

建構數位典藏「主題索引典」：簡介與實務

2002年3月27日

一、索引典簡介

1. 定義

- *Guidelines for the Construction, Format, and Management of Monolingual Thesauri* (美國國家標準ANSI/NISO Z39.19-1993, URL:

<http://www.niso.org/standards/resources/Z39-19.html>)

‘A **controlled vocabulary** arranges in a known order in which equivalence, homographic, hierarchical, and associative relationships among terms are clearly displayed and identified by standardized relationship indicator, which must be employed reciprocally.’ (將控制詞彙利用已知的清楚地展現詞彙間同義、同形異義、層級和關聯的關係，而這些關係也需反指使用。)

‘Its purposes are to promote consistency in the indexing of documents, predominantly for postcoordinated information storage and retrieval systems, and to facilitate searching by linking entry terms with descriptors.’ (其目標是促進文件**標引**的一致、控制後組合資訊**儲存與檢索**系統、藉由款目詞與描述語的連結而便於**搜尋**。)

2. 索引典在數位典藏的功能與目的

- 功能：提供標引(indexing)與檢索(retrieval)時的控制詞彙以達到資料的檢準(precision)和檢全(recall)
- 目的：[\(ANSI/NISO Z39.19-1993\)](#)
 - (1) 轉換：將自然語言轉換成控制詞彙
 - (2) 一致性：促進標引詞的一致性
 - (3) 指示關係：指出詞彙間的語意關係
 - (4) 檢索：作為檢索文件時的搜尋輔助

3. 結構

(1) 索引典的結構須展現詞彙間的三種關係：

等同(equivalence)關係—表達術語和術語間同指相同概念的關係，其中一術語選為描述語，其他則為非描述語。等同關係具有以下幾類：

- 同義詞(synonyms)，例如：犬/狗、老鼠/耗子
- 詞彙變異(lexical variants)，例如：熊貓/貓熊
- 類同義詞(quasi-synonyms)，例如：冷/熱(一連續面上的不同點)
- 複合詞內的互指(cross-references to elements of compound terms)，例如：coal mining/coal + mining

層級(hierarchical)關係—表達概念與概念間廣義與狹義的關係(broader terms vs. narrower terms)或整部關係(whole-part relationships)

- 屬種關係(generic relationship)，例如：鳥類/麻雀

- 整部關係，例如：消化系統/胃
- 類例關係(the instance relationship)，例如：中央山脈/玉山
- 多層級關係(polyhierarchical relationships)，例如：生物化學/生物/化學

關聯(associative)關係—描述語間概念相關聯(如同位語(sibling terms)或分屬不同層級)，但意義上不是等同或具有層級上下的關係

- 同一層級(含義部分重疊)，例如：機車/腳踏車
- 不同層級間但彼此的關係是「物件/研究物件的學科」，例如：植物/植物學、「過程/成品」，例如：編織/衣服、「工具/使用工具的過程」，例如：畫筆/繪畫等。(Thesaurus principles and practice, URL: <http://www.willpower.demon.co.uk/thesprin.htm>)

(2) 索引典基本內容及其縮寫符號

名稱	英文縮寫
描述語(Descriptor, Preferred Terms)	USE
非描述語(Non-preferred Terms)	UF (Used For)
範圍註(Scope Note)	SN
上位詞((Broader Terms)	BT
下位詞(Narrower Terms)	NT
關聯詞(Related Terms)	RT

範例：ERIC 索引典

CONTENT AREA READING

SN Indications concerned with reading assignments – also, instructional material in such subject areas as social studies, mathematics, science, and English

UF Content Reading (1967, 1980)
Readign in Content Areas

BT Reading
Reading Instruction

RT Critical Reading
Directed Reading Activity

4. 索引典編製的基本步驟

根據《索引典理論與實務》(1994)一書的說明，索引典編製的基本步驟可分為五項：

- (1) 確定索引典的主題範圍及決定特性
- (2) 建立知識概念的基本架構
 - [ANSI/NISO Z39.19-1993](#)建議的概念類別分為七類：
 - a) 人事物和其物理部分(例：鳥類、山、油畫、泰迪熊)

- b) 材質(例：黏著劑、芥氣、油漆、水)
 - c) 活動或過程(例：酸化、繪畫、潛水、縫紉)
 - d) 事件或發生的事情(例：生日、內戰、假期、革命)
 - e) 人、事、材質或活動的屬性或狀態(例：意識、人格、速度、紋理)
 - f) 學門或主題領域(例：人類學、資訊科學、有機化學、神學)
 - g) 測量單位(例：赫茲、公里)
- (3) 選定詞彙
- (4) 建立這些詞彙概念間的關係及完成各描述語款目與參照款目
- 同一概念結構下，概念涵蓋較廣(general)作為上位詞，越專門(specific)的概念作為下位詞
 - 從首位詞(top terms)到下位詞的層級級數以五級為宜
- (5) 測試與修正

三、藝術類主題索引典發展現況：以藝術與建築索引典(Art & Architecture Thesaurus, AAT)為例

1. 背景簡介

(1) AAT 範圍

- 術語內容：藝術、建築、裝飾藝術、物質文化、檔案資料
- 術語總數：包含 26000 個概念，約 125000 個詞彙
- 收詞時空範圍：古代到現代，遍及全球(但較重西方概念)

(2) AAT 功能

1. 作為資料輸入的權威詞彙或控制詞彙
2. 作為檢索輔助

2. AAT的結構與內容

(1) 內容

- 內容呈現遵循索引典國際標準
- AAT 的一筆記錄(record)表現一種「概念」(concept)
 - 「概念」的定義：包含抽象的概念、現象以及實際的材質和工藝屬性，可分成五類：
 - a) 風格、時期、文化
 - b) 描述個人和組織的術語
 - c) 描述活動和職業的術語
 - d) 描述原材料或應用材質過程的術語
 - e) 描述物件的術語，例如藝術作品、裝飾藝術、建築和相關物件

(2) 結構

- 著錄端的欄位(REC 的格式)

標籤	欄位名稱
LEN	記錄總字數(length of record)
STATUS	異動記錄(record status)
IDNO	唯一識別號(unique identification)
DATCHG	最後更動日期(date changed)
DATENT	建檔日期(date entered)
CN	歸類記錄(classification notation)
ST	特定索引[碼代號](special thesaurus code)
TERM	描述語(preferred term, descriptor)
ALT	其他描述語(alternate descriptor)
UK	英式描述語(British equivalent)
UKALT	英式其他描述語(British alternate)
UKUF	英式非描述語(British "use for")
UF	非描述語(variant term ("use for" term))
BT	上位詞(broader term)
RT	關聯詞詞(related term)
SN	定義(scope note)
IN	標引說明(indexing note)
HN	更動說明(history note)
SOURCE	來源(sources)
LINK	LCSH 對等詞彙(LCSH term)

■ End-User 端的欄位

欄位名稱	定義	詞間關係
Record ID (識別號)	唯一識別號	
Descriptor (描述語)	從眾多表達相同概念的詞彙裡，選擇權威資源裡最常為人所用，又稱 preferred terms	
Note (說明)	描述概念，例如術語的意義、範圍或用法	
Hierarchy (層級)	以結構化方式展現概念的層級關係	層級關係
Terms (術語)	與描述語概念相同的術語，可能是拼法上的差異、同義詞、一般用法和地方用法上的不同等	等同關係
Related concepts (相關概念)	語意上相關概念的詞彙	關聯關係
Sources (來源)	參考書目	

(3) 表現方式

AAT 是以鋸齒狀的方式表示階級結構，同層次的概念(同位語)則依字母排列或依時間先後排列。(當沒有術語來銜接較廣義的概念與較狹義的概念時，則以「連結語(guide term, node labels)」表達，表示方式為「<...>」，連結語是

不被索引，也不是「術語」)

