

使用可擴展標示語言製作台灣社會地圖（初步報告）

張藝鴻，莊庭瑞

中央研究院 資訊科學研究所

台北市 南港區 115 研究院路二段 128 號

公元 2001 年 8 月 31 日

摘要

本文件描述「台灣社會地圖」（Taiwan Social Map）系統建置之經驗。「台灣社會地圖」為一套使用「可擴展標示語言」（XML）來整理、標示、處理、以及呈現有關台灣的社經統計資料，並採用數位地圖的介面，讓使用者以互動與視覺化的方式探索社經資料和其地域關連之系統。此計畫之主要目的包括：一、檢驗 XML 技術規範以及其搭配的軟體工具，二、實際使用 1990 年的全國戶宅普查資料（近兩千萬筆資料），搭配詳細的行政區域劃分圖（七千餘村里）來進行系統製作，最終目標在於發展並累積以 XML 為底的「複雜資料集」（complex datasets）的處理技術。這個系統並藉由將統計資料透過 XML 標示與視覺化呈現等工作，使台灣社經統計資料得被方便地交換、再利用與探索。本文件中描述了建置「台灣社會地圖」之過程、遭遇的問題與我們的解決方案，並且對於有關各種資料集取得與交換之議題，進行了簡短的討論。

An XML-based Taiwan Social Map

(Preliminary Report)

Yi-Hong Chang, Tyng-Ruey Chuang

Institute of Information Science

Academia Sinica

Nankang, Taipei 115

Taiwan

August 31, 2001

Abstract

We report our experience in building Taiwan Social Map: An XML-based Web system for interactive exploration of Taiwan population datasets. Our motivation in building such a system is to evaluate the feasibility of using XML technologies, with currently available free software tools and systems, to prepare, markup, process, and visualize socioeconomic datasets about Taiwan population. The goal is to cultivate our technological expertise in processing complex datasets, and to build a system that is actually useful to social science researchers. The Taiwan Social Map system currently combines the 1990 national census dataset (about 20,000,000 survey records) with a detailed administration boundary map (with more than 7,000 contours of villages and precincts). It allows interactive visualization of summary statistics of the census data on the administration map, and is very easy to use. We report here the problems we encountered in building the system and the solutions we adapted. We also briefly discuss issues related to datasets acquisition and exchange.

目次：

0. [如何取得此份文件與軟體](#)
 1. [前言與計畫目標](#)
 2. [系統概述](#)
 3. [原始資料集的整理](#)
 - 3.1 [1990 年主計處戶宅普查](#)
 - 3.2 [地圖資料](#)
 4. [現行系統的建製](#)
 - 4.1 [原始資料集的 XML 產生](#)
 - 4.2 [視覺化顯示子系統](#)
 - 4.3 [原始資料集與視覺化顯示子系統的整合](#)
 5. [目前的結論](#)
 - 5.1 [原始資料的取得問題](#)
 - 5.2 [XML 與其他軟體工具的限制](#)
- [致謝](#)
- [註解](#)
- 附錄
1. [censusdata.dtd](#)
 2. [displaydata.dtd](#)
 3. [本系統之架構與原始程式說明](#)

如何取得此份文件與軟體

本文件版本：http://quad.iis.sinica.edu.tw/~tsm/doc/2001/doc_tsm_20010831.html

文件中所描述之「台灣社會地圖」系統，可在 <http://quad.iis.sinica.edu.tw/~tsm> 線上測試：

注意：

1. 目前在上述網址中的台灣社會地圖之最新版本為 0.10 beta，此版本僅在 Microsoft Windows 平台上的 Internet Explorer (版本 5.0 以上) 測試，已知在任一版本的 Netscape 上均無法運作。我們將會設法使本系統能跨越各種作業系統與瀏覽器平台。
2. 本系統需有 SVG (見下述) 的瀏覽器外掛軟體 (plug-in) 方能執行，目前我們僅在 [Adobe SVGViewer](#) 中測試，其他的 SVG 軟體均未測試。同樣的，我們將會設法使本系統能跨越各種環境。

本系統之原始程式碼與資料庫結構定義，可在 <http://quad.iis.sinica.edu.tw/~tsm/sourcecode/> 取得，有關各程式之作用，請參見 [附錄三](#)。

注意：本系統尚屬實驗性質，故各原始程式碼中可能含有錯誤。

對本文件與軟體有任何的意見，請寫信到 trc@iis.sinica.edu.tw 與我們一起討論。

版權：本文件與程式均為自由軟體，歡迎重複使用。惟引用本文與程式，請著名出處。但，1990 年戶宅普查的權利屬行政院主計處所有，如欲使用本系統之資料，請參見該單位之相關規定。

1. 前言與計畫目標：

「台灣社會地圖」(TaiwanSocial Map) 的建置動機在實際使用「可擴展標示語言」(XML, Extensible Markup language) [1]來整理、標示、處理、以及呈現有關台灣的社經統計資料，並採用數位地圖的介面，讓使用者以互動與視覺化的方式探索(explore) 社經資料和其地域關連。這個計畫有兩個目的：第一是檢驗 XML 技術規範以及其搭配的軟體工具(如瀏覽器、XML 文件處理器、伺服軟體系統等) 是否已經成熟，足以建立一套以「全球資訊網」(WWW, World-Wide Web) 為執行環境的「自由軟體」(free software)，以數位地圖方式，來處理並呈現社經資料。其次是實際使用 1990 年的全國戶宅普查資料(近兩千萬筆原始問卷資料)，搭配詳細的行政區域劃分圖(共有七千多個村里的邊界圖)，探討在系統製作上的可行性、相關議題、以及解決方法。我們的最終目標在於發展並累積以 XML 為底的「複雜資料集」(complex datasets) 的處理技術，並期望我們所實際發展出來的系統——台灣社會地圖——有助於社會科學的研究。

本計畫從 2001 年 4 月 1 日至今(2001 年 8 月 31 日)，已進行五個月，並有部分成果。成果雖然並不完備，但多有超乎預期之處。這份初步報告記載了我們的作法與經驗，以供大眾參考。

採用 XML 以製作台灣社會地圖，至少包含了以下三項優點：

一、XML 的可擴充性以及格式公開性可使各項社經資料得被使用者簡便的交換與再利用。

二、統計資料的視覺化得以使各種社會現象直觀地被使用者所理解，並且提供再研究的可能。

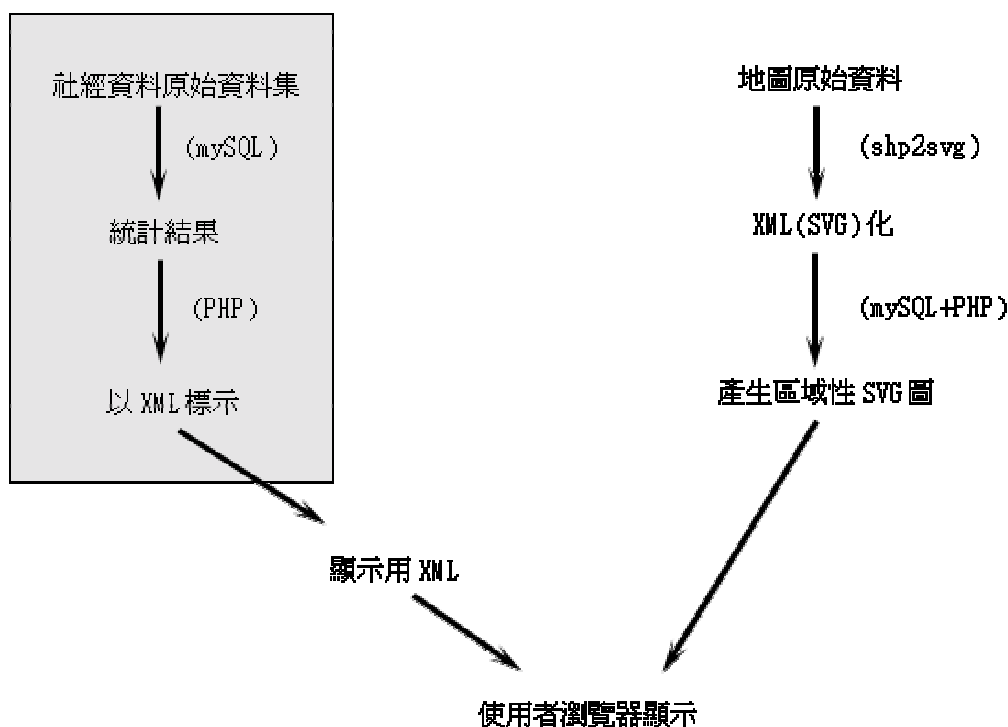
三、經由 XML 為媒介，有可能整合各種社經統計資料，使易於交叉比較與呈現。

這三個優點也是我們建置本系統的目標，在目前的實驗架構中，我們嘗試先由行政院主計處 1990 年的戶宅普查資料著手，建立一套可以透過網路瀏覽器直接呈現在使用者眼前的台灣社會地圖系統。

2. 系統概述

目前「台灣社會地圖」雖是一套前後連貫的系統，但實際上可切分為兩個子部分，第一個子部分是有關原始資料— 1990 年戶宅普查資料—的整理；第二個子部分則是統計資料的視覺化呈現。這兩個子部分可以各自獨立存在，但又可彼此整合。

本系統的運作程序如下圖所示：



有關各部分的實際運作過程，將於下面各節中描述。

本系統目前的建置環境、採用的軟硬體設備為：

硬體部分：

中央處理器（CPU）：Pentium Pro 200 Mz * 4

記憶體（RAM）：1G bytes

軟體部分：

作業系統採用 RedHatLinux.7.1 (<http://www.redhat.com/>.)

