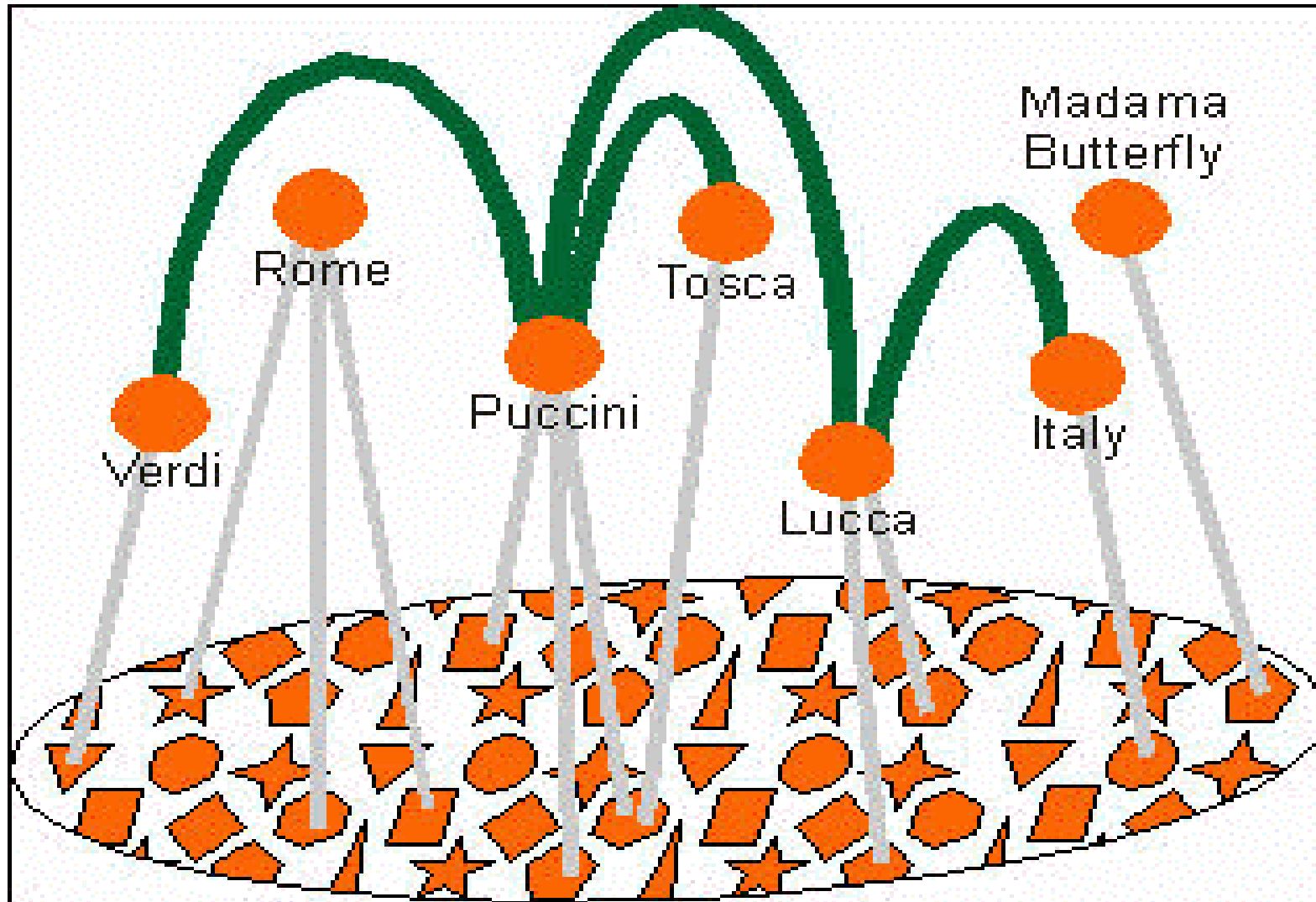


Topic Map Approach⁵



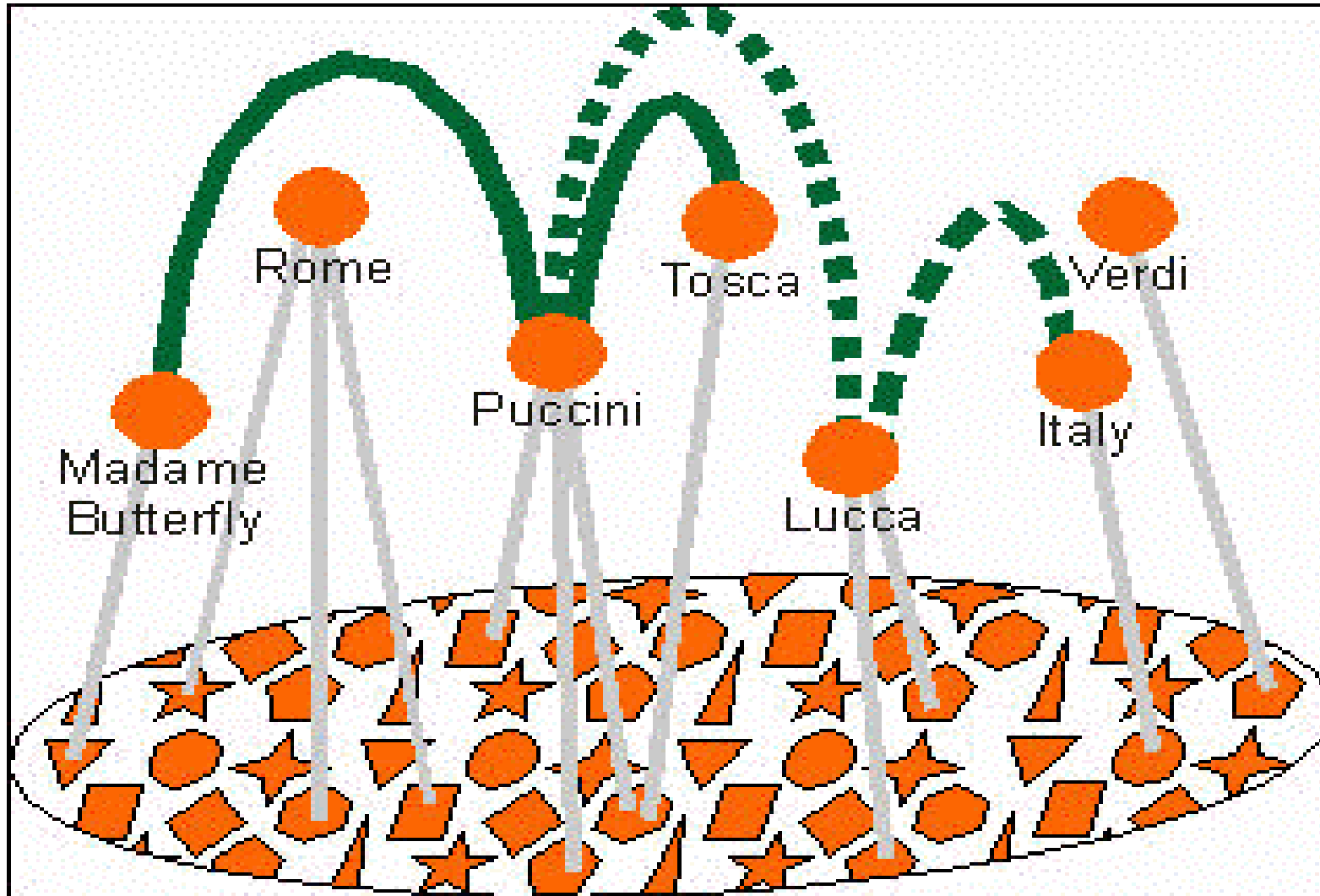


Topic Map Approach⁶





Topic Map Approach⁷





Topic Map Approach⁸ — HighWire

The screenshot shows the HighWire Press website interface. At the top, there is a navigation bar with the following links: [Adv. Search](#), [My Email Alerts & PDA](#), [For Institutions](#), [For Publishers](#), [About](#), [Contact](#), and [Help](#). Below the navigation bar, a welcome message reads: "Welcome to HighWire Press! [About our features...](#)" and a prompt to "Sign in or register for access to all HighWire Press customization features".