例如：

...containers
<containers by form>
bags
vessels
alembics
askoi
basins
louteria

3. AAT如何呈現三種詞間關係

- 在 end-user 端的「術語 terms」欄位列出所有的「非描述語」，以表現對等關係(Equivalence Relationship)
- 在 end-user 端的「層級 Hierarchy」欄位列出此概念的上位語與下位語，表現層級關係(Hierarchical Relationship)。
 - AAT 層級關係的類別大多以「屬種關係」為主，少部分為「整部關係」
 - 多層級關係在 AAT 裡是呈現於「相關概念」欄位裡
 - 七種類別—AAT 把所有的概念分成七種面向(facets)，七種類別之下再區分有 33 種層次(hierarchies)：
 - 相關概念面向(Associated concepts facet)—包含有關研究和實行人類思想和活動的抽象概念和現象，包括各種建築和藝術以及相關學門；也包括重要的理論與批評、意識形態、態度、社會和文化運動等。例如：美、平衡、鑑賞、隱喻、自由、社會主義。
 層次：相關概念
 - 物理屬性面向(Physical attributes facet)—是有關材質和工藝品的觸覺和測量特徵、以及材質和工藝品的特色，例如大小、形狀、材質的化學特色、織品質料和硬度、表面紋飾以及顏色等。例如：帶狀飾、滾邊、圓形、浸水的、易碎品。
 層次：屬性和性質、條件和效果、設計元素、顏色
 - 風格與時期面向(Styles and periods facet)—提供和藝術、建築以及裝飾藝術相關的風格團體以及特定時代廣為人所使用的術語。例如：法國、路易十四、Xia、Black-figure、抽象表現主義。
 層次：風格與時期
 - 人物面向(Agents facet)—包含人物、團體名稱以及藉由職業、活動、物理或精神特徵、或是社會角色或條件所識別的組織名稱。例如：製版者、景觀設計者、法人、religious orders
 層次：人物、組織
 - 活動面向(Activities facet)—包含身體與精神的活動、個別的事件、一系列系統性的活動、達成結果的方法以及發生在材質和物件上的過程。活動範圍可以從學習或職業領域到特殊生活事件、從心智上

的構思到實施在材質或物件上的過程、或是從單一物理上的活動到複雜的競賽。例如：考古學、工程學、分析、比賽、展覽、跑步、繪圖(影像製作)、腐蝕。

層次：學門、功能、事件、物質活動、過程與技術

- 材質面向(Materials facet)—處理物質，不論自然或合成。其範圍從特定材質到為功能而設計的材質類型，例如著色劑、或從生材質到用來建造建築或物件的材料。例如：鐵、黏土、黏著劑、感光乳劑、人造象牙、水車機械。

層次：材質

- 物件面向(Objects facet)—在 AAT 裡是最大的面向，包含以人力完成、非生物性、可感覺到或看到的事物。就物質形式來說，從建築作品到影像和文字檔案；就目的來說，從實用主義到唯美主義。也包含提供建築環境內涵的景觀。

層次：物件群組和系統、物件類型、構件

建築環境

聚落和景物、建築物 and 地區、單一建築作品、開放空間和地點元素

家具和設備

家具、服裝、工具和設備、武器和軍火、測量裝置、容器、聲音裝置、娛樂工具、交通工具

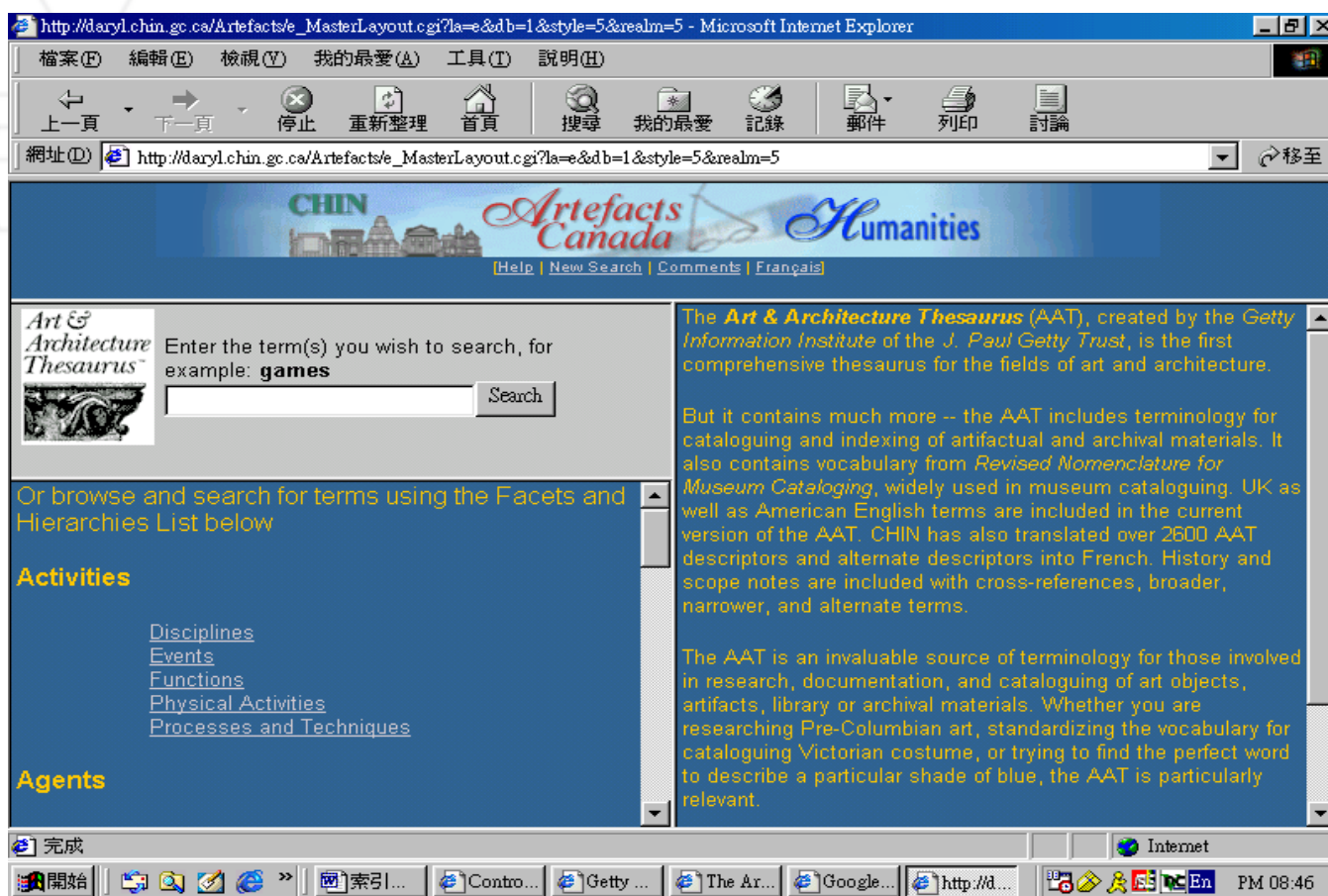
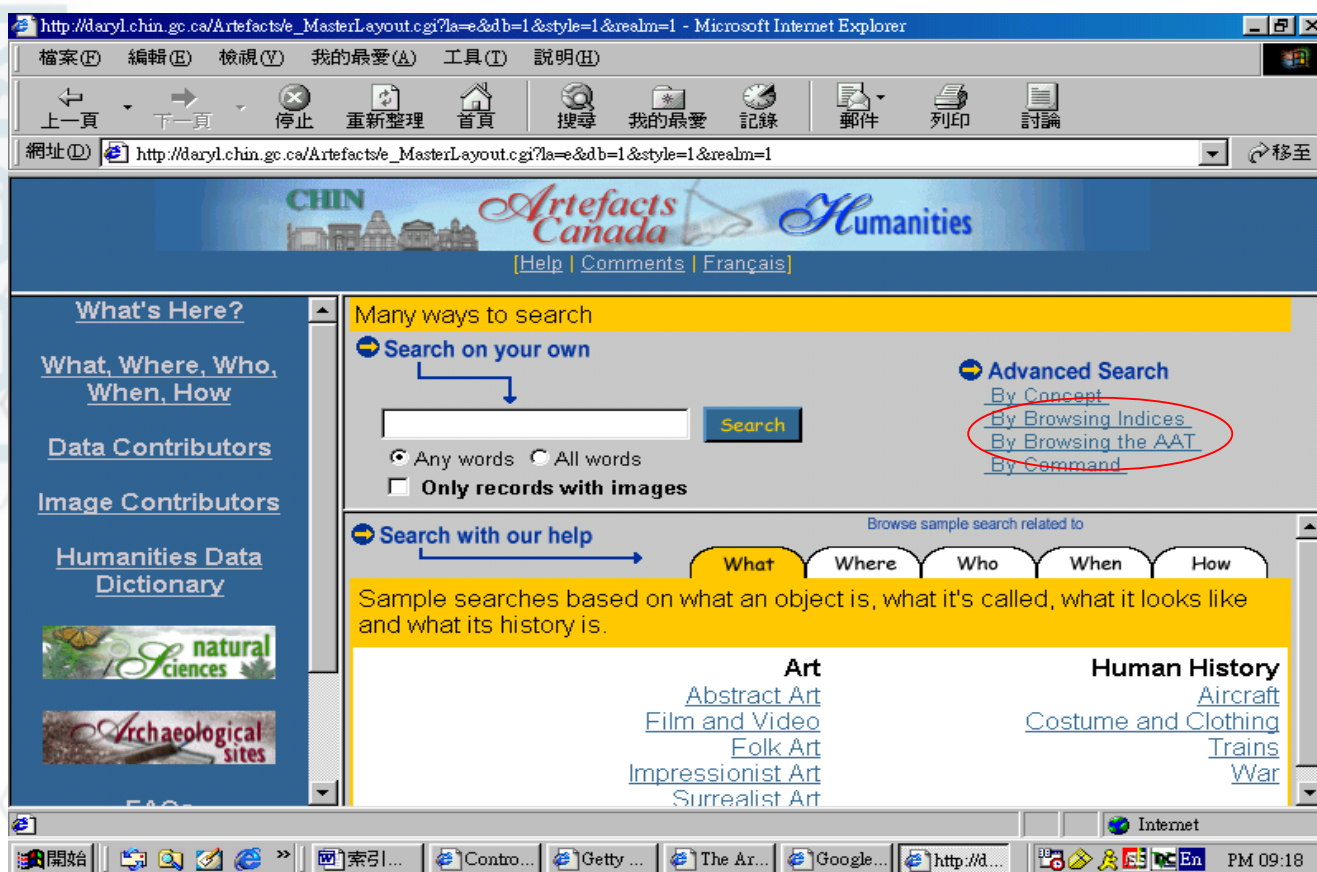
視覺和口語溝通