資料庫軟體採用 mySQL3.23 (<http://www.mysql.com/>)

部分程式在 JDK 1.3 上開發

網頁伺服器與動態網頁產生：

Apache Web Server 1.3019 (<http://httpd.apache.org/>)

PHP 4.05 (<http://www.php.net>)

XML 的轉換 (Transformation)：

James Clark 開發的 xt (<http://www.jclark.com/>)

**上述軟體均為自由軟體。*

原始資料集的整理與網頁的撰寫等則在 windows 2000 Advance Server + Internet Explorer 5.5 sp1 上進行。

本系統由於是實驗性質，因此測試了各種軟體，而導致目前採用的軟體紛雜，我們將在未來對軟體環境進行整合，初步認為部分程式若採用 Java 平台將較易於進行。此點僅在此處提出以供參考。

3. 原始資料集的整理

台灣社會地圖需要社經資料與地圖資料兩種資料集，由於這兩類的資料在過往的運用中被視為是兩種不同性質的資料，且擁有這些資料的單位均有其自身的標示方式與軟體規格，使資料集雖然電子化，但卻無法輕易地被再利用與直接整合。本計畫的最初步驟即在進行原始資料集的解讀與整理工作，這個工作的目標在使資料集能夠用 XML 的方式標示，採用 XML 標示，已如前述具有可交換的優點，因此在本節中，我們詳盡描述了對原始資料集進行標示的步驟與其中所面臨的問題，以提供所有願意以 XML 來標示資料的單位、個人參考。

3.1 1990 年主計處戶宅普查

台灣每十年由官方[2]進行一次戶口與住宅普查（以下簡稱「戶宅普查」）。由於調查範圍含括全體住民，調查項目又包括如經濟、衛生、住宅等項目，戶宅普查的結果成為許多有關台灣社經情景研究之基本資料。但是過往的調查最後多以紙本的方式發行，一方面導致再利用上的不便，另一方面由於僅發佈由固定選項所組成的統計結果，間接使其豐富資訊之利用反受到限制。因此，本系統計畫以戶宅普查的最原始資料[3]為統計核心，並且賦予使用者最大的可能來決定其所需的統計條件。

就原先的構想，是希望能夠直接使用最近一次（2000 年）的戶宅普查資料，但由於 2000 年的資料約在 2003 年後才能整理妥當，因此我們先以前次（1990 年）的戶宅普查作為本計畫先期研究的核心資料。

1990 年戶宅普查原始資料電子檔可向行政院主計處免費索取[4]。該電子檔是以純文字模式儲存，其資料表現方式完全依照戶宅普查表登錄，即每筆（每行）資料即為一張普查表，其中包含屬於同一戶口的多人（至多五人，超出五人則另成一筆）。

就本系統的目標來看，我們認為原始資料必須重新整理以方便運用，整理的原則與步驟如下：

1. 以每一張普查表為一單位的資料表示方式，在統計上並沒有實質意義。就戶宅普查的調查方式而言，普查表可以被抽出、重新整理為以個人（individual）、以家戶（household）以及以住宅（house）為單位的三種資料集。本計畫的先期研

究尚無法同時處理此三種資料集，而僅以最常使用的個人資料集為基底資料（以下所稱之戶宅普查原始資料集均指稱此種個人資料集）。

2. 以每一個個人為單位的資料集本身，可以以 XML 的方式來標示，我們也進行了類似的嘗試：

●每個個人的 XML 資料集

```
<?xml version="1.0" encoding="big5" ?>
<!DOCTYPE people (View Source for full doctype...)>
- <people>
- <individual>
- <household>
  <county>台東縣</county>
  <town>蘭嶼鄉</town>
  <village>椰油村</village>
  <conduct_area>001</conduct_area>
  <census_area>1</census_area>
  <houseid>001</houseid>
  <householdid>01</householdid>
  <category>普通住戶</category>
</household>
- <person serialnumber="001">
  <sex>男</sex>
  <terminology>戶長</terminology>
  <birthday beforeDec15="True">1946</birthday>
- <hometown>
  <hometownarea>台閩地區</hometownarea>
  <county>台東縣</county>
</hometown>
- <fiveyrago>
  <fiveyrago_place>同縣不同鄉</fiveyrago_place>
</fiveyrago>
  <marriage>離婚分居</marriage>
- <education>
  <status>畢業</status>
  <level>專科</level>
  <department>12</department>
</education>
- <physical>
  <pstatus>良</pstatus>
</physical>
- <economic>
  <isemployed>有工作</isemployed>
  <industry>931</industry>
  <occupation>133</occupation>
  <title>政府雇用</title>
  <isfamilylivelihood>主要家計人</isfamilylivelihood>
</economic>
- <workorstudy>
  <isworkorstudy>有工作或求學</isworkorstudy>
  <ws_area>同鄉市鎮</ws_area>
</workorstudy>
</person>
</individual>
</people>
```

[以 XML 檔案方式呈現](#)

但由於戶宅普查的資料量太大（19,974,329 筆，每筆資料 31 個欄位），採用 XML 標示更必須負擔額外的標示空間，同時將 XML 作為一個資料集的查詢語法目前仍然在發展中，因此本系統現階段仍然採用將原始資料集置入現有的關連式資料庫軟體中的方式。目前本系統中原始資料集的格式如下：

資料庫軟體中戶宅普查資料各欄位格式 (mysql)

```
mysql> show fields from census1990_people;
```

Field	Type	NULL	Key	Default	Extra
code	char(10)	YES	MUL	NULL	
conduct_area	smallint(5) unsigned zerofill	YES		NULL	
census_area	tinyint(3) unsigned zerofill	YES		NULL	
houseID	smallint(5) unsigned zerofill	YES		NULL	
householdID	tinyint(3) unsigned zerofill	YES		NULL	
householdcategory	tinyint(3) unsigned zerofill	YES		NULL	
personID	tinyint(3) unsigned zerofill	YES		NULL	
sex	tinyint(3) unsigned zerofill	YES		NULL	
terminology	tinyint(3) unsigned zerofill	YES		NULL	
birth_year	smallint(5) unsigned	YES		NULL	
birth_beforeDec25	tinyint(3) unsigned	YES		NULL	
hometown_area	tinyint(3) unsigned zerofill	YES		NULL	
hometown_county	tinyint(3) unsigned zerofill	YES		NULL	
fiveyago_place	tinyint(3) unsigned zerofill	YES		NULL	
fiveyago_county	tinyint(3) unsigned zerofill	YES		NULL	
marriage	tinyint(3) unsigned zerofill	YES		NULL	
edu_status	tinyint(3) unsigned zerofill	YES		NULL	
edu_level	tinyint(3) unsigned zerofill	YES		NULL	
edu_department	tinyint(3) unsigned zerofill	YES		NULL	
physical_status	tinyint(3) unsigned zerofill	YES		NULL	
physical_condition	tinyint(3) unsigned zerofill	YES		NULL	
isemployed	tinyint(3) unsigned zerofill	YES		NULL	
unemployedreason	tinyint(3) unsigned zerofill	YES		NULL	
isfamilylivelihood	tinyint(3) unsigned zerofill	YES		NULL	
industry	smallint(5) unsigned zerofill	YES		NULL	
occupation	smallint(5) unsigned zerofill	YES		NULL	
title	tinyint(3) unsigned zerofill	YES		NULL	
isworkorstudy	tinyint(3) unsigned zerofill	YES		NULL	
isworkorstudy_area	tinyint(3) unsigned zerofill	YES		NULL	
isworkorstudy_countyortown	smallint(5) unsigned zerofill	YES		NULL	
livingcostfrom	tinyint(3) unsigned zerofill	YES		NULL	

```
31 rows in set (0.00 sec)
```

事實上，即使捨棄 XML 資料集而採用既有的資料庫軟體來儲存所有資訊，仍然有資料量過大以致於難以進行查詢動作等的問題，這個問題以及我們所嘗試解決的方式，將在後文中說明。

3. 為節省儲存空間與方便查詢，資料庫中所儲存的資料（除出生年外）均以代碼的方式表示。代碼的類型可以分成兩大類：

a. 專屬於本份資料的代碼，譬如有關於婚姻狀況，以 1（數字格式）表示「未婚」、2 表示「有配偶同居」、3 表示「離婚分居」、4 表示「喪偶」等。在某些不適用的欄位中，則以 NULL 表示未調查。

b. 由行政單位所發佈的標準分類代碼，這類代碼可作為與其他社經資料整合的參照標準。包括如行政區域代碼（「中華民國臺閩地區各市、鎮、鄉、村里代碼」[\[5\]](#)）以及行業、職業代碼。

有關標準分類代碼，在處理中我們面臨許多問題，其中主要的問題在，此三種代碼均在 1990 年戶宅普查之後才建立（或修正），因此在原始資料的轉換過程中，我們必須將戶宅普查的原始資料中採用的代碼轉換為現行的代碼。然而，不論就行政區代碼或職業與行業代碼來看，原始資料都不一定能夠直接對應到現行的代碼。

這個問題，在職業與行業部分，由於目前的系統並不依靠其作為參照，因此我們尚未進行此部分代碼的轉換工作。但就行政區代碼之部分，則因其為後續之視覺化工作的主要與唯一參照，而成為首要解決的工作。

「中華民國臺閩地區各市、鎮、鄉、村里代碼」之代碼表是由行政院主計處於 1991 年制訂，並於 1995 年之後陸續修訂，其制訂時為各個行政區賦予一唯一編號，即使該行政區消失、改變，其編號仍然繼續存在。行政區編號的獨一性使得即使資料之調查年不同，也不會因此產生混亂，彼此能相互參照。但理論上的優點，在本系統的實際使用上，仍有一些問題待解，包括：

a. 目前採用的資料集為該代碼表制訂前所統計，因此某些行政單位並未出現在該份代碼表中。這個問題我們僅能給予其另一編號來試圖解決。幸運的是，這個部分只出現在金馬地區，而金馬地區，由於地圖取得的問題，目前尚未能在我們的系統中呈現。

b. 台灣地區村里的變動率極頻繁，常數個村里集合為一個村里或者反之一個（或一個以上）村里分割為數個村里。不管何種情形，仍有一代碼指稱原來的村里，但顯然變動前後之（同一代碼）村里的範圍、人口等均有變化。這部分的變化如何在本系統中呈現（或者不需呈現）仍待解決。

c.由於台灣的行政區域劃分採取省—縣市—鄉鎮市區—村里（四級）與直轄市—區—村里（三級）兩軌制，在實際運用上我們面臨了兩者無法對應的狀況，亦即如台北市、高雄市常同時與台灣省為一比較單位，也常與台北縣等縣市為同一比較單位。此種不對應也表現在行政區代碼的編制上，即直轄市所屬之區同時與縣市與鄉鎮市相對應。我們目前暫時只能忽略此問題，而在本系統中略去省級，直接將直轄市視作一般縣市來處理。