The main content area is divided into three columns:

- Left Column:** Titled "FULL TEXT I HAVE ACCESS TO", it lists the number of journals with various access options: 57 free trials, 28 free journals, 195 free back issues, and 706 pay-per-view. It also highlights "plus! FREE access to the full-text of cited references in all 847 HighWire-hosted journals". At the bottom, there is a banner for "Washington DC Principles for FREE ACCESS to SCIENCE".
- Middle Column:** Titled "Quick search HighWire + Medline: [help](#) / [advanced search](#)", it contains a search form with fields for Author (e.g., Smith, JS), Keywords (gene), Year, Vol, Page, and In (My Favorite Journals, HighWire-hosted journals, HighWire-hosted journals + Medline). Below the search form, there is a section for "Browse articles:" with links for Biological Sciences, Physical Sciences, Medical Sciences, and Social Sciences. A "TopicMap" link is circled in pink. At the bottom, there is a section for "Browse HighWire-hosted journals:" with links for alphabet, topic, publisher, and other lists, followed by a grid of letters A through M and N through Y, and an "all" link.
- Right Column:** Titled "TO GET MORE FREE FEATURES", it includes a "Register in 1 minute, or sign in:" section with email and password fields, a "remember me" checkbox, and a "sign in" button. Below this is a link for "[I forgot my password]". Further down, there is a section for "SIGN IN FOR FREE SPECIAL FEATURES" with sub-sections for "My Favorite Journals" (Track your favorite journals), "My Email Alerts" (Monitor the publication of articles and issues matching your criteria), and "My Access" (Sign in to all your personal subscriptions with one click).

At the bottom of the right column, there is a "MY SITEBAR" section with the text: "A tool for quick access to searching and your Favorite".



Topic Map Approach⁹ — HighWire

Microsoft PowerPoint - [001]

TopicMap - Microsoft Internet Explorer

新細
大綱

96
97
98
99
100
101

Search for Topics: Topics: 367

CLICK: to select a topic, DRAG: to move, DOUBLE-CLICK: to view document lists in your web browser
TYPE: partial topic name to search, ENTER: to go to each matching topic
BOTTOM BUTTONS: change font size, reset page and get additional help

繪圖(R) Applet HWTree.HyperTreeApp started 網際網路

投影片 99 / 121 預設簡報設計 中文(台灣)

開始 收件匣 - O... HighWire P... Microsoft P... TopicMap - ... 下午 05:49



Topic Map Approach¹⁰ — HighWire

HighWire Press - Microsoft Internet Explorer

TopicMap - Microsoft Internet Explorer

Search for Topics: Topics: 366

CLICK: to select a topic, DRAG: to move, DOUBLE-CLICK: to view document lists in your web browser
TYPE: partial topic name to search, ENTER: to go to each matching topic
BOTTOM BUTTONS: change font size, reset page and get additional help

Applet HWTree.HyperTreeApp started

開始 | 收件匣 - O... | HighWire P... | Microsoft P... | TopicMap - ... | 下午 05:49



Metadata Approach

- 全球趨勢
 - Locan Dempsey & Rachel Heery
 - 陳雪華教授
 - Anne Swetland-Gilliand
 - MetaMap
- 數位典藏國家型科技計畫(National Digital Archives Programme in Taiwan, NDAP)現況



全球發展現況—文獻分析¹

- Dempsey and Heery (1998) 依據下列要點進行Metadata分析
 - 使用環境(environment of use)
 - 功能(function)
 - 產生(creation)
 - 結構(designation)
 - 相關的查詢協定(associated search protocols)
 - 狀態(status)



全球發展現況－文獻分析²

	Band 1	Band 2	Band 3
使用環境	Global Internet search services	Selective Internet search services; directory service	Descriptions of scholarly collections; other importance repositories
功能	Location	Discovery, location, selection	Location, selection, evaluation, analysis, documentation