視覺作品、交換媒介

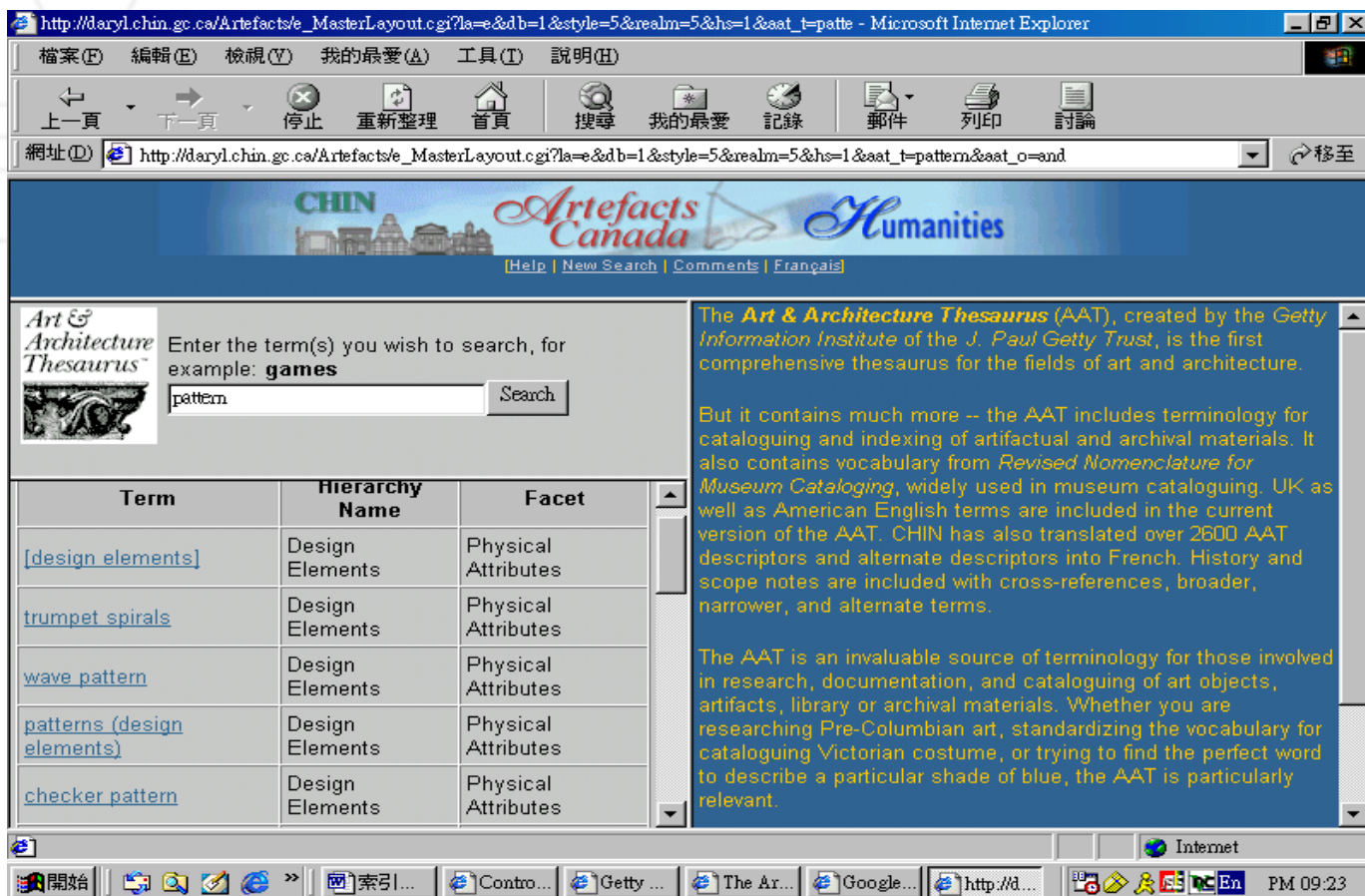
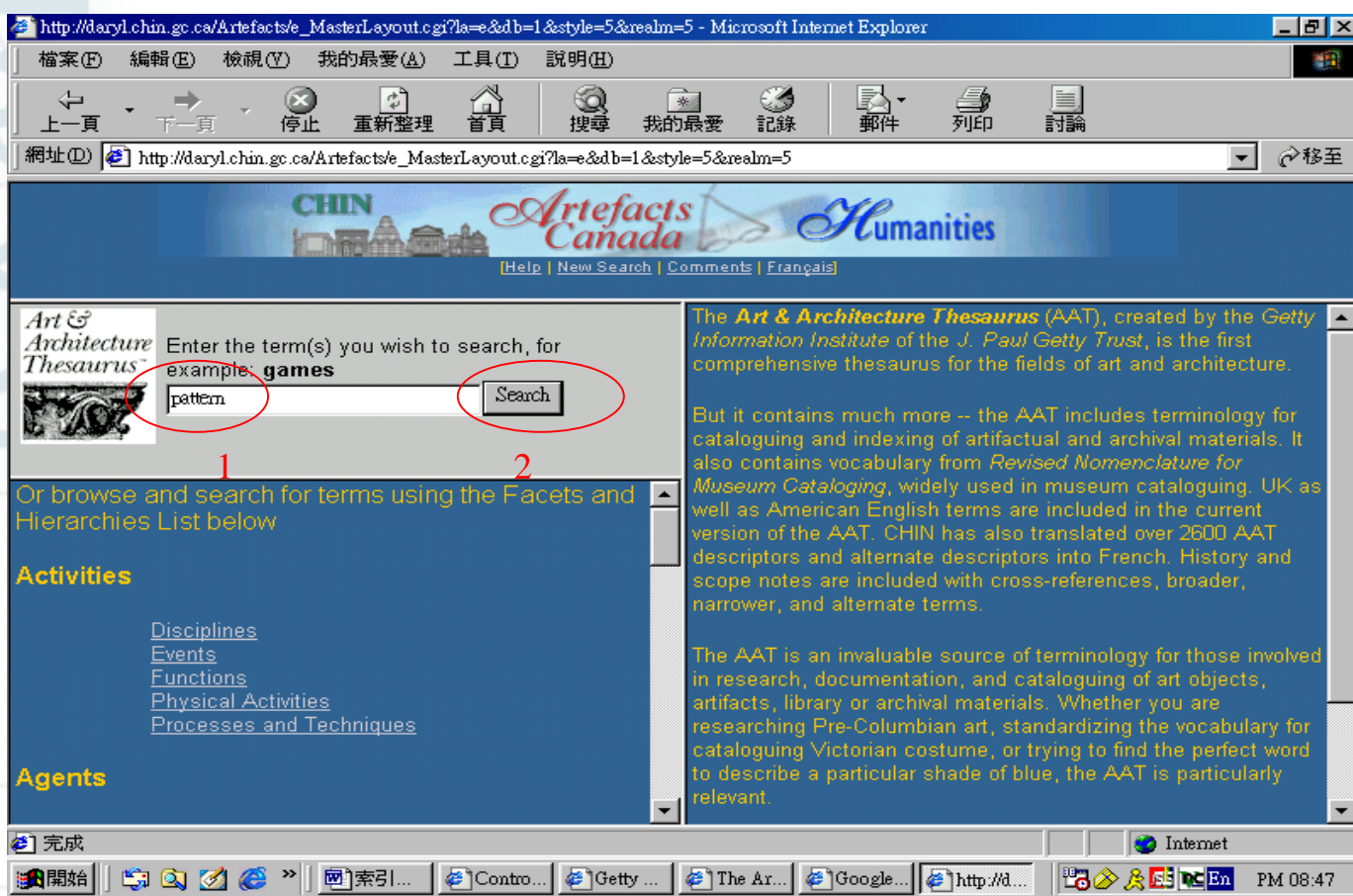
- **關聯關係(Associative Relationship)**—用來連結相關聯的概念，呈現在 end-user 端為「相關概念」(Related concepts)欄位