統計資料的 XML 化

根據上述原則與步驟，我們對原始資料進行初步整理並且將之轉存入資料庫軟體中，下一階段的工作即在將統計的結果以 XML 標示。由於一份戶宅普查資料具有許多不同的問題選項（類目），而能有各種不同的統計方式，因此其所產生出的 XML 必然是具極大可變性的。以「婚姻」、「國籍」與「性別」三個類目為例，可以產生如未婚、本國籍之男性，與本國籍、男性且未婚等不同的、具有統計意義的排列組合。因此我們嘗試建構出一個得以描述各種統計方式的「文件結構定義」（DTD，Document Type Definition）。

就統計資料的性質來看，各統計類目可分為獨立類目與非獨立類目兩種，前者不依附於其他類目而為可獨立存在、具有統計意義的單位；後者則是必須臣屬於某個獨立類目（或其子類目）而不能包含其父類目的統計資訊。經由這些類目的組合，即可得到使用者所需的統計資訊。

如「婚姻」為一獨立類目，它可以包含其他的獨立類目，其 DTD 表示法則如下：

```
<!ENTITY % count "count CDATA #REQUIRED">
<!ENTITY % 婚姻外的獨立類別 "性別*出生年*籍別*教育程度*教育狀況*身體狀況*勞動力狀況*行業*職業*從業身分*是否為主要家計人*就業求學狀況*生活費來源*">
<!ELEMENT 婚姻 (%婚姻外的獨立類別;)>
<!ATTLIST 婚姻
    類型 (未婚|有配偶同居|離婚分居|喪偶) #REQUIRED
    %count;>
```

非獨立類目，則必須屬於某個獨立類目，如下例中的本國籍與籍別之關係：

```
<!ENTITY % 籍別外的獨立類別 "婚姻*性別*出生年*教育程度*教育狀況*身體狀況*勞動力狀況*行業*職業*從業身分*是否為主要家計人*就業求學狀況*生活費來源*">
<!ELEMENT 籍別 (本國籍*|外國籍*|%籍別外的獨立類別;)>
<!ATTLIST 籍別
```

```

    國籍(本國籍|外國籍) #REQUIRED
    %count;>
<!ELEMENT 本國籍 (台閩地區*|他省市*|%籍別外的獨立類別;)>
<!ATTLIST 本國籍
    地區 CDATA #REQUIRED
    %count;>
<!ELEMENT 外國籍 (%籍別外的獨立類別;)>
<!ATTLIST 外國籍
    地區 (美國|日本|東南亞地區|其他國家) #REQUIRED
    %count;>
<!ELEMENT 台閩地區 (%籍別外的獨立類別;)>
<!ATTLIST 台閩地區
    縣市 CDATA #REQUIRED
    %count;>
<!ELEMENT 他省市 (%籍別外的獨立類別;)>
<!ATTLIST 他省市
    省市 CDATA #REQUIRED
    %count;>

```

有關於非獨立類目，必須再做說明。以上述籍別為例，「籍別」為一獨立類目元素（element），其下包含著「本國籍」、「外國籍」、「台閩地區」、「他省市」等非獨立類目元素，這類非獨立類目元素和獨立類目元素之差異性在於，非獨立類目必然得伴隨其父元素（parent element）出現，如「本國籍」、「外國籍」必然得伴隨「籍別」元素，「台閩地區」、「他省市」必然得隨「本國籍」出現等。元素間的關係即是具有階序性的，但是並不必然每次進行含有非獨立類目之類目的統計時，都得將全部包含的非獨立類目都進行統計，而是隨時可終止的。亦即，如上述之例，使用者可以只統計到「本國籍」（與「外國籍」）之層級，無須統計「台閩地區」、「他省市」等而可直接統計如「婚姻」、「性別」等其他獨立類目。

有關於獨立與非獨立類目之關係，可參照以下 XML 文件：

範例一、僅有獨立類目（婚姻、性別）

```
<?xml version="1.0" encoding="big5" ?>
<!DOCTYPE TaiwanSocialMap (View Source for full doctype...)>
- <TaiwanSocialMap 調查單位="行政院主計處" 資料名稱="戶宅普查" 調查時間="1990">
- <行政區域 代碼="_6300100001" 類型="村里" 名稱="台北市松山區金泰里">
- <基本資訊>
  <土地面積 面積="2.6996" 單位="平方公里" />
</基本資訊>
- <人口統計 描述="統計" count="4525" 單位="人">
- <婚姻 類型="未婚" count="2383">
  <性別 性別="男" count="1376" />
  <性別 性別="女" count="1007" />
</婚姻>
- <婚姻 類型="有配偶同居" count="1814">
  <性別 性別="男" count="927" />
  <性別 性別="女" count="887" />
</婚姻>
- <婚姻 類型="離婚分居" count="120">
  <性別 性別="男" count="81" />
  <性別 性別="女" count="39" />
</婚姻>
- <婚姻 類型="喪偶" count="208">
  <性別 性別="男" count="67" />
  <性別 性別="女" count="141" />
</婚姻>
</人口統計>
</行政區域>
</TaiwanSocialMap>
```

[以 XML 檔案方式呈現](#)

範例二、獨立類目的另一種統計方式（性別、婚姻）

```
<?xml version="1.0" encoding="big5" ?>
<!DOCTYPE TaiwanSocialMap (View Source for full doctype...)>
- <TaiwanSocialMap 調查單位="行政院主計處" 資料名稱="戶宅普查" 調查時間="1990">
- <行政區域 代碼="_6300100001" 類型="村里" 名稱="台北市松山區金泰里">
- <基本資訊>
  <土地面積 面積="2.6996" 單位="平方公里" />
</基本資訊>
- <人口統計 描述="統計" count="4525" 單位="人">
- <性別 性別="男" count="2451">
  <婚姻 類型="未婚" count="1376" />
  <婚姻 類型="有配偶同居" count="927" />
  <婚姻 類型="離婚分居" count="81" />
  <婚姻 類型="喪偶" count="67" />
</性別>
- <性別 性別="女" count="2074">
  <婚姻 類型="未婚" count="1007" />
  <婚姻 類型="有配偶同居" count="887" />
  <婚姻 類型="離婚分居" count="39" />
  <婚姻 類型="喪偶" count="141" />
</性別>
</人口統計>
</行政區域>
</TaiwanSocialMap>
```

[以 XML 檔案方式呈現](#)

範例三：包括非獨立類目（籍別、性別）

```
<?xml version="1.0" encoding="big5" ?>
<!DOCTYPE TaiwanSocialMap (View Source for full doctype...)>
- <TaiwanSocialMap 調查單位="行政院主計處" 資料名稱="戶名普查" 調查時間="1990">
- <行政區域 代碼="_6300100001" 類型="村里" 名稱="台北市松山區金泰里">
- <基本資訊>
  <土地面積 面積="2.6996" 單位="平方公里" />
</基本資訊>
- <人口統計 描述="統計" count="4525" 單位="人">
- <籍別 國籍="本國籍" count="4525">
- <本國籍 地區="台灣地區" count="4280">
  <性別 性別="男" count="2267" />
  <性別 性別="女" count="1993" />
</本國籍>
- <本國籍 地區="他省市" count="245">
  <性別 性別="男" count="164" />
  <性別 性別="女" count="81" />
</本國籍>
</籍別>
- <籍別 國籍="外國籍" count="0">
- <外國籍 地區="美國" count="0">
  <性別 性別="男" count="0" />
  <性別 性別="女" count="0" />
</外國籍>
- <外國籍 地區="日本" count="0">
  <性別 性別="男" count="0" />
  <性別 性別="女" count="0" />
</外國籍>
- <外國籍 地區="東南亞地區" count="0">
  <性別 性別="男" count="0" />
  <性別 性別="女" count="0" />
</外國籍>
- <外國籍 地區="其他國家" count="0">
  <性別 性別="男" count="0" />
  <性別 性別="女" count="0" />
</外國籍>
</籍別>
</人口統計>
</行政區域>
</TaiwanSocialMap>
```

[以 XML 檔案方式呈現](#)

每個元素間關係之定義方式已如上述，元素內，則均具有兩個屬性（attribute），其一是依據元素類型而有不同名稱的選項，這個選項屬性的值，在可能的範圍內均以列舉的方式來表示；另一是統計數據擺放的 count 屬性。這兩個屬性均是必須的（#REQUIRED）。

根據前述的 DTD 制訂原則，我們可以保留各種統計方式、允許使用者進行各種的統計組合。甚至，使用者亦可（不透過本系統）自行進行資料統計之工作，只要其統計時依據本 DTD 的規範來建構一份有效的 XML 統計文件，理論上即可被後端的系統處理。

不過，由於 DTD 只規範了父子節點（parent-child node）之間的關係，事實上仍有許多有效但無意義的情形，譬如：

a. 本 DTD 允許每個獨立類目元素下接其他除其自身之外的獨立類目元素，以保留統計的多樣性，但卻也使得無法經由 DTD 的驗證來確保一個獨立類目元素是否會在其兩階之後出現。以婚姻、籍別二元素為例，透過 DTD 我們可以確認如婚姻-婚姻-籍別之 XML 文件無效，但婚姻-籍別-婚姻之文件卻會被視為有效，雖然其顯然不具統計意義；

b. 再者，理論上每個層級的資料都必須平行發展，以「婚姻」、「籍別」、「性別」、「教育狀況」四個元素為例，若我們要求得知其中「婚姻」、「籍別」、「性別」的統計結果，產生的 XML 樹狀結構應如下：

「婚姻」、「籍別」、「性別」組合

```
<?xml version="1.0" encoding="big5" ?>
<!DOCTYPE TaiwanSocialMap (View Source for full doctype...)>
- <TaiwanSocialMap 調查單位="行政院主計處" 資料名稱="戶宅普查" 調查時間="1990">
- <行政區域 代碼="_6300100001" 類型="村里" 名稱="台北市松山區金泰里">
- <基本資訊>
  <土地面積 面積="2.6996" 單位="平方公里" />
</基本資訊>
- <人口統計 描述="統計" count="4525" 單位="人">
- <婚姻 類型="未婚" count="2383">
- <籍別 國籍="本國籍" count="2383">
  <性別 性別="男" count="1376" />
  <性別 性別="女" count="1007" />
</籍別>
- <籍別 國籍="外國籍" count="0">
  <性別 性別="男" count="0" />
  <性別 性別="女" count="0" />
</籍別>
</婚姻>
- <婚姻 類型="有配偶同居" count="1814">
- <籍別 國籍="本國籍" count="1814">
  <性別 性別="男" count="927" />
  <性別 性別="女" count="887" />
</籍別>
- <籍別 國籍="外國籍" count="0">
  <性別 性別="男" count="0" />
  <性別 性別="女" count="0" />
</籍別>
</婚姻>
- <婚姻 類型="離婚分居" count="120">
- <籍別 國籍="本國籍" count="120">
  <性別 性別="男" count="81" />
  <性別 性別="女" count="39" />
</籍別>
- <籍別 國籍="外國籍" count="0">
  <性別 性別="男" count="0" />
  <性別 性別="女" count="0" />
</籍別>
</婚姻>
- <婚姻 類型="喪偶" count="208">
- <籍別 國籍="本國籍" count="208">
  <性別 性別="男" count="67" />
  <性別 性別="女" count="141" />
</籍別>
- <籍別 國籍="外國籍" count="0">
  <性別 性別="男" count="0" />
  <性別 性別="女" count="0" />
</籍別>
</婚姻>
</人口統計>
</行政區域>
</TaiwanSocialMap>
```