全球發展現況－文獻分析³

	Band 1	Band 2	Band 3
產生	Robot generated	Robot plus manual input	Intellectual expertise required, often involving dedicated 'information' staff
結構	Unstructured	Attribute value pairs; limited structure	Subfields; qualifiers, structured mark up



全球發展現況－文獻分析⁴

	Band 1	Band 2	Band 3
相關查詢協定	http with CGI form interface	http with CGI form interface; directory service protocols (WHOIS++, LDAP) with query routing (CIP)	Z39.50 SGML browsers and querying
狀態	Proprietary	Emerging Internet Standards	Domain specific standardisation



全球發展現況－文獻分析⁵

- 陳雪華教授(民86)將Metadata區分爲
 - － 早已普遍使用－MARC, PICA
 - － 描述科技文獻－BibTex, EELS, EEVL, RFC 1807
 - － 描述人文及社會科學資源－ ICPSR SGML Codebook Initiative, TEI Headers
 - － 描述政府資訊－ GILS
 - － 描述地理空間性資源－ CSDGM/FGDC
 - － 描述博物館藏品與檔案特藏－ CDWA, CIMI, EAD
 - － 描述大量網路資源－ Dublin Core(DC), IAFA/WHOIS++ Templates, LDIF, SOIF, URCs
 - － 其他－ Warwick Framework



全球發展現況－文獻分析⁶

- Gilliland-Swetland (2000)將Metadata區分爲
 - 管理(administrative)
 - 描述(descriptive)
 - 長期保存(preservation)
 - 技術(technical)
 - 應用(use)





全球發展現況－文獻分析⁷

- Gilliland-Swetland (2000)將Metadata屬性/特質區分為
 - Metadata來源－internal (如file names and header information) vs. external (如cataloging recs, legal information)
 - 產生方式－automatic vs. manual
 - 性質－creator vs. expert
 - 狀態－static vs. dynamic, long-term vs. short-term
 - 結構－structured vs. unstructured
 - 語意－controlled vs. uncontrolled
 - 層次－collection vs. item



全球發展現況－文獻分析⁸

- Gilliland-Swetland (2000)依數位資訊物件的生命週期將Metadata區分為幾種層面
 - 產生(creation and multi-versioning)
 - 組織(organization)
 - 查詢與檢索(searching and retrieval)
 - 長期保存與淘汰(preservation and disposition)