四、AAT 在數位典藏中的應用

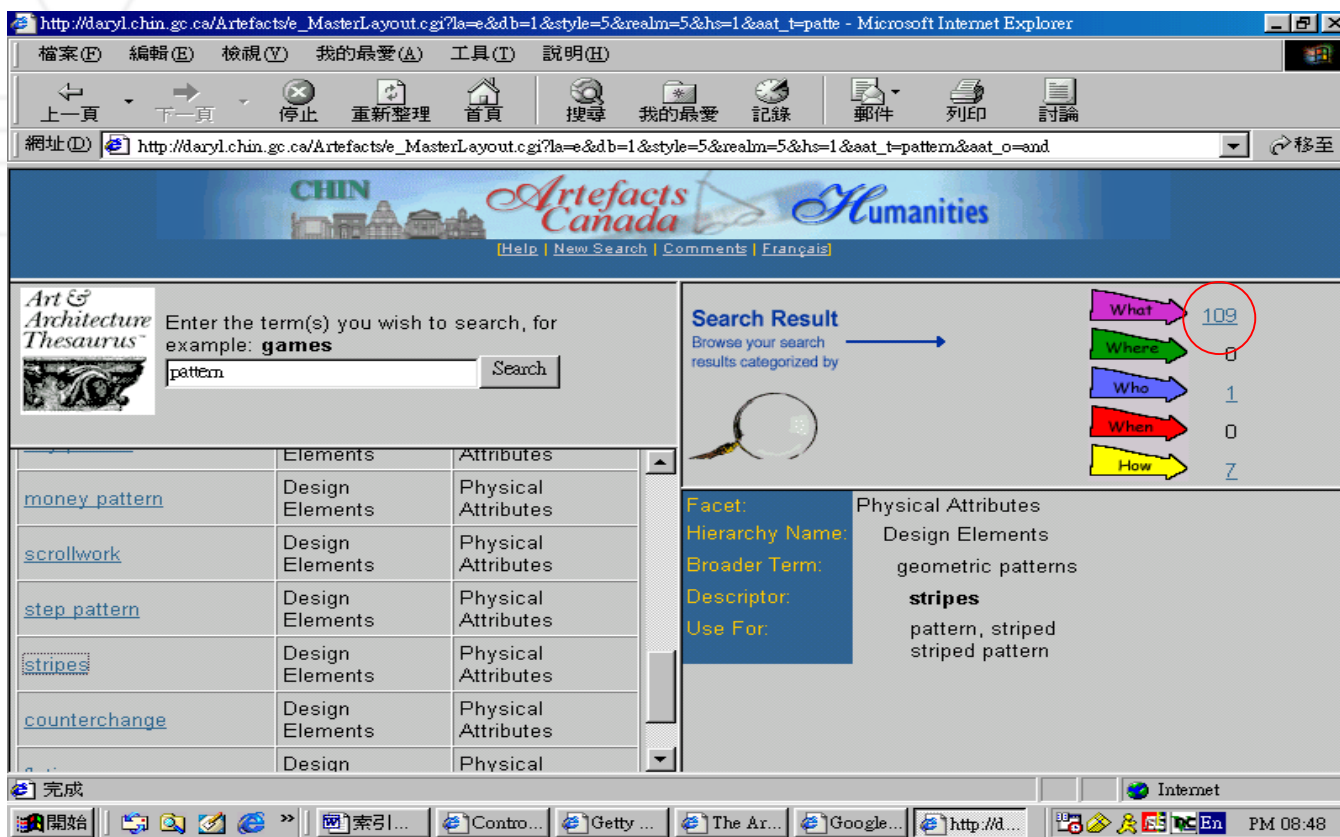
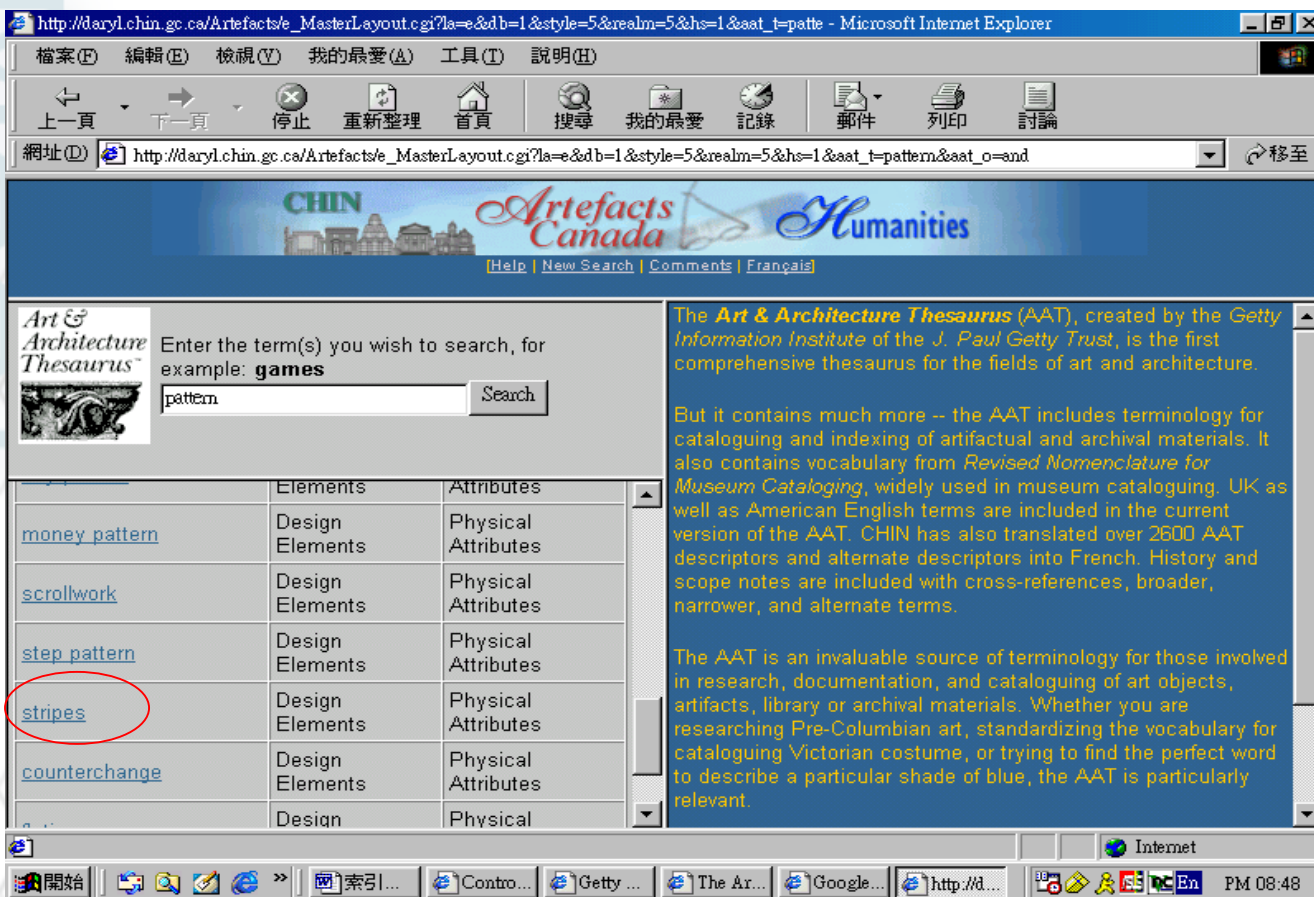
1. 從搜尋面應用AAT的範例：Canadian Heritage Information Network (CHIN, http://daryl.chin.gc.ca/Artefacts/e_MasterLayout.cgi)