[以 XML 檔案方式呈現](#)

但根據 DTD 的規範，若其中穿插「教育狀況」元素而產生的 XML 文件亦為有效：

有效但不合法之組合

```
<?xml version="1.0" encoding="big5" ?>
<!DOCTYPE TaiwanSocialMap (View Source for full doctype...)>
- <TaiwanSocialMap 調查單位="行政院主計處" 資料名稱="戶宅普查" 調查時間="1990">
- <行政區域 代碼="_6300100001" 類型="村里" 名稱="台北市松山區金泰里">
- <基本資訊>
  <土地面積 面積="2.6996" 單位="平方公里" />
</基本資訊>
- <人口統計 描述="統計" count="4525" 單位="人">
- <婚姻 類型="未婚" count="2383">
- <籍別 國籍="本國籍" count="2383">
  <性別 性別="男" count="1376" />
  <性別 性別="女" count="1007" />
</籍別>
- <籍別 國籍="外國籍" count="0">
  <性別 性別="男" count="0" />
  <性別 性別="女" count="0" />
</籍別>
</婚姻>
- <婚姻 類型="有配偶同居" count="1814">
- <籍別 國籍="本國籍" count="1814">
  <性別 性別="男" count="927" />
  <性別 性別="女" count="887" />
</籍別>
- <籍別 國籍="外國籍" count="0">
  <性別 性別="男" count="0" />
  <性別 性別="女" count="0" />
</籍別>
</婚姻>
- <婚姻 類型="喪偶" count="208">
- <籍別 國籍="本國籍" count="208">
  <教育狀況 就學狀況="在學" count="0" />
  <教育狀況 就學狀況="畢業" count="63" />
  <教育狀況 就學狀況="肄業" count="10" />
  <教育狀況 就學狀況="檢定合格" count="0" />
  <教育狀況 就學狀況="未就學" count="135" />
</籍別>
- <籍別 國籍="外國籍" count="0">
  <教育狀況 就學狀況="在學" count="0" />
  <教育狀況 就學狀況="畢業" count="0" />
  <教育狀況 就學狀況="肄業" count="0" />
  <教育狀況 就學狀況="檢定合格" count="0" />
  <教育狀況 就學狀況="未就學" count="0" />
</籍別>
</婚姻>
</人口統計>
</行政區域>
</TaiwanSocialMap>
```

[以 XML 檔案方式呈現](#)

而雖然此兩種情形對 DTD 而言都是有效的，但後者的統計並不具有實質意義。

儘管有上述這些有效但在統計上不具意義的情形，但是我們認為這些錯誤，可以很容易的由其他程式來驗證，並且也可撰述其他的文件以為規範。

根據主計處所做的統計，我們目前暫時整理出 14 項獨立類別以為試驗。哪些統計項目能夠完整表現出原始資料的性質，實須經由社會科學家的共同參與。目前本系統使用的 DTD 請參見[附錄一](#)。

難題與待解決之事

除前文中以描述者外，戶宅普查原始資料集的整理工作，仍然有許多額外的待解決難題。這些難題，在目前的實驗系統中尚不構成問題，但如將本系統適用範圍擴大，則必得解決。已知的難題與我們所設想的解決方式如下所述：

1. 資料量過於龐大以致於查詢速度過於緩慢。理論上，我們可以從儲存（整理過後的）原始資料集的資料庫軟體之資料表（table）中直接透過 SQL 語法取得所需的統計資料，但是由於原始資料集是以每個個人為單位，總資料共達 19974329 筆（人），每次統計均在如此龐大的資料表格中計算，在目前的軟硬體設備中，使用者必須久待。為改善此情形，我們目前採取的方式是將常用的統計資訊先行統計成另一資料表，並將資料表依最小行政單位（村里）切割（有關此統計方式將在第四節中描述）。但此方式僅可部分加快搜尋速度，仍非根本的解決之道。

2. 含括多種類型之資料的難題。目前本系統目前所採用的資料僅包括 1990 年戶宅普查之人口部分一項，因此所有的統計與 DTD 設定方式均以單一年度單一資料為試驗模式。但我們最終的目標之一是能夠統合各種社經資料，因此如何去設計一個能處理多年度、多資料類型的程序，是進行下一階段工作時必須面對的議題。

3. 原始資料的統計問題。在整理完原始資料集後，本系統已經能夠透過程式計算出某些統計的數據，為了驗證這些數據是否正確，於是我們比對了 1990 年戶宅普查的報告書的統計結果[6]，但此二統計結果並不相符，經過檢視戶宅普查報告書的統計數據，我們發現部分統計項目無法經由原始資料集得知。另外，我們又比對了行政院主計處所提供的 1990 戶宅普查線上統計系統（七十九年戶口普查台灣電子地圖查詢系統 <http://140.129.146.192/census~n/six/lue5/census-n.htm>）的統計結

果，其結果卻又與本系統之統計大致相符[7]。這兩份資料間的差異如何發生，我們仍未進一步詳究。

4. 編碼（encoding）的問題。目前中文文字的編碼多半採用大五碼（big5），不過許多處理 XML 的工具軟體並不支援 big5 編碼，甚至由於大五碼中包含了部分控制碼，即使如 mySQL 等資料庫軟體在儲存資料時也會發生錯誤。目前本系統採用的編碼方式，由於因應所試驗的工具軟體而做過多次修改，因此不同檔案的編碼方式並不一致，我們認為，這些編碼系統應該整合，可行的方式則是採用標準萬國碼（unicode）[8]。這個問題，將在第五節結論中再做探討。

3.2 地圖資料

既然本系統的目標之一是將有關台灣之社經資訊以視覺化的方式呈現，而最直覺的視覺化呈現方式之一是透過地圖來表達，因此尋找一份可搭配社經資訊之地圖即成爲我們所必須進行的工作之一。

由於目前本系統採用 1990 年戶宅普查爲核心資料，其資料可以村里等行政區爲統計單位，因此搭配此資料集的地圖應爲行政區域的劃分圖。在最初的認知中，找到這份地圖應非難事，我們也查詢到各種某些行政單位存有此份資料的資訊，但是最後卻發現實際上目前並沒有任何的行政單位可提供如此的地圖。

最後，我們幸運的得到本院（中研院）計算中心 GIS 小組的協助，取得了 1996 年的行政區域劃分圖[9]。雖然由於行政區域的變動頻仍，1996 年的行政區劃分圖無法直接完整套入 1990 年的統計資料中，但若要將此地圖完全校正回 1990 年的樣態，實非我們單薄的人力所能完成。因此本系統仍暫時以此份地圖爲試驗基準[10]。

地圖的 XML 化

現行的各種電子地圖系統，大多採用商用軟體（如 AcrView, Autocad 等），這些商用軟體提供了強大的功能。建構者只要去學習該軟體的使用方式，就能夠雖不算輕鬆但也可說是簡單的建立電子地圖的系統。不過商用軟體一般來說耗費昂貴，關於系統功能的改變也絕大部分得透過系統廠商的升級、更新，更重要的是，其資料規格多未公開，對電子資訊的流通反而是種阻礙。本系統最初採用 XML 的原因

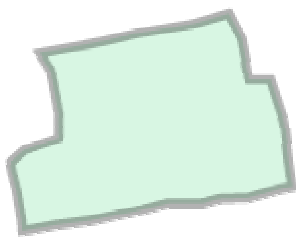
之一，即在於使電子資料公開化，以使使用者能夠輕易的再利用。在這種考量下，取得地圖之後的工作即在將地圖資料以 XML 標示。

我們採用 Scalable Vector Graphics (SVG)[11]來進行地圖資料的標示，SVG 為 W3C 所制訂的一種採用 XML 來描繪二維圖形的語言，使用 SVG 描述的圖形資料也保留了 XML 資料公開的特性，並且可直接在使用者的瀏覽器端呈現[12]。

我們所取得的地圖是以 ArcView 檔 (*.shp)包裝，透過 Nedjo Rogers 所撰述的一套 script[13]，直接將其轉換為 SVG，為便利於再利用，我們又將之轉換為相對座標並計算出每個區域的範圍 [14]，這些轉換出來的地圖資訊，均是以純文字的方式儲存。在經過進一步的整理之後，我們將每個區域的地圖資訊存入資料庫中，並以行政區域代碼為其索引參考值。以台北市松山區莊敬里為例，標示後的地圖 SVG 描述為：

```
<?xml version="1.0" encoding="UTF-8"?>
<svg height="400" width="400" viewBox="2080930 266554 4516 8217">
<g id="villageBoundary" stroke="#000000" stroke-width="100" stroke-opacity="0.3" fill="#d8f6e3"
  fill-opacity="1" stroke-antialiasing="true" class="showme">
<path id="6300100002" d="M2081321,2701171413,-731262,-511361,-831580,-951374,-
1081133,263177,4271228,11146,463153,
454133,2151-141,241-325,891-299,111-202,511-347,-11-302,751-511,1051-364,471-351,621-9,-1071-50,-3091-68,-3061146,
-1281361,-12415,-2561-84,-4591-119,-187z"/>
```

其顯示結果為：



[以 SVG 檔案方式呈現 \(需支援 SVG 之瀏覽器\)](#)

難題與待解決之事

將地圖轉換為 SVG 以供再利用的過程中，我們依舊遇到了一些難題，包括如：

1. 由於 SVG 是以純文字的方式儲存，當所包含的資訊越細膩時，其檔案大小必然會倍增。這雖然可透過將 SVG 檔案壓縮來去得部分的改散，但是對某些地圖來看，仍然必須多耗時間來下載此圖。