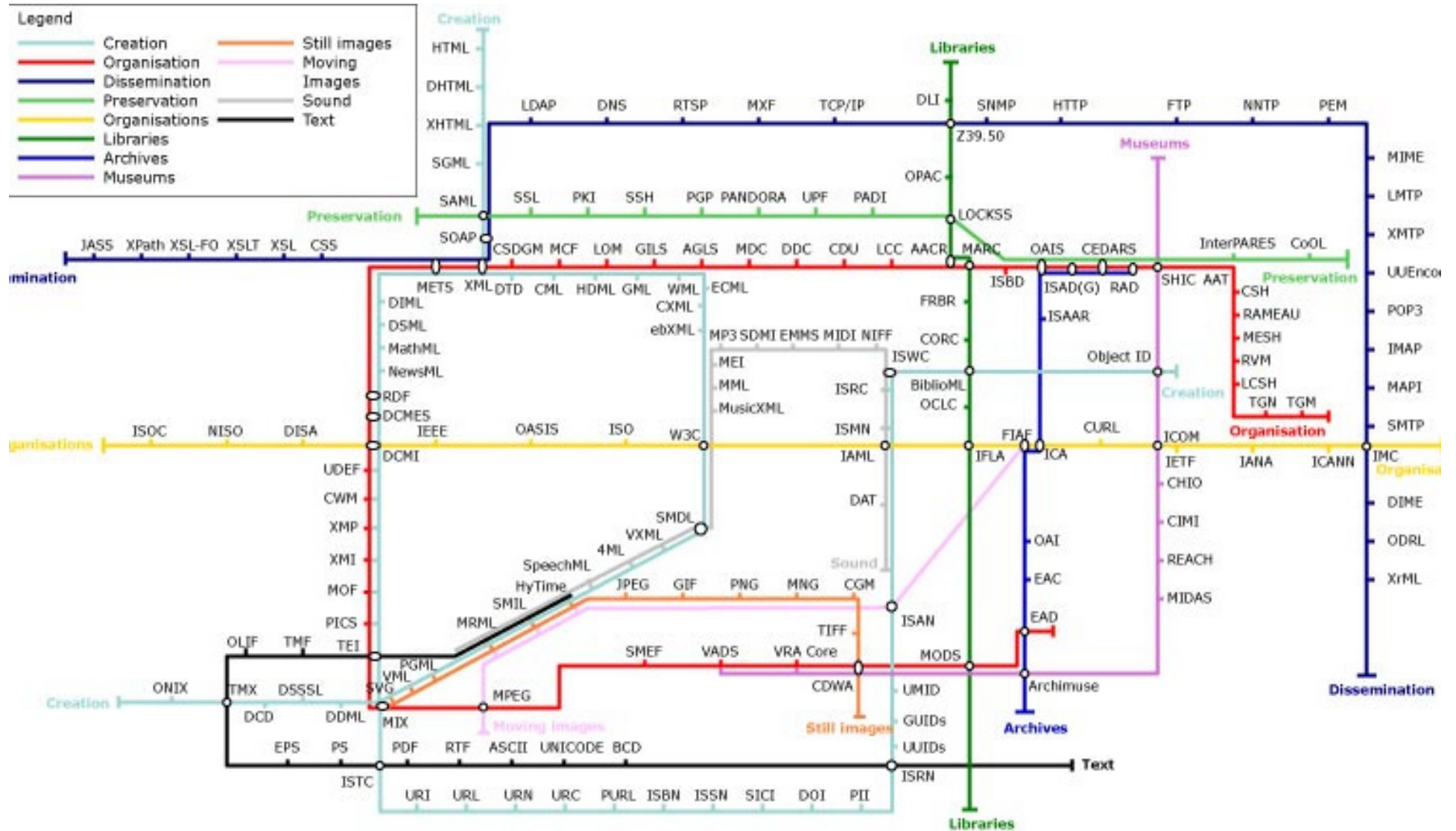


全球發展現況－文獻分析⁹

- MetaMap將Metadata區分爲
 - －功能－產生(creation), 組織(organisation), 散播(dissemination)與長期保存(preservation)
 - －社群－圖書館, 博物館, 檔案館等
 - －檔案類型－圖片(still images), 動畫(moving images), 聲音(sound), 文字(text)
 - －機構組織－ISOC, NISO, DISA, DCMI, IEEE, OASIS, ISO, W3C, IAML, IFLA, FIAF, ICA, CURL, ICOM, IANA, ICANN, IMC



全球發展現況—文獻分析¹⁰





全球發展趨勢之分析

- **Functionality** — Resources Discovery, Description, Administrative, Technical, Use, Preservation, & Rights
- **Domain** — Digital Library/Museum/Archive, eLearning/LOM, E-Government, Science, E-Commerce & E-Business
- **Community** — Library, Museum, Archives, & Internet Community
- **Creation** — automation, semi-automation
- **Level** — Data Structure, Data Content, Data value
- **Granularity** — Item, Collection/CLD, & Full-text/TEI Lite
- **Identifier** — DOI, GUID, ISBN, ISAN, ISSN, MIX, PII, PURL, SICI, UMID, URC, URI, URL, URN, UUID.
- **File Format** — GIF, HyTime, JPEG, MNG, MPEG, PNG, SMIL, SVG, TIFF, VML.
- **Conceptual Model** — FRBR, CIDOC CRM, Records Continuum Model, OAIS Reference Model, RDF Model
- **Interchange & Interoperation** — XML DTD and Schema, and RDF
- **Protocol** — OAI-PMH, SOAP, LDAP
- **Framework** — Warwick Framework, RDF



NDAP Current Status¹ — MD Stds.