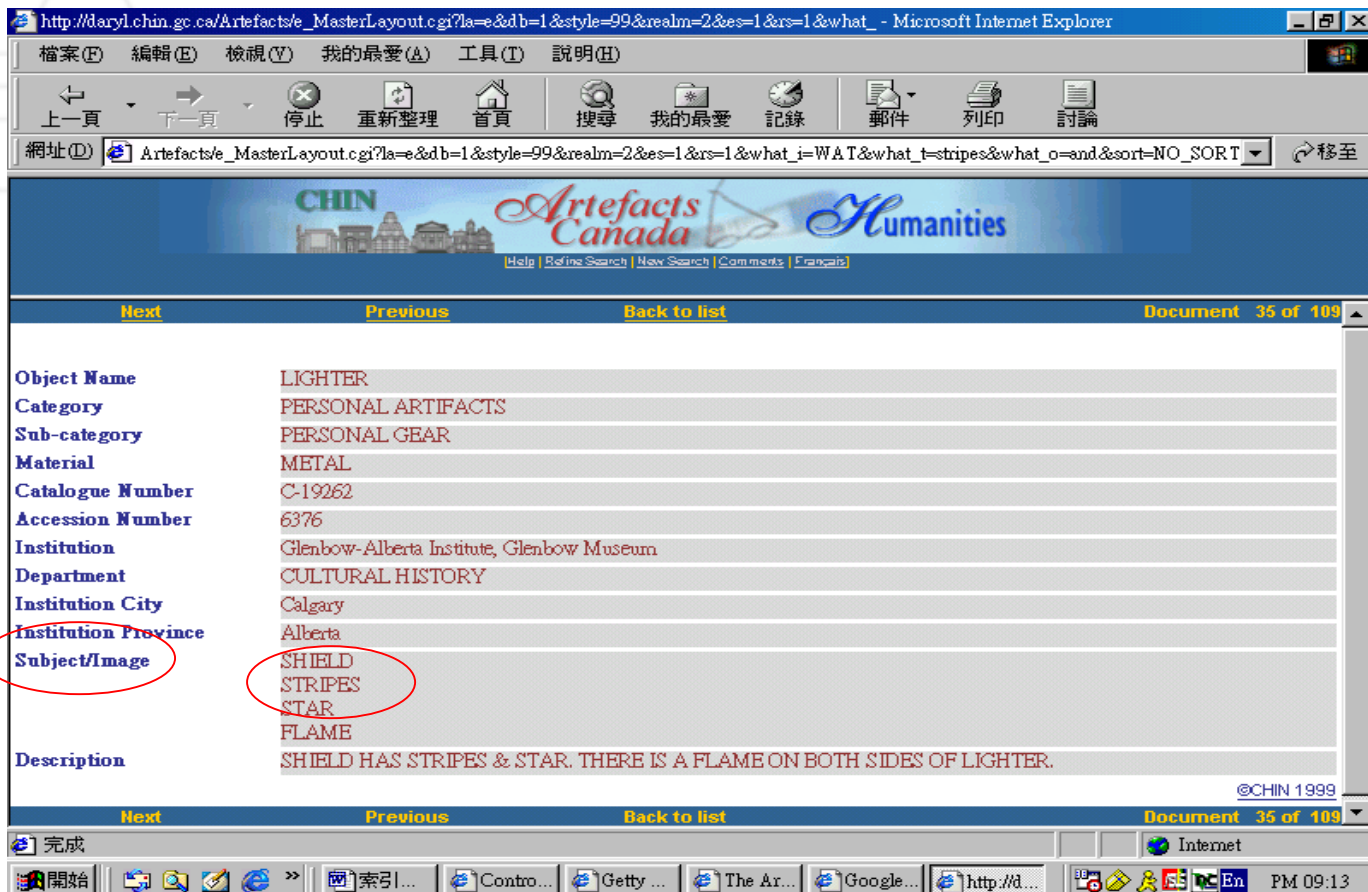
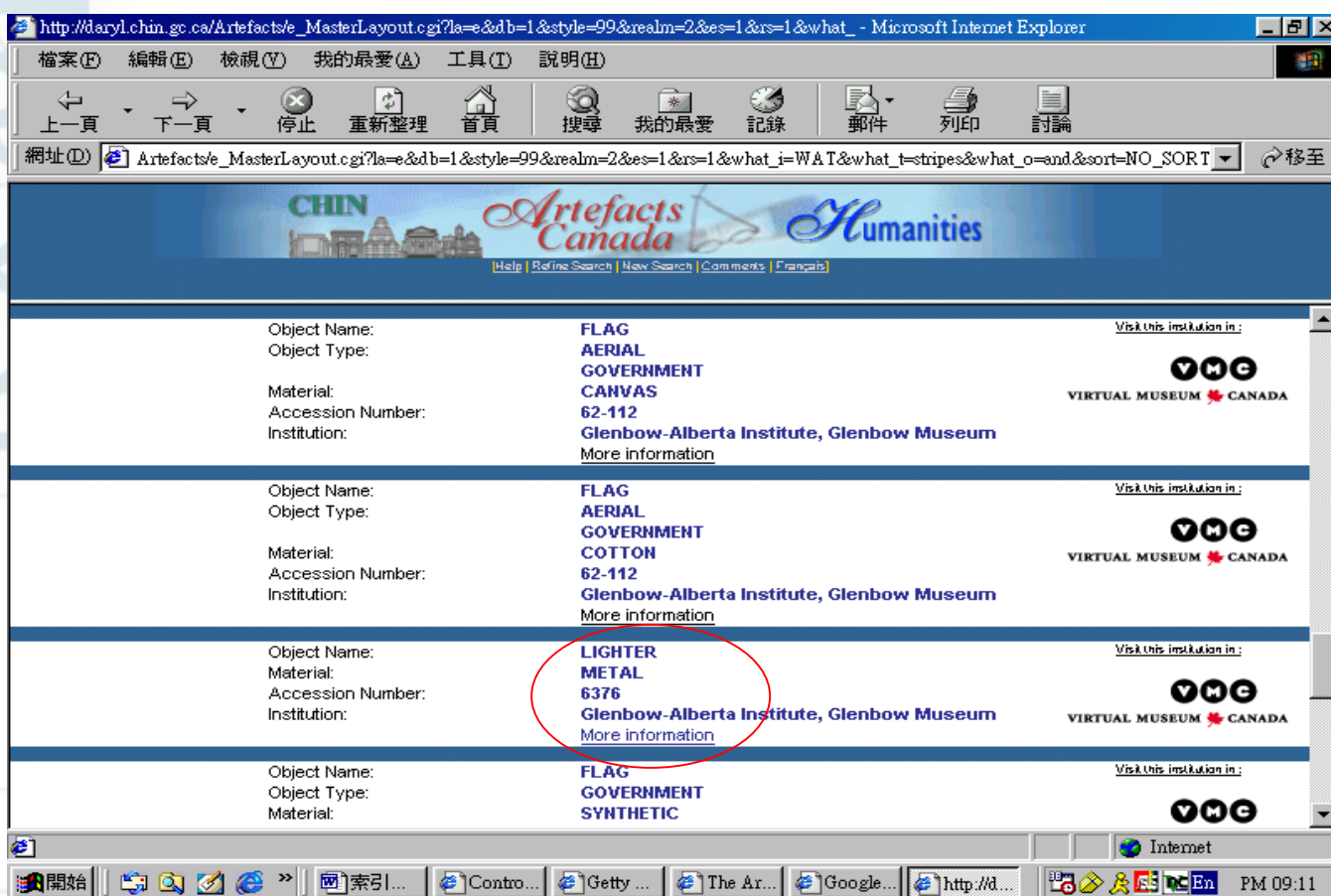