2. 或許是由於目前關於 SVG 的顯示程式仍在發展中，SVG 地圖在最終使用者的顯示時，十分耗用資源。以一張包含台北縣市各村里的地圖來看，我們初步的計算是其約需佔用 30M bytes 的動態記憶體。對於某些硬體環境來說，會是沈重的負擔。

3. 最複雜的難題是在地圖的本身，如同前述，1996 年的行政劃分圖與 1990 年的戶宅普查資料本即有整合上的問題，有些村里消失，有些村里新生，這導致某些 1990 年的統計資訊無法在地圖上呈現，或者會呈現錯誤的訊息。有關此問題，仍待進一步去取得各村里的實際邊界異動狀況方能解決。另一方面，目前的行政區劃分方式中，土地和村里並非必然勾連的，亦即某些區域會是不屬於任何村里的，但在地圖上仍有這些區域存在，因此我們仍然必須給予其一個編號以便識別。

4. 如同戶宅普查資料的整理，本系統目前僅只處理單一年度單一類型的資料，若要處理多年度多類型時，仍如前節中所描述的須處理整合的問題。

4. 現行系統的建置

在第二節中已經敘述，本系統可以區分為兩個子系統，兩者可獨立存在，本節因此也分為兩部分來敘述，第三部分則描述此二子系統的整合。

4.1 原始資料集的 XML 產生

由於普查資料之原始資料集過於龐大，我們因此先從目前存在的 14 個獨立類目中，抽出婚姻、性別、出生年、籍別（包括國籍與籍貫）、教育狀況、教育程度（包括程度與學歷）、身體狀況（包括有無不良與不良種類）等七個類目來做實驗。對這七個類目，先做初步的統計，並且依村里區分為不同的資料表。

在初步統計工作中，我們所進行的是將此七個類目所包含之各種屬性的可能組合進行人口數的統計。換言之，原始資料集中是以個人為每筆資料的單位，如：

code	conduct_area	census_area	houseID	householdID	householdcategory	personID	sex
1000101004	00001	001	00001	001	001	001	002
1000101004	00001	001	00001	001	001	002	001

terminology	birth_year	birth_beforeDec25	hometown_area	hometown_county	fiveyago_place
001	1908	1	001	004	001
006	1950	1	001	004	001

fiveyago_county	marriage	edu_status	edu_level	edu_department	physical_status	physical_condition
NULL	004	005	008	NULL	002	003
NULL	002	002	006	014	001	NULL

isemployed	unemployedreason	isfamilylivelihood	industry	occupation	title	isworkorstudy
003	006	002	NULL	NULL	NULL	002
001	NULL	001	00630	00400	001	001

isworkorstudy_area	isworkorstudy_countyortown	livingcostfrom
NULL	NULL	004
003	00004	NULL

2 rows in set (0.08 sec)

透過 SQL 語法，先將七個項目均具有同樣屬性的資料整理為：

```
mysql> select
  code,sex,birth_year,marriage,howntown_area,edu_status,edu_level,physucal_status,physical_condition,
  count(*) as mycount
from census1990_people
group by
  code,sex,birth_year,marriage,howntown_area,edu_status,edu_level,physucal_status,physical_condition;
```

code	sex	birth_year	marriage	hometown_area	edu_status	edu_level	physical_status	physical_condition	mycount
6300100001	001	1901	002	001	005	009	001	NULL	0000000001
6300100001	001	1901	004	001	005	008	001	NULL	0000000001
6300100001	001	1902	004	001	002	002	001	NULL	0000000001

3 rows in set (0.00 sec)

系統最後則直接由此統計結果資料表中取得進一步的統計資訊。透過這個事先整理的工作，可將資料集數量大幅減低，並減低統計資訊所需資訊。

在資料範圍上，本系統則先以台北市各村里之資料為限，每次只能查詢一個區，只會顯示村里資料。每次且只能查詢三項的獨立類目。在此情形下，每次查詢約略可在 2 分鐘之內完成[15]，為便於之後使用者的方便，我們並將每次的查詢存成暫存檔，如此當之後使用者欲查詢同樣的條件資料時，就毋須再長久等待。

如同前述，採用此種方式僅能暫時達成加快統計速度的目的，一旦擴展到全部的類目，統計資料集的縮減幅度並不明顯，而若又包含全國的資料，則所需查詢時間更可能高達數十分鐘以上。有關此問題，可能必須透過改善硬體配置（如採用磁碟矩陣）或實驗他種資料庫軟體來因應。

產生此 XML 的程式，是透過 Document Object Model (DOM)。其細節，請參見[附錄三](#)。

4.2 視覺化顯示子系統

視覺化顯示之子系統是用於資料在經過處理後呈現在使用者之瀏覽器端，並且可提供使用者進行一些額外操作的系統。這個系統包括了：

1. 資料的來源是一個符合視覺化顯示之 DTD（以下簡稱顯示用 DTD）的 XML 檔案。爲了能夠處理各種類型的資料，顯示用 DTD 的規範十分鬆散，關於資料的類型、性質等，我們都要求使用者在 XML 檔中定義（見[附錄二](#)）。

2. 雖然部分的瀏覽器（如 Microsoft Internet Explorer）可以透過其內建[\[16\]](#)的 XML 解析器處理 XML 檔案，但是我們認爲最終不應將本系統限制於特定的瀏覽器或環境中，因此有關 XML 的解析我們仍然在伺服器端先行處理，將 XML 轉換成 script（Javascript 或 EMACscript）檔案之後交由瀏覽器處理。

將 XML 轉換爲其他類型檔案的方式有多種，目前本系統採用 XSLTransformation（XSLT）[\[17\]](#)。

3. 轉換爲 script 之後，餘下的工作在統計資料如何與地圖整合[\[18\]](#)，並且產生各種使用者所需之功能的問題。這兩者，我們都是透過 script，經由 XML DOM 來進行各種視覺上的應用，如繪製各種依顏色程度來表明統計數字的地圖，繪製圓餅、長條等圖表以及繪製表格等。這部分的工作仍可再繼續發展。請直接參考網站上所提供的測試。

4.3 原始資料集與視覺化顯示子系統的整合

如同前述，原始資料集能夠產生一個符合其特定 DTD 的 XML 檔案，視覺化顯示則需要一個符合顯示用的 DTD 的 XML 檔，因此我們仍然必須要將前者轉換爲後者。這部分的工作仍然是透過 XSLT 來進行，只是由於 XSLT 的語法有許多的限制，原始資料集的選項又有些複雜，因此我們最後決定採用動態產生 XSLT，而非一個能夠處理所有原始資料集的 XSL 之方式來將前者轉換爲後者。有關於動態產生 XSLT 的細節，請參閱[附錄三](#)。

將台灣社會地圖的兩個子系統整合之後，使用者即可自行定義查詢條件，而後透過瀏覽去取得描述統計結果的 XML 檔案，此檔案再經過轉換成視覺化顯示用 XML 與 Script 檔案後，搭配地圖呈現。除了地圖之外，使用者並可根據統計數據，直接由系統取得各種統計圖表。最終的整個流程與結果，如以下各圖所示：

一、選擇查詢條件



二、第一次轉換



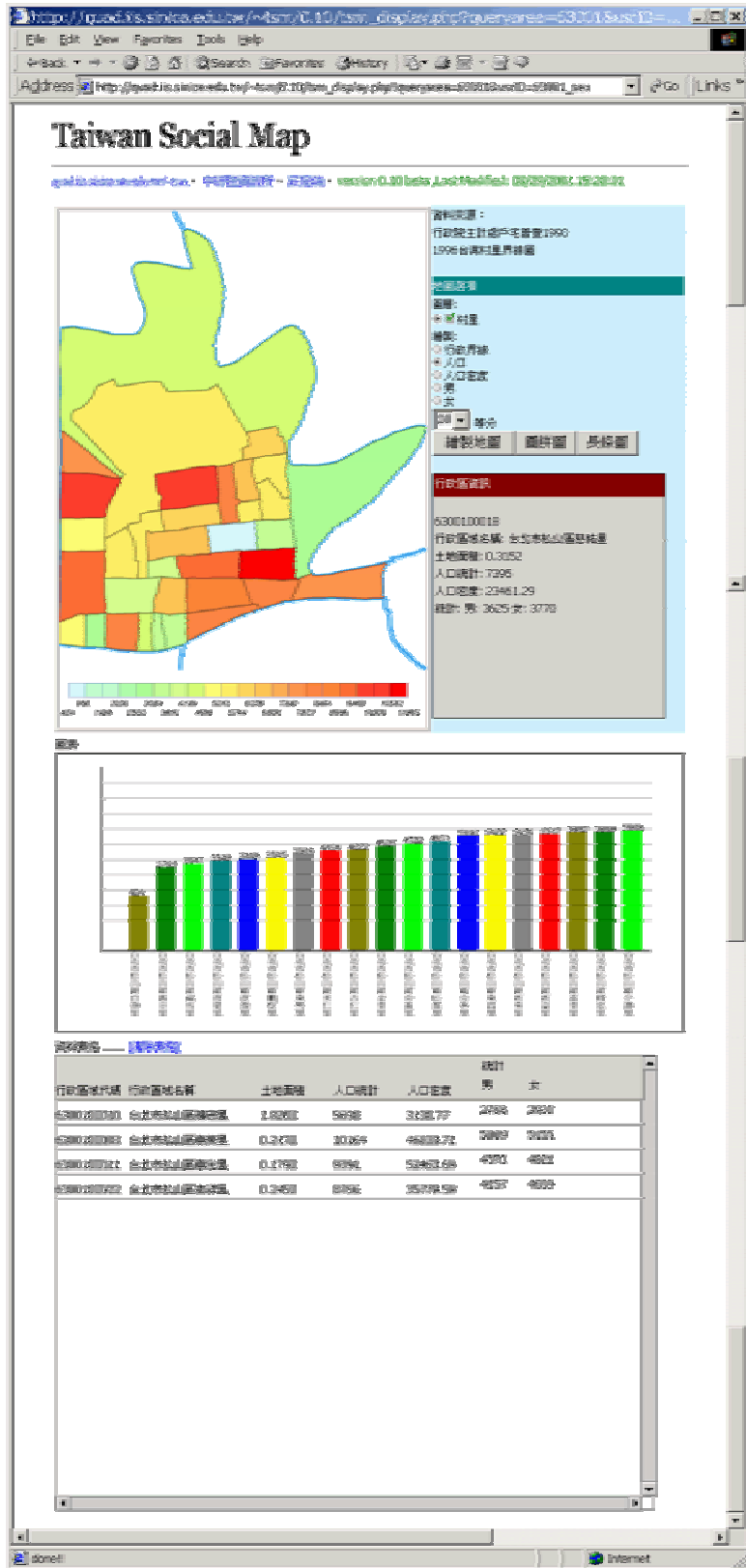
三、第二次轉換



四、轉換前後檔案列表



五、視覺化顯示結果



5. 目前的結論

經由以上的步驟，本系統因此試驗了包括原始社經與地圖資料集的蒐集，以及各種 XML 工具之試驗等工作。從這兩個部分來看，現行的體系與工具都有許多不足之處，這些不足之處也是造成各種資料雖電子化，但卻無法達到電子化資訊之流通性功能。在此節中，我們將簡單描述此二部分的問題。