- 藝術博物館
- 檔案館
- 圖書館
- 生物
- 空間資訊
- 電子商務
- 概念模式
- OAI-PMH
- DC
- XML DTD





NDAP Current Status²

Authority Control

- 人名居多 – MARC21-Authoriy Format
- 藝術類的詞彙控制
- 生物 – Species 2000
- GIS – CSDGM & ADL GCS





NDAP Current Status³

Selection of MD Stds.

- Data Structure
 - CDWA與EAD最多。
 - Data Value
 - MARC21-Authority Format與Species 2000最多。
 - Data Content
- 描述與組織為主
 - 以學科、社群、資料類型為三大主軸
 - 整體而言，著重在深化的學科知識組織，同時兼具社群與資料類型等多重取向的描述。



NDAP Results

- 功能訴求 — 結構化描述與組織為主
- 領域與社群 — 朝多元化發展
- 數位檔案與產生方式 — 以人工方式為主
- 描述深度 — 以資料結構與資料值為主
- 描述層次 — 以典藏品為主
- 其他 — XML DTD, 應用特徵檔與Metadata Registries



Application Profile¹

Simple description

Rich description

More general

More specific



DC, VRA, etc.

Qualifier
Simplification

CDWA, EAD, etc.



Less cost

More cost

Less structured

More structured

Less value-added

More value-added



全球趨勢與NDAP的MD比較¹

	全球	NDAP
功能	發掘、描述、管理、技術、使用、長期保存與著作權管理	以發掘與描述為主，其餘為輔
領域	數位典藏、數位學習、電子化政府、科學、電子商務	以數位典藏為主 另有一計畫已跨入電子商務
社群	圖書館、博物館、檔案館、標本館、電腦科技與網路虛擬自治社群	圖書館、博物館、檔案館與標本館



全球趨勢與NDAP的MD比較²

	全球	NDAP
產生方式	自動、人工與半自動	人工
描述深度	館藏合集、典藏品、內容全文	典藏品與內容全文
描述層次	資料結構、資料內容與資料值	以資料結構與資料值為主
資源識別碼	數位與傳統	傳統



全球趨勢與NDAP的MD比較³

	全球	NDAP
檔案格式	靜態與動態	靜態與動態
概念模式	FRBR、CIDCO CRM、Records Continuum Model、 OAIS Reference Model、RDF Model	FRBR Model
檢索與界面協定	OAI-PMH、SOAP、 LDAP；CSS、XSL、 XSLT	OAI-PMH



全球趨勢與NDAP的MD比較⁴

	全球	NDAP
框架	RDF	無
應用特徵檔	▲	▲
註冊與登記	▲	▲



NDAP — Findings¹

- 應用情形 — 多元化、精緻化，並以結構性後設資料為主
 - 社群: 圖書館、博物館、檔案館、標本館
 - 學科: 藝術、人文、社會與生物多樣性
 - 資料類型: 拓片、善本、照片、影音、書畫、田野調查、古文契書等
 - 功能: 發掘、描述、技術、管理、使用等