Metadata Architecture and Application Team

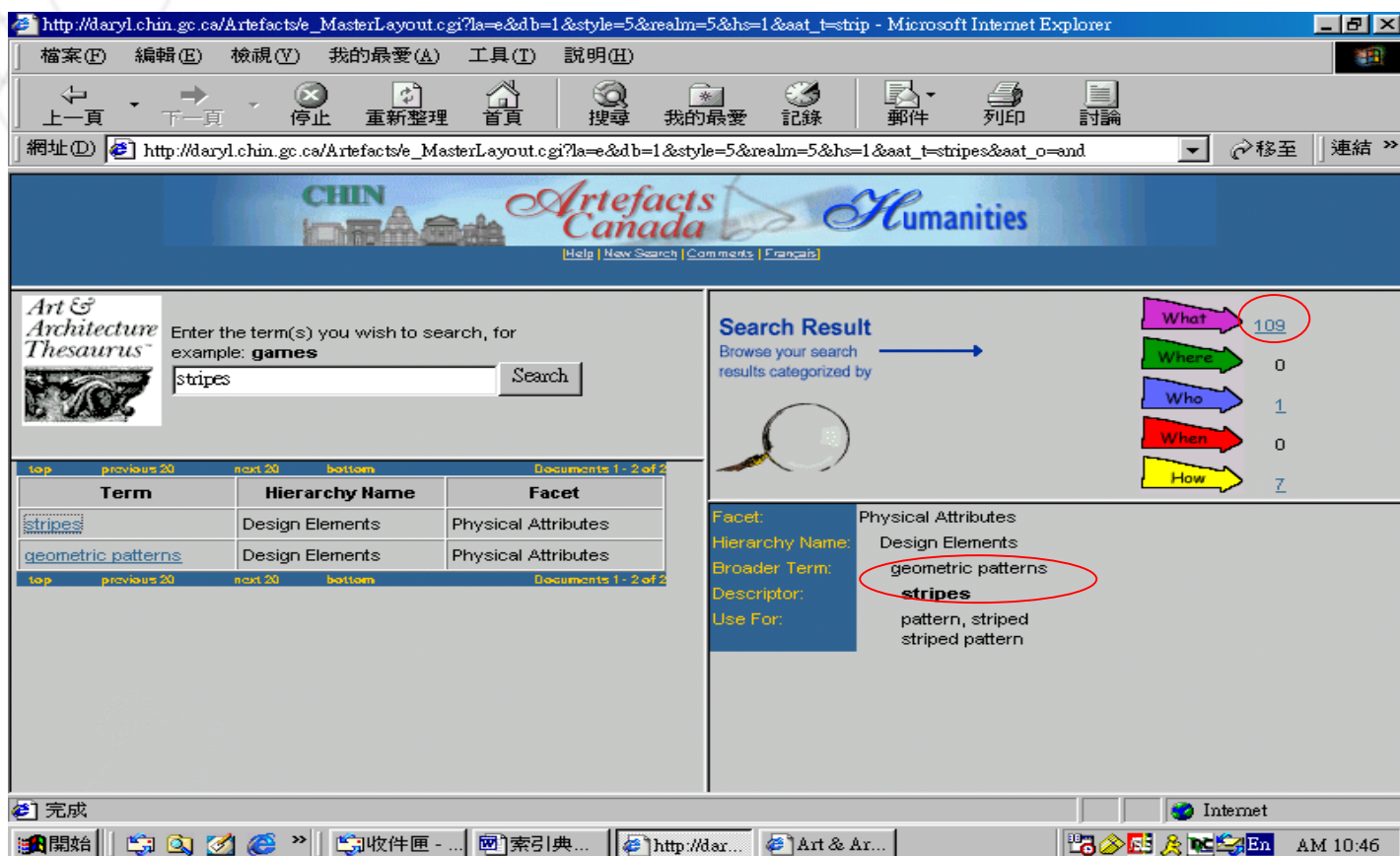
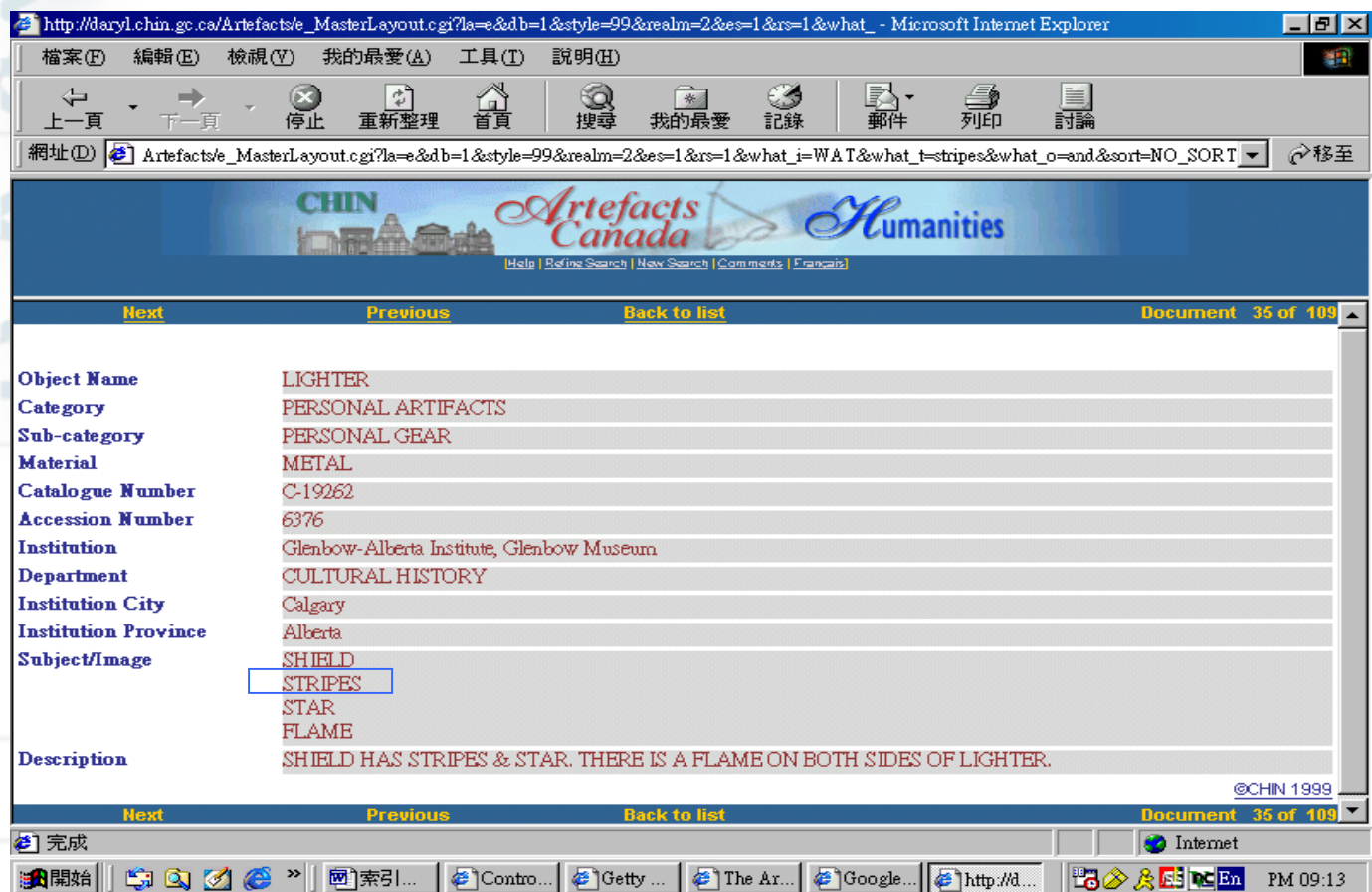