5.1 原始資料的取得問題

2000 年，行政院通過行政機關電子資料流通實施要點，根據該要點，該要點之立意在促使政府機關之資訊得以便利流通，各機關所藏之資料必須電子化並且將之做成目錄擺置在網路上使一般機關、個人得公開查詢並取得[19]。

不過，在實務運作上卻仍然有許多困難。本計畫取得 1990 年普查資料的過程雖然十分順利，但由於行政機關並未有專人負責資料的提供，且各種資料的取得均需透過公文往返，以致於我們仍然花費時間以取得資料。同時，我們所取得的資料本身雖然已經數位化，但處理該資料所需的其他資訊，如各種代碼，卻仍然是紙本資料，以致於我們也必須額外花費自行將這些資料電子化的工作。

在取得電子地圖的過程中，我們則面臨了較大的障礙。首先，雖然前述要點規範政府資訊需數位化並提供流通，但也同時有不溯及既往的原則，因此在該要點訂立之前的各種資料，或未電子化，或未提供流通。加上政府機構各部門的業務移轉情形複雜，更加深了取得政府數位資訊的難度。以我們所需使用的電子地圖為例，根據戶宅普查法規定，戶宅普查時必須同時繪製普查區圖，但這份地圖一方面是由地方單位（戶政事務所）所繪製，其最終統整單位不明，另一方面在普查完成後，也未能得知是否有最終保存的單位，因此造成資料再利用的困難。而有關於行政區劃分圖本身，我們也未能取得何單位保存此類地圖的資訊。

行政機關電子資料流通實施要點雖然是立意良善的法規，但由於政府單位的複雜組織不易為一般民眾所理解，因此實需更易於使用的資料處理方式。我們建議或可將所有的數位（與非數位）資訊交由專責單位來處理（如國家檔案局），並且可經由網路直接提供使用者下載使用[20]。

5.2 XML 與其他軟體工具的限制

採用 XML 的方式可以達到資料格式公開化與流通的效果。由於資料的特性之一是其可能採用不同的語言，爲了能夠同時處理各種語言，XML 規範中允許 XML 文件採用各種文字編碼方式。以中文爲例，我們即可直接以中文作爲元素名稱，並且可以使用既有的大五碼編碼。

不過，由於大五碼中包含了控制碼，同時大五碼本身也非多語系編碼（multi-language），與其他語言並不完全相容，許多軟體也不提供大五碼的處理方式，因此我們計畫採用 unicode 來處理所有中文文件。在 unicode 中又可區分爲最常用的 utf-8（八位元編碼）與 utf-16（十六位元編碼）兩種，前者爲可變長度的編碼，在處理如英文之文字時，採用一個位元一個字元的方式，但在處理如中文文字時，一個字元則需三個以上的位元；後者則將所有字元均以兩個位元來編碼。經由這兩種方式的比較，我們認爲描述中文資料應以 utf-16 爲準，以避免資料儲存空間耗費過大的問題。

但現行的許多能夠處理 XML 文件的工具軟體，並不一定都完整支援 utf-16 的編碼方式[21]，以本系統所採用的 php 爲例，本身僅提供 utf-8 的編碼函數，無法提供 utf-16 編碼。在將中文元素名稱轉換爲 utf-8 時，也會同時出現問題。

另一方面，即使部分軟體已可使用 unicode 編碼，但是在後端處理上（如 javascript）仍然無法以中文作爲變數名稱等，以致於我們必須將原先用中文標示的社經統計 XML 檔案，轉換爲英文的元素名稱才能處理。

編碼的問題，其實也造成資料流通的阻礙。這是在強調 XML 的格式公開與可交換性時，也必須處理的問題。

儘管有這些問題，透過台灣社會地圖的系統，我們嘗試將有關台灣之社經資料 XML 化以公開交換，並且透過 SVG 等方式將這些資料在使用者端的瀏覽器中展現。這整套系統的最基本原則是資料的可再利用性，也就是任何使用者應可輕易的取得原始資料、並且進行其所需的各種處理。我們認爲，只有在如此的公開原則下，各種研究才有可能有快速的發展，並且能除去過往不同學域甚或不同學者間之資料無法彼此交換，造成重複資源的浪費等問題。

致謝

台灣社會地圖的系統建立有賴以下各單位與個人的協助。

1990 年戶宅普查資料由行政院主計處普查局提供。地圖由中研院計算中心提供。本計畫之經費並由中央研究院「資訊科技與社會轉型」主題計畫支援。

特別感謝中央研究院社會學研究所吳齊殷副研究員與我們一起討論、建立這套系統。除此之外，張欽隆協助處理伺服器端的程式處理，資訊所平行處理研究室的諸位同仁亦對本系統提供許多協助。謹此致謝。

註解

- [1] Extensible Markup language (XML) 1.0 (Second Edition), W3C Recommendation 6 October 2000.
<http://www.w3.org/TR/REC-xml>
- [2] 目前負責進行戶宅普查之調查與保存工作之政府單位為行政院主計處普查局。
- [3] 經除名化。
- [4] 索取方法、使用條件與費用等請參見行政院主計處普查局（第四局）<http://www.dgbas.gov.tw/>。
- [5] 各類標準分類代碼，請參見行政院主計處第三局（<http://www.dgbas.gov.tw/dgbas03/stat-n.htm>）。
- [6] 行政院戶口普查處，1992，中華民國七十九年臺閩地區戶口及住宅普查報告。台北市：行政院戶口普查處。又可參見行政院主計處 <http://www.dgbas.gov.tw/census~n/Six/lue5/HT4621.HTM>
- [7] 由於該系統中將某些村里合併為系統建置時的村里單位，因此產生部分差異。
- [8] 見 www.unicode.org
- [9] 計算中心 GIS 小組收集了各種地圖，其資訊可在 <http://www.sinica.edu.tw/cc/gis/index1/ccmap.htm> 查詢。又，本份 1996 年行政區域劃分圖並非由計算中心 GIS 小組所繪製，因此我們必須對其進行校正工作。
- [10] 這導致了目前系統中部分村里的統計資料無法表現在地圖之上。
- [11] Scalable Vector Graphics (SVG) 1.0 Specification, W3C Proposed Recommendation 19 July, 2001,
<http://www.w3.org/TR/SVG/>
- [12] 但目前並沒有任何商用的瀏覽器軟體可直接顯示 SVG 圖形，不過使用者可以下載如 Adobe SVG Viewer（<http://www.adobe.com/svg>）等免費的 plug-in 軟體以顯示 SVG 圖。
- [13] 參見 <http://www.carto.net> 中之 shp2svg。
- [14] 該 script 轉出的 svg 檔為絕對座標，在某些情形下，我們只想抽出某個區域的地圖，將之轉換為相對座標是較為方便的。
- [15] 若是出生年與其他選項混用，可能會超過此時限。
- [16] 實際上是作業系統（windows）內建而非瀏覽器內建。關於 Microsoft 的 xml 解析器，可參考 www.msdn.com。
- [17] XSL Transformations (XSLT) Version 1.0, W3C Recommendation 16 November 1999,
<http://www.w3.org/TR/xslt>
- [18] SVG 的地圖也可透過 script 每次動態產生，實際上我本系統是透過 script 來產生，只是為節省時間，並不每次都動態產生。產生 SVG 的原始程式碼，亦請參見：
- [19] 電子資料流通目錄 http://www.gov.tw/uw/uw_disp.exe/rdec2/edir/index.html。
- [20] 這部分或許牽涉到著作權的問題。但是我們認為政府單位的資訊不應該有著作權的限制。同時著作權也不應成為資料在利用的障礙。
- [21] 這其實是一種歧視，因為採用 utf-16 在處理英文文件時會使文件大小倍增（中文則不會），因此諸多軟體均以 utf-8 作為其預設或唯一的編碼方式。