MD Used in NDAP

*Community
Layer*

Library, Archives, Museum, Herbarium

*Data Type
Layer*

**Rubbing, Rare Book, Photo, Audio, Film
Specimen, Archive, Field Report, etc.**

*Discipline
Layer*

**Art, Archaeology, History, Ethnology
Linguistics, Biodiversity, Genomics, etc.**

*Function
Layer*

**Resource Discovery, Description, Technical,
Management, and Use, etc.**

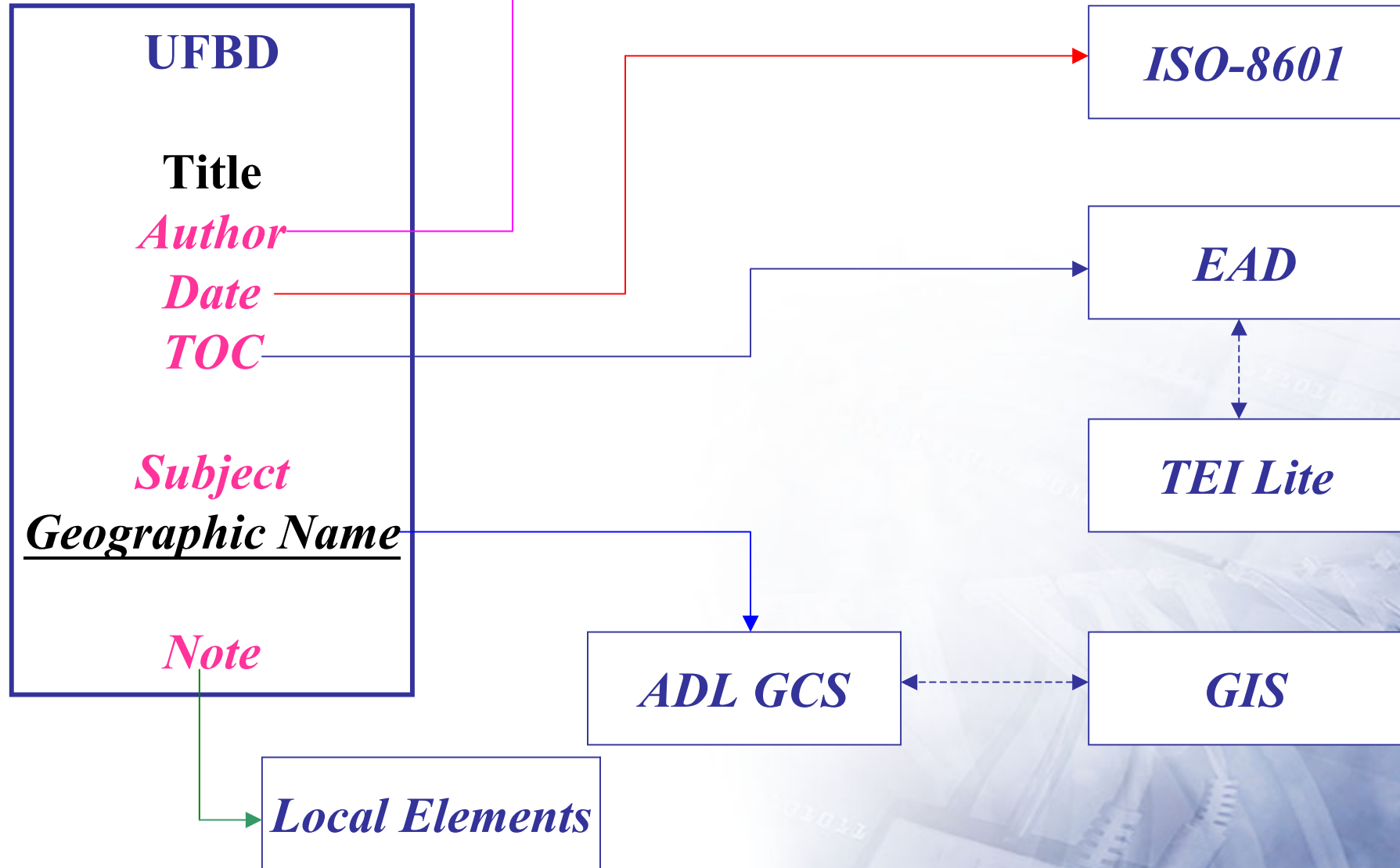


NDAP — Findings²

- 應用方式 — 多種標準與平行政策
 - 2套以上的資料結構
如臺北藝術大學電子影音資料庫增值應用之研究 — ECML與XrML
 - 資料結構與資料值各1
如臺灣原住民數位典藏計畫 — CIDOC Ethno Core Dada Stds., CDWA, MARC21 — Authority Format
 - 2套以上的資料值
故宮文物數位典藏系統之開發與建立 — AAT & MARC21 — Authority Format
 - DC & OAI-PMH