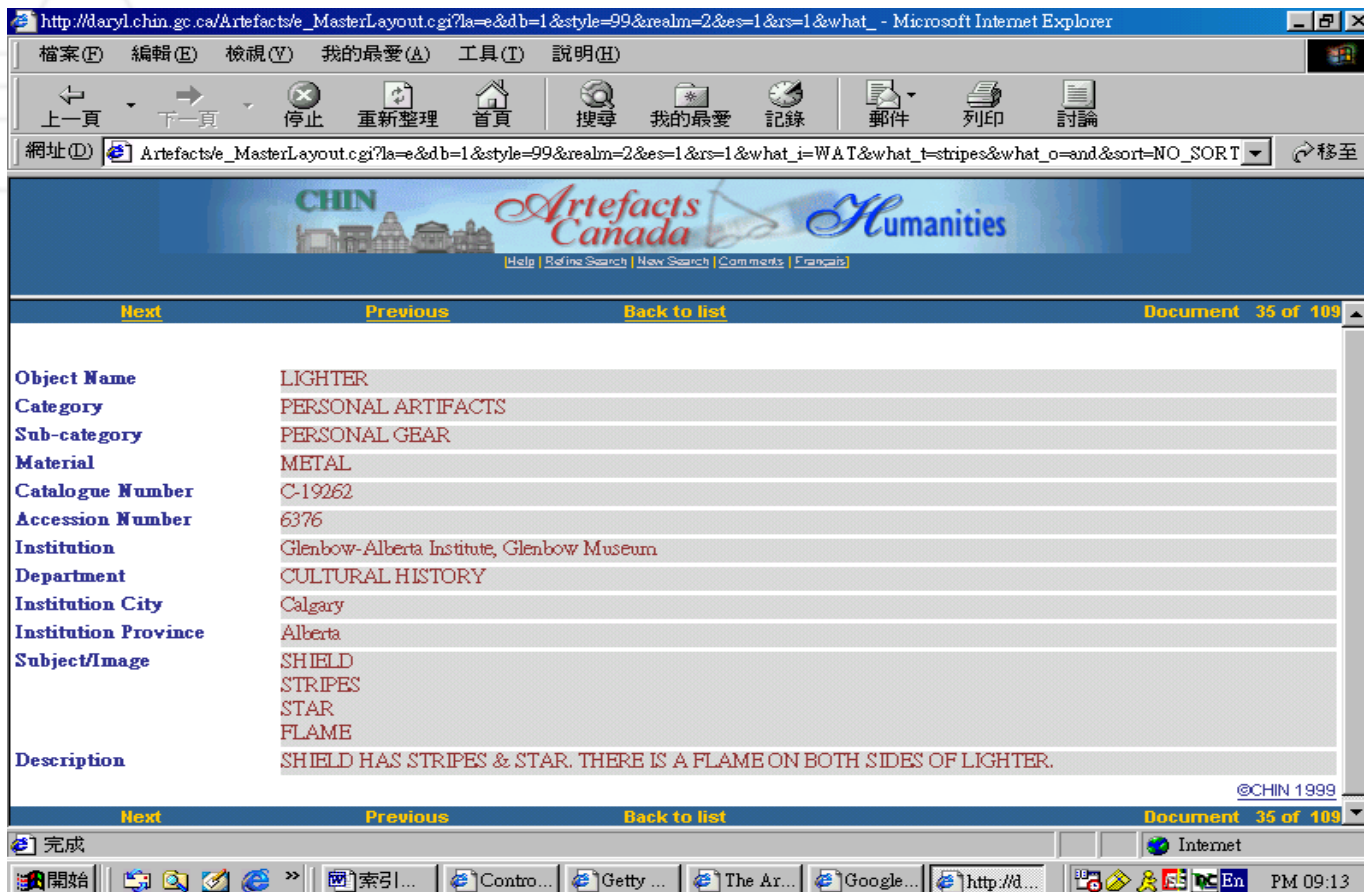
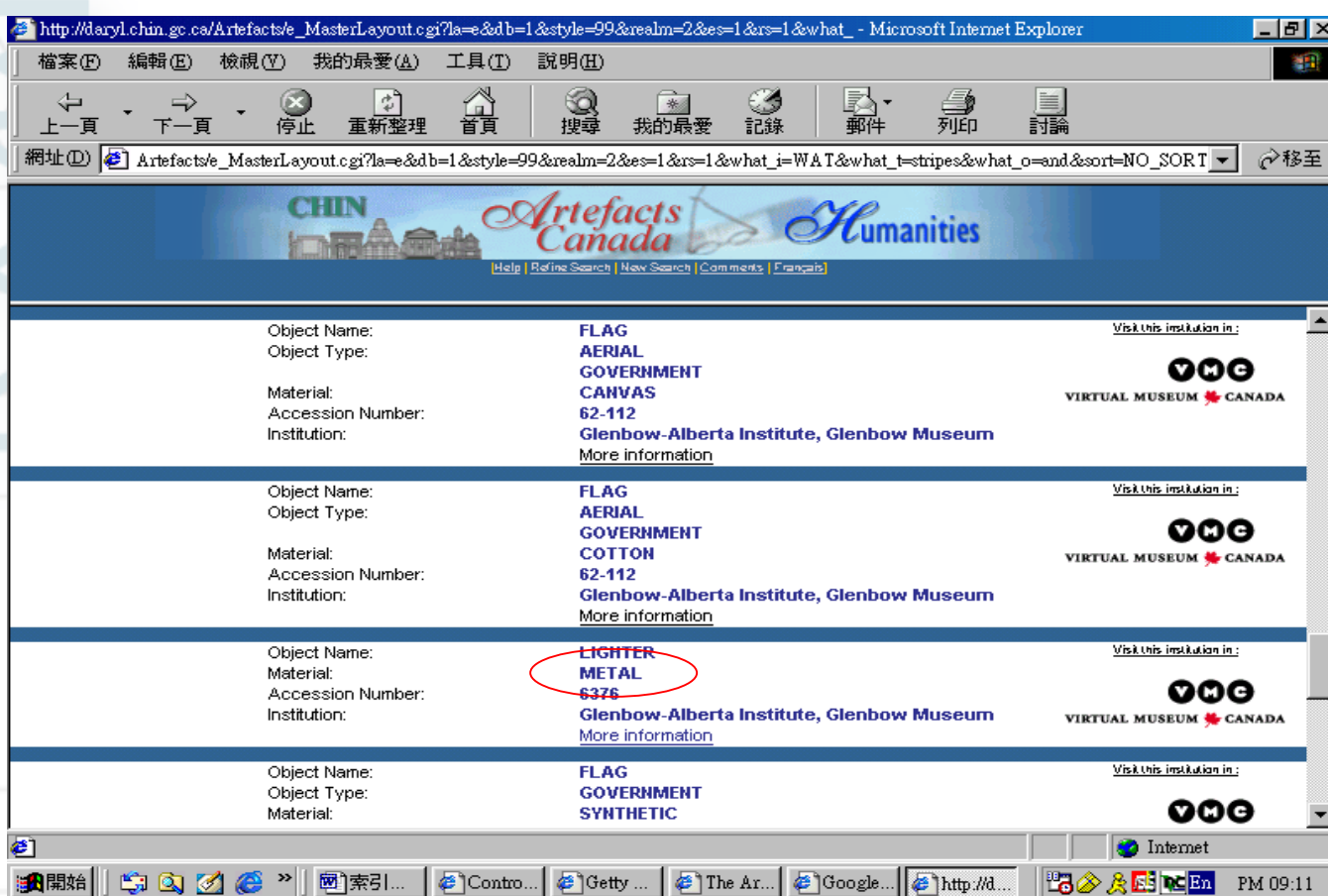
Metadata Architecture and Application Team