附錄

1. censusxml.dtd

```
<?xml version="1.0" encoding="UTF-16"?>
<!ENTITY % count "count CDATA #REQUIRED">
<!ENTITY % 人口統計的全部獨立類別 "婚姻*性別*出生年*籍別*教育程度*教育狀況*身體狀況*勞動力狀況*行業*職業*從業身分*是否為主要家計人*就業求學狀況*生活費來源*">
<!ENTITY % 婚姻外的獨立類別 "性別*出生年*籍別*教育程度*教育狀況*身體狀況*勞動力狀況*行業*職業*從業身分*是否為主要家計人*就業求學狀況*生活費來源*">
<!ENTITY % 性別外的獨立類別 "婚姻*出生年*籍別*教育程度*教育狀況*身體狀況*勞動力狀況*行業*職業*從業身分*是否為主要家計人*就業求學狀況*生活費來源*">
<!ENTITY % 出生年外的獨立類別 "婚姻*性別*籍別*教育程度*教育狀況*身體狀況*勞動力狀況*行業*職業*從業身分*是否為主要家計人*就業求學狀況*生活費來源*">
<!ENTITY % 教育狀況外的獨立類別 "婚姻*性別*出生年*籍別*教育程度*身體狀況*勞動力狀況*行業*職業*從業身分*是否為主要家計人*就業求學狀況*生活費來源*">
<!ENTITY % 行業外的獨立類別 "婚姻*性別*出生年*籍別*教育程度*教育狀況*身體狀況*勞動力狀況*職業*從業身分*是否為主要家計人*就業求學狀況*生活費來源*">
<!ENTITY % 職業外的獨立類別 "婚姻*性別*出生年*籍別*教育程度*教育狀況*身體狀況*勞動力狀況*行業*從業身分*是否為主要家計人*就業求學狀況*生活費來源*">
<!ENTITY % 從業身分外的獨立類別 "婚姻*性別*出生年*籍別*教育程度*教育狀況*身體狀況*勞動力狀況*行業*職業*是否為主要家計人*就業求學狀況*生活費來源*">
<!ENTITY % 是否為主要家計人外的獨立類別 "婚姻*性別*出生年*籍別*教育程度*教育狀況*身體狀況*勞動力狀況*行業*職業*從業身分*就業求學狀況*生活費來源*">
<!ENTITY % 就業求學狀況外的獨立類別 "婚姻*性別*出生年*籍別*教育程度*教育狀況*身體狀況*勞動力狀況*行業*職業*從業身分*是否為主要家計人*生活費來源*">
<!ENTITY % 生活費來源外的獨立類別 "婚姻*性別*出生年*籍別*教育程度*教育狀況*身體狀況*勞動力狀況*行業*職業*從業身分*是否為主要家計人*就業求學狀況*">
<!ENTITY % 身體狀況外的獨立類別 "婚姻*性別*出生年*籍別*教育程度*教育狀況*勞動力狀況*行業*職業*從業身分*是否為主要家計人*就業求學狀況*生活費來源*">
<!ENTITY % 勞動力狀況外的獨立類別 "婚姻*性別*出生年*籍別*教育程度*教育狀況*身體狀況*行業*職業*從業身分*是否為主要家計人*就業求學狀況*生活費來源*">
<!ENTITY % 教育程度外的獨立類別 "婚姻*性別*出生年*籍別*教育狀況*身體狀況*勞動力狀況*行業*職業*從業身分*是否為主要家計人*就業求學狀況*生活費來源*">
<!ENTITY % 籍別外的獨立類別 "婚姻*性別*出生年*教育程度*教育狀況*身體狀況*勞動力狀況*行業*職業*從業身分*是否為主要家計人*就業求學狀況*生活費來源*">
<!ELEMENT TaiwanSocialMap (行政區域+)>
  <!-- ATTLIST TaiwanSocialMap
    調查單位 CDATA #REQUIRED
    資料名稱 CDATA #REQUIRED
    調查時間 CDATA #REQUIRED -->
<!ELEMENT 行政區域 (基本資訊,人口統計*,戶口統計*)>
  <!-- ATTLIST 行政區域
    代碼 ID #REQUIRED
    類型 (村里鄉鎮市區縣市省市) #REQUIRED
    名稱 CDATA #REQUIRED -->
<!ELEMENT 基本資訊 (土地面積)>
<!ELEMENT 土地面積 EMPTY>
<!-- ATTLIST 土地面積
```

```

        單位 CDATA "平方公里"
        面積 CDATA #IMPLIED>
<!ELEMENT 人口統計 (%人口統計的全部獨立類別);>
<!ATTLIST 人口統計
        單位 CDATA "人"
        %count;>
<!ELEMENT 婚姻 (%婚姻外的獨立類別);>
<!ATTLIST 婚姻
        類型 (未婚有配偶同居|離婚分居|喪偶) #REQUIRED
        %count;>
<!ELEMENT 性別 (%性別外的獨立類別);>
<!ATTLIST 性別
        性別 (男|女) #REQUIRED
        %count;>
<!ELEMENT 出生年 (%出生年外的獨立類別);>
<!ATTLIST 出生年
        西元年 CDATA #REQUIRED
        %count;>
<!ELEMENT 教育狀況 (%教育狀況外的獨立類別);>
<!ATTLIST 教育狀況
        就學狀況 (在學|肄業|畢業|檢定合格|未就學) #REQUIRED
        %count;>
<!ELEMENT 行業 (%行業外的獨立類別);>
<!ATTLIST 行業
        行業別 CDATA #REQUIRED
        %count;>
<!ELEMENT 職業 (%職業外的獨立類別);>
<!ATTLIST 職業
        職業別 CDATA #REQUIRED
        %count;>
<!ELEMENT 從業身分 (%從業身分外的獨立類別);>
<!ATTLIST 從業身分
        身分 (雇主|自營作業|私人雇用|政府雇用|無酬工作者) #REQUIRED
        %count;>
<!ELEMENT 是否為主要家計人 (%是否為主要家計人外的獨立類別);>
<!ATTLIST 是否為主要家計人
        是否為主要家計人 (不適用|主要家計人|非主要家計人) #REQUIRED
        %count;>
<!ELEMENT 就業求學狀況 (就業*|求學*|%就業求學狀況外的獨立類別);>
<!ATTLIST 就業求學狀況
        狀態 (不適用|就業|求學|未就業也未求學) #REQUIRED
        %count;>
<!ELEMENT 就業 (%就業求學狀況外的獨立類別);>
<!ATTLIST 就業
        地點 (現住處所|同鄉鎮市|不同鄉鎮市|非台閩) #REQUIRED
        %count;>
<!ELEMENT 求學 (%就業求學狀況外的獨立類別);>
<!ATTLIST 求學
        地點 (現住處所|同鄉鎮市|不同鄉鎮市|非台閩) #REQUIRED
        %count;>
<!ELEMENT 生活費來源 (%生活費來源外的獨立類別);*>
<!ATTLIST 生活費來源
        來源 (六十五足歲以下不適用|積蓄|工作所得|配偶子女|其他親屬朋友|退休金保險金|社會救助|其他)
#REQUIRED
        %count;>
<!ELEMENT 身體狀況 (有不良狀況*|無不良狀況*|%身體狀況外的獨立類別);>

```

```

<!ATTLIST 身體狀況
    是否有不良狀況 (是|否) #REQUIRED
    %count;>
<!ELEMENT 有不良狀況 (%身體狀況外的獨立類別);>
<!ATTLIST 有不良狀況
    不良種類 (多重障礙|視覺障礙|聽覺障礙|語言障礙|肢體障礙|智能障礙|重要器官失去功能|顏面傷殘|植
    物人|老年痴呆|自閉症|無自顧能力|其他) #REQUIRED
    %count;>
<!ELEMENT 無不良狀況 (%身體狀況外的獨立類別);>
<!ATTLIST 無不良狀況
    不良種類 (無不良狀況) #REQUIRED
    %count;>
<!ELEMENT 勞動力狀況 (勞動力*|非勞動力*|%勞動力狀況外的獨立類別);>
<!ATTLIST 勞動力狀況
    是否為勞動力 (十五足歲以下不適用|勞動力|非勞動力) #REQUIRED
    %count;>
<!ELEMENT 勞動力 (失業*|%勞動力狀況外的獨立類別);>
<!ATTLIST 勞動力
    狀態 (就業|現役軍人|失業) #REQUIRED
    %count;>
<!ELEMENT 非勞動力 (%勞動力狀況外的獨立類別);>
<!ATTLIST 非勞動力
    狀態 (求學|料理家務|殘障|久病衰老|監禁|收容|不願工作|其他) #REQUIRED
    %count;>
<!ELEMENT 失業 (%勞動力狀況外的獨立類別);>
<!ATTLIST 失業
    狀態 (初尋工作|重尋工作) #REQUIRED
    %count;>
<!ELEMENT 教育程度 (學前教育及其他*|初等教育*|中等教育*|高等教育*|不識字+|自修+|%教育程度外的獨立
    類別);>
<!ATTLIST 教育程度
    程度 (不識字|自修|學前教育及其他|初等教育|中等教育|高等教育) #REQUIRED
    %count;>
<!ELEMENT 不識字 (%教育程度外的獨立類別);>
<!ATTLIST 不識字
    學歷 CDATA #FIXED "不識字"
    %count;>
<!ELEMENT 自修 (%教育程度外的獨立類別);>
<!ATTLIST 自修
    學歷 CDATA #FIXED "自修"
    %count;>
<!ELEMENT 初等教育 (%教育程度外的獨立類別);>
<!ATTLIST 初等教育
    學歷 CDATA #FIXED "國小"
    %count;>
<!ELEMENT 中等教育 (高職*|%教育程度外的獨立類別);>
<!ATTLIST 中等教育
    學歷 (國中|高中|高職) #REQUIRED
    %count;>
<!ELEMENT 高職 (教育類*|藝術類*|商業及管理類*|醫藥衛生類*|工業技藝類*|工程類*|建築及都市規劃類*|農
    林漁牧類*|家政類*|運輸通信類*|觀光服務類*|大眾傳播類*|其他學類*
    +%教育程度外的獨立類別);>
<!ATTLIST 高職
    科系類別 (教育類|藝術類|商業及管理類|醫藥衛生類|工業技藝類|工程類|建築及都市規劃類|農林漁牧類
    |家政類|運輸通信類|觀光服務類|大眾傳播類|其他學類) #REQUIRED
    %count;>