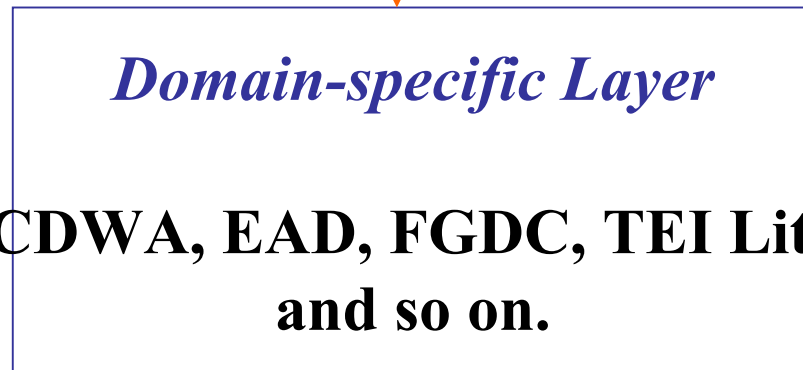
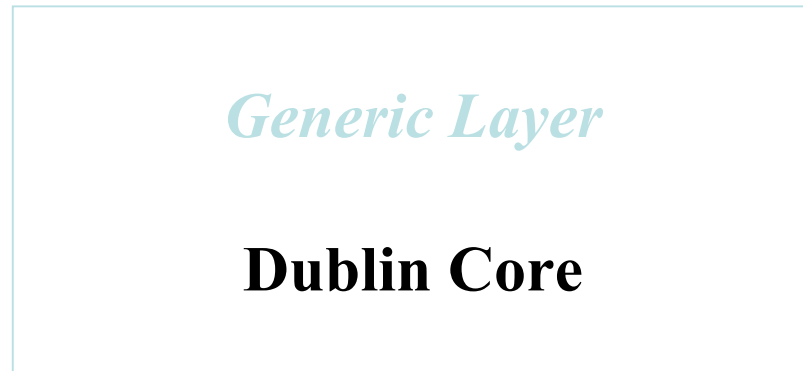


Application Profile²





Parallel Strategy



For the public, and
a base for cross-domain
interoperability.

For a discipline-oriented,
in-depth, and precise search.



NDAP — Findings³

- 描述層面 — 以物(典藏品)、人與地理空間為主
- 以學科與社群為導向的後設資料核心集
 - 以某一種標準為基礎進行分析與篩選
CDWA(書畫主題小組)、EAD(檔案主題小組)、
HISPID(植物主題小組)、MARC21(善本主題小組)、
動物主題小組(Species 2000)
 - 採用某一學科或社群為導向的既有核心元素集為依歸
動植物主題聯合小組(Darwin Core)
 - 直接採用某學科的既有核心元素集為計畫標準
臺灣視覺記憶數位典藏計畫(VRA Core Categories)



Suggestion for NDAP

- 應用領域 — eLearning, eCommerce, eBusiness
- 層面 — 時間與事件
- 功能 — 長久保存
- 產生方式 — 半自動，乃至於全自動
- 描述深度 — 資料內容(如: 著錄規範)、資料值(如: 索引典)
- 描述層次 — 採取廣(館藏合集)且深(全文標引)的雙向策略
- 資訊交換 — XML Schema
- 多種標準的整合與應用 — 資料轉換機制與服務



參考書目

- 陳雪華。(民86)。網路資源組織與Metadata之發展。圖書館學刊，12，19-37。
- 陳亞寧、陳淑君。(民90年)。以知識探索為本之知識組織方法論及研究分析。圖書與資訊學刊，39，頁36-51。
- Dempsey, L., & Heery, R. (1998). Metadata: A current view of practice and issues. Journal of Documentation, 54(2), 145-172.
- Gilliland-Swetland, A. (2000). Introduction to metadata: Setting the stage. Retrieved December 24, 2003, from http://www.getty.edu/research/conducting_research/standards/intro_metadata/pdf/swetland.pdf
- Heery, R., & Patel, M. (2000). Application profiles: Mixing and matching metadata shcemas. Ariadne, 25. Retrieved September 27, 2000, from <http://www.ariadne.ac.uk/issue25/app-profiels/>
- MetaMap. (2003). Why a MetaMap? Retrieved February 23, 2005, from <http://mapageweb.umontreal.ca/turner/meta/english/whatfor.html>