2. 從 Metadata 記錄連結 AAT 範例(藉由 CHIN 查詢介面虛擬主題查詢結果)





五、故宮器物紋飾權威檔

1. 故宮器物主題/紋飾代碼分析

- 紋飾描述欄位分有兩層級，第一層代碼為較廣義的詞彙，分有 45 類；第二層代碼具有 4 類到 30 多類不等，為較狹義的詞彙
- 第一層和第二層代碼的關係分析

屬種關係	獸面紋、龍紋、動物紋、昆蟲、水族、鳥類、花卉、植物、草葉紋、瓜果、人物、文字、吉祥文字、錦地紋、幾何紋、開光、雲紋、雷紋
整部關係	道釋仙者、道家雜寶、黻黼十二章、八寶、七珍、暗八仙、山水風景、建築舟車
類例關係	十二生肖、八仙、吉祥圖案

2. AAT紋飾(patterns)的層級結構

- 紋飾(patterns)在 AAT 的結構裡是歸類在「Design Elements」層次下；「Design Elements」則屬於「物質屬性層面(PHYSICAL ATTRIBUTES FACET)」

<Design Elements>

<design elements>

applied decoration

<design elements by function>

motifs

openwork

<ornament areas>

patterns (design elements)

- 紋飾下分有三級

patterns

...patterns by general type

.....allover patterns

.....diaper patterns

.....counterchange

.....running ornament

...patterns by specific type

.....arabesques

.....cloud bands

.....filigree

.....floral patterns

.....

3. 故宮紋飾代碼的問題

- (1) 就概念類別上，僅有事物細類的劃分(紋飾)，沒有整體概念上分類
- (2) 就詞間關係上，紋飾位置沒有層級劃分；紋飾描述的第一層代碼和第二層代碼沒有顯示出上位詞與下位詞的區別，因此造成有些屬種關係在同一

層級，例如：植物(屬)與花卉(種)；類例關係在同一層級，例如：釋道仙者(類)與八仙(例)

4. 建議

- (1) 故宮器物
 - A. 就紋飾主題作概念分類的再整理，如：植物、草葉紋、花卉、瓜果可 group 為植物(建議可參考美國國家標準七種概念類別)
 - B. 整理紋飾主題之上下層級關係
 - C. 歸納紋飾主題之相關關連關係
 - D. 羅列紋飾主題別名清單(非描述語清單)
- (2) 後設資料工作組
 - A. 實際以故宮器物紋飾主題進行索引典之應用分析
 - B. 提供主題索引典參考元素清單

六、索引典參考資源

1. 建立索引典的國際、國內標準

- 國際標準
Documentation -- Guidelines for the establishment and development of monolingual thesauri (ISO 2788: 1986, URL: <http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=7776&ICS1=1&ICS2=140&ICS3=20>, 付費取得)
- 美國國家標準
Guidelines for the Construction, Format, and Management of Monolingual Thesauri (ANSI/NISO Z39.19-1993, URL: <http://www.niso.org/standards/resources/Z39-19.html>, pdf檔)
- 英國國家標準
Guide to establishment and development of monolingual thesauri (BS 5723: 1987, <http://bsonline.techindex.co.uk/>, 付費取得)
- 中國國家標準(CNS)
西文單一語文索引典編製標準 (CNS 13224: 1993, Z70246, URL: <http://www.cnsppa.com.tw/>, 付費取得)

2. 博物館相關主題線上索引典

- 英國 MDA 機構博物館術語的資源
 - MDA Archaeological Objects Thesaurus (<http://www.mda.org.uk/archobj/archcon.htm>)
 - British Museum Object Names Thesaurus (<http://www.mda.org.uk/bmobj/Objintro.htm>)
 - British Museum Materials Thesaurus (<http://www.mda.org.uk/bmmat/matintro.htm>)
 - Vocabulary of Basic Terms for Cataloguing Costume (<http://www.mda.org.uk/costume/vbt00e.htm>)

- 美國國會圖書館構圖材料索引|典 I, II
 - Thesaurus for Graphic Material I: Subject Terms (TGM I)
(<http://lcweb.loc.gov/rr/print/tgm1/toc.html>)
 - Thesaurus for Graphic Material I: Genre & Physical Characteristic Terms (TGM II)
(<http://lcweb.loc.gov/rr/print/tgm2/>)
- 澳洲繪畫索引典(由澳洲洲立圖書館和澳洲線上博物館合力出資發展)
 - Australian Pictorial Thesaurus (APT)
(<http://www.picturethesaurus.gov.au/index.html>)
- 圖像類目系統索引典(由 Leiden 大學教授 Henri van de Waal 所發展)
 - The ICONCLASS Home Page
(<http://www.iconclass.nl/>)

3. 彙整索引典資源的網站

- Controlled Vocabularies Resource Guide
(http://sky.fit.qut.edu.au/~middletm//cont_voc.html#Databases)
- Publications on thesaurus construction and use
(<http://www.willpower.demon.co.uk/thesbibl.htm>)

4. 參考文獻

- *索引典理論與實務*
美國資訊科學學會台北分會、農業科學資訊服務中心、國立中央圖書館、中國圖書館學會
台北：美國資訊科學學會台北分會，1994
- “Guidelines for Constructing a Museum Object Name Thesaurus”
MDA, 1997-1998, URL: <http://www.mda.org.uk/holm.htm>
- “Thesaurus principles and practice”
Willpower Information, 1998, URL:
<http://www.willpower.demon.co.uk/thesprin.htm>
- *User's Guide to the AAT Data Releases* (Release Version 4.0)
Patricia Harpring, (ed.)
The J. Paul Getty Trust, 2001, URL:
<http://www.getty.edu/research/tools/vocabulary>