```

```

<!ELEMENT 高等教育 (專科*|大學*|研究所*|%教育程度外的獨立類別);>
<!ATTLIST 高等教育
    學歷 (專科|大學|研究所) #REQUIRED
    %count;>
<!ELEMENT 專科 (人文類*|大眾傳播類*|工程類*|工業技藝類*|自然科類*|其他學類*|法律類*|建築及都市規劃
類*|家政類*|商業及管理類*|教育類*|經社及心理類*|農林漁牧類*|運輸通信類*|數學及電算機類*|醫藥衛生類*|
藝術類*|觀光服務類*
    |%教育程度外的獨立類別);>
<!ATTLIST 專科
    科系類別 (人文類|大眾傳播類|工程類|工業技藝類|自然科類|其他學類|法律類|建築及都市規劃類|家政
類|商業及管理類|教育類|經社及心理類|農林漁牧類|運輸通信類|數學及電算機類|醫藥衛生類|藝術類|觀光服務類)
#REQUIRED
    %count;>
<!ELEMENT 大學 (人文類*|大眾傳播類*|工程類*|工業技藝類*|自然科類*|其他學類*|法律類*|建築及都市規
劃類*|家政類*|商業及管理類*|教育類*|經社及心理類*|農林漁牧類*|運輸通信類*|數學及電算機類*|醫藥衛生類
*|藝術類*|觀光服務類*
    |%教育程度外的獨立類別);>
<!ATTLIST 大學
    科系類別 (人文類|大眾傳播類|工程類|工業技藝類|自然科類|其他學類|法律類|建築及都市規劃類|家政
類|商業及管理類|教育類|經社及心理類|農林漁牧類|運輸通信類|數學及電算機類|醫藥衛生類|藝術類|觀光服務類)
#REQUIRED
    %count;>
<!ELEMENT 研究所 (人文類*|大眾傳播類*|工程類*|工業技藝類*|自然科類*|其他學類*|法律類*|建築及都市規
劃類*|家政類*|商業及管理類*|教育類*|經社及心理類*|農林漁牧類*|運輸通信類*|數學及電算機類*|醫藥衛生類
*|藝術類*
    |%教育程度外的獨立類別);>
<!ATTLIST 研究所
    科系類別 (人文類|大眾傳播類|工程類|工業技藝類|自然科類|其他學類|法律類|建築及都市規劃類|家政
類|商業及管理類|教育類|經社及心理類|農林漁牧類|運輸通信類|數學及電算機類|醫藥衛生類|藝術類)
#REQUIRED
    %count;>
<!ELEMENT 學前教育及其他 (%教育程度外的獨立類別);>
<!ATTLIST 學前教育及其他
    學歷 (托兒所|幼稚園|嬰兒及其他) #REQUIRED
    %count;>
    <!ENTITY % 科系別 "科系別 CDATA #REQUIRED">
    <!ELEMENT 人文類 (%教育程度外的獨立類別);>
    <!ATTLIST 人文類
        %科系別; %count;>
    <!ELEMENT 自然科類 (%教育程度外的獨立類別);>
    <!ATTLIST 自然科類
        %科系別; %count;>
    <!ELEMENT 法律類 (%教育程度外的獨立類別);>
    <!ATTLIST 法律類
        %科系別; %count;>
    <!ELEMENT 經社及心理類 (%教育程度外的獨立類別);>
    <!ATTLIST 經社及心理類
        %科系別; %count;>
    <!ELEMENT 數學及電算機類 (%教育程度外的獨立類別);>
    <!ATTLIST 數學及電算機類
        %科系別; %count;>
    <!ELEMENT 教育類 (%教育程度外的獨立類別);>
    <!ATTLIST 教育類
        %科系別; %count;>
    <!ELEMENT 藝術類 (%教育程度外的獨立類別);>
    <!ATTLIST 藝術類

```

```

%科系別; %count;>
<!ELEMENT 商業及管理類 (%教育程度外的獨立類別);>
<!ATTLIST 商業及管理類
%科系別; %count;>
<!ELEMENT 醫藥衛生類 (%教育程度外的獨立類別);>
<!ATTLIST 醫藥衛生類
%科系別; %count;>
<!ELEMENT 工業技藝類 (%教育程度外的獨立類別);>
<!ATTLIST 工業技藝類
%科系別; %count;>
<!ELEMENT 工程類 (%教育程度外的獨立類別);>
<!ATTLIST 工程類
%科系別; %count;>
<!ELEMENT 建築及都市規劃類 (%教育程度外的獨立類別);>
<!ATTLIST 建築及都市規劃類
%科系別; %count;>
<!ELEMENT 農林漁牧類 (%教育程度外的獨立類別);>
<!ATTLIST 農林漁牧類
%科系別; %count;>
<!ELEMENT 家政類 (%教育程度外的獨立類別);>
<!ATTLIST 家政類
%科系別; %count;>
<!ELEMENT 運輸通信類 (%教育程度外的獨立類別);>
<!ATTLIST 運輸通信類
%科系別; %count;>
<!ELEMENT 觀光服務類 (%教育程度外的獨立類別);>
<!ATTLIST 觀光服務類
%科系別; %count;>
<!ELEMENT 大眾傳播類 (%教育程度外的獨立類別);>
<!ATTLIST 大眾傳播類
%科系別; %count;>
<!ELEMENT 其他學類 (%教育程度外的獨立類別);>
<!ATTLIST 其他學類
%科系別; %count;>
<!ELEMENT 籍別 (本國籍*|外國籍*|%籍別外的獨立類別);>
<!ATTLIST 籍別
    國籍 (本國籍|外國籍) #REQUIRED
    %count;>
<!ELEMENT 本國籍 (台閩地區*|他省市*|%籍別外的獨立類別);>
<!ATTLIST 本國籍
    地區 CDATA #REQUIRED
    %count;>
<!ELEMENT 外國籍 (%籍別外的獨立類別);>
<!ATTLIST 外國籍
    地區 (美國|日本|東南亞地區|其他國家) #REQUIRED
    %count;>
<!ELEMENT 台閩地區 (%籍別外的獨立類別);>
<!ATTLIST 台閩地區
    縣市 CDATA #REQUIRED
    %count;>
<!ELEMENT 他省市 (%籍別外的獨立類別);>
<!ATTLIST 他省市
    省市 CDATA #REQUIRED
    %count;>

```


2. displaydata.dtd

```
<?xml version="1.0" encoding="UTF-16"?>
<!ELEMENT TSM (docdesc,def,mapLayer,mapShow,region+)>
<!ELEMENT docdesc (dataSource)>
  <!ELEMENT dataSource (source+)>
    <!ELEMENT source (#PCDATA)>
  <!ELEMENT def (datadef|groupdef)*>
<!ELEMENT groupdef (datadef+)>
  <!ATTLIST groupdef
    type CDATA #REQUIRED
    name CDATA #REQUIRED
    width CDATA #REQUIRED
    desc CDATA #REQUIRED
    unit CDATA #IMPLIED
  >
<!ELEMENT datadef EMPTY>
  <!ATTLIST datadef
    type CDATA #REQUIRED
    name CDATA #REQUIRED
    width CDATA #IMPLIED
    desc CDATA #REQUIRED
    unit CDATA #IMPLIED
    expression CDATA #IMPLIED
  >
<!ELEMENT mapLayer (layer+)>
<!ELEMENT layer EMPTY>
  <!ATTLIST layer
    name CDATA #REQUIRED
    desc CDATA #REQUIRED
  >
<!ELEMENT mapShow (show+)>
<!ELEMENT show EMPTY>
  <!ATTLIST show
    name CDATA #REQUIRED
    decimal CDATA #REQUIRED
    desc CDATA #REQUIRED
  >
<!ELEMENT region (data*|group*)*>
  <!ATTLIST region
    type CDATA #REQUIRED
    code CDATA #REQUIRED
  >
<!ELEMENT group (data+)>
  <!ATTLIST group
    name CDATA #REQUIRED
  >
<!ELEMENT data (#PCDATA)>
  <!ATTLIST data
    name CDATA #REQUIRED
    value CDATA #IMPLIED
```

3.本系統之架構與原始程式說明

主要程式：

1. tsm_query.php

本檔案為單純的 html 檔案，提供使用者經由表格 (from) 自訂條件以建立普查資料集 (censusdata) 的 XML 文件。

2. tsm_censusxmlmaker.php, tsm_censusxmlmaker_m1.php

此二檔均用於建立普查資料集，前者為判斷是否有暫存檔案，後者則實踐建立 XML 文件。

在程式設計上我們採用 PHP DOM 的函數庫 (請參見

<http://devil.medialab.at/projects/phpdom/doc/phpDOM.html>)

來建立 XML 文件，惟請注意，phpDOM 採用 phpxml 的函數庫，但迄本系統所使用的 PHP 版本 (4.01) 為止，phpxml 函數

庫都仍是測試函數庫。已知可能因此而導致網路伺服器之資源被佔用的問題。本程式建立 XML DOM Tree 的方式是先

將查詢的村里資訊自 mySQL 資料庫中讀出，在根據取得的資訊內容決定必須統計的資料表格。當統計資料從資料表格

讀出後，再經由遞迴的方式，逐一建立節點 (node)。

3. tsm_censusdata2displaydataxmlmaker.php, tsm_censusdata2displaydataxmlmaker_m1.php

本檔案之目的在動態建立可將普查資料集之 XML 轉換為顯示用 XML 之 XSLT 檔案。由於我們在進行轉換時需得知普查資

料集 XML 之最底層結構與其各節點之屬性值，因此必須動態產生 XSLT 檔案。

4. tsm_censusxmlresult.php

顯示普查資料集 XML 之 HTML 畫面

5. tsm_censusxml2displayxml.php

本檔案將呼叫外部程式 xt，以將普查資料集之 XML 透過 XSLT 轉換成顯示用 XML。

6. tsm_displayxmlresult.php

顯示顯示用 XML 之 HTML 畫面

7. tsm_displayxml2displayjs.php

由於本系統最終是將資訊以 HTML+SCRIPT 的方式傳送至使用者端以瀏覽器介面呈現。本檔案因此呼叫外部程式 xt，以

將轉換成顯示用 XML 透過 XSLT 轉換成 SCRIPT 檔案。

8. tsm_displayjsresult.php

縣市在所有轉換過程中的原始與結果檔案

9. tsm_display.php

本檔案為最後顯示於使用者端之 HTML 檔案。在使用者端進行的各種功能，均於此檔案及其包含之檔案中描述。

這些檔案被分置於幾個目錄中：

/image 圖像檔案 (略)

/svg 空白之 svg 定義 (略)

/script 顯示用 script

domtool.js 常用的 DOM 函數庫

graph.js 繪製圖表用之函數庫

mapdraw.js 繪製地圖之函數庫

tabledraw.js 繪製表格之函數庫

mouseevent.js 處理滑鼠動作之函數庫

scroll.js 處理表格拖曳之函數庫

selectdraw.js 處理顯示選單之函數庫

(註：以上兩個函數庫參照 dynapi，參見 <http://dynapi.sourceforge.net/dynapi/>)

svgdraw.js 雜項函數庫

/userdata 透過系統建立之 XML、DTD、XSLT、SVG 與 JS 檔案

censusdata.dtd 普查資料集之 DTD

displaydata.dtd 顯示用 XML 之 DTD

map*.svg SVG 地圖，* 表示行政區域編號 (以下同)

censusdata2displaydata*_?.xml 轉換普查資料集至顯示資料集之 XSLT 檔，??表示查詢條件（以下同）
censusdata*_?.xml 普查資料集
displaydata*_?.xml 顯示資料集
displaydata2js.xml 轉換顯示資料集至顯示用 script 之 XSLT 檔
data*_?.js 顯示用 script 檔案

其他檔案

1. createsvgfile.php

本檔案用於建立地圖 SVG

2. transfer.java

從最原始資料中抽出個人資料集與轉換行政區域代碼之程式