

以功能為導向之分散式 WebGIS 架構

彭逸帆

中央研究院計算中心

台北市南港區研究院路二段 128 號
886-2-27898038

yfpeng@gate.sinica.edu.tw

廖泫銘

中央研究院計算中心

台北市南港區研究院路二段 128 號
886-2-27898012

veevee@gate.sinica.edu.tw

范毅軍

中央研究院歷史語言所

台北市南港區研究院路二段 128 號
886-2-265231848

mhfanbbc@ccvax.sinica.edu.tw

摘要

近幾年來，國內各相關單位紛紛且積極投入數位典藏計畫的執行工作，迄今已累積了相當龐大的數位化資料量；如此大量的典藏資訊必須藉由各種不同的系統工具，以清楚且完整地呈現其建置後的成果。觀之現今各種不同的呈現方式，以地理資訊系統（Geographic Information System, GIS）可稱為係一較具親和力且深具發展潛力的新興服務介面。然而，在各單位建置 GIS 的過程中，往往是針對單一主題即建置一套單一的對應系統，此一做法對於各種不同計畫但可重複使用同一套圖資的部份而言，不僅既有的資源無法達成有效地應用，同時，各單位各自建置自己的典藏系統，無法套用及相互交流，則其所耗費之金錢、時間與人力成本等，等同於是一種重覆性的投入與浪費，相當不經濟。且如此，亦有違數位時代的資訊特性與推動典藏計畫概念的本意。

本論文即是以中央研究院計算中心空間資訊技術小組為例，介紹 GIS 小組對於 WebGIS 所採取之功能性分散式處理架構作法，以達到有效降低主機負載、提升網站執行效率，以及重複使用已建構完成之典藏資訊於多種計畫中之規畫與解決方案。

關鍵字

GIS、分散式 WebGIS

壹、前言

數位典藏國家型科技計畫自民國九十一年起執行，至今，各個單位所累積之數位化典藏資料量已頗具相當規模，接踵而來所面臨的問題，即是如何將典藏建置的成果具體呈現於使用者面前。目前，一般常見之呈現方式多為單純之表單結構，藉由查詢資料庫的型式來展現其建置內容。亦即，由使用者輸入相關之資訊，經由查詢資料庫的作法，以搜尋出使用者感興趣之結果。

而如果典藏之物品或資料與時間或空間產生關聯，那麼，地理資訊系統則是另一種合適的檢索介

面。此一方法不僅可以方便地讓使用者藉由選取地理位置的方式搜尋出結果；除此之外，並可將搜尋出之結果連結至更多的資訊，而後再回傳給使用者。這對於使用者來說，相同的付出卻可以得到同等、甚至更多、更廣的資訊回報，不啻為一個可簡單瞭解典藏成果的友善推廣服務介面。

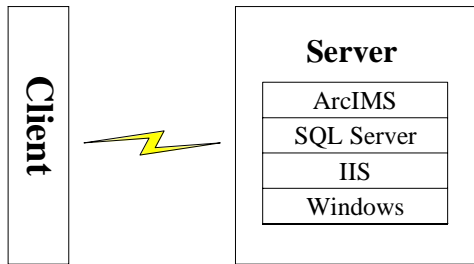
然而，如同我們所瞭解，在一般已建構的網際網路地理資訊系統(WebGIS)中，其作法往往是單一計畫即建置一套單一的對應系統，分屬不同計畫，即需分別建置多套不同的服務系統，亦即，對於不同計畫但相同的資訊(如：台灣行政區界...等)往往都需要重複建置，以致造成資料量日益浩繁，不僅使得系統負荷相對擴大，同時亦提高了資訊檢索的困難度，落入資訊超載以致無法妥善運用的困境。這些都是沒有做到資源重複使用的例子。

為避免上述問題產生，中央研究院計算中心空間資訊技術小組採行有別於一般單一計畫即建置單一系統的方式，採取以功能為導向的分散式架構，來建構網際網路地理資訊系統，以達到資訊之可交換性與重複使用性，同時可以有效降低單一主機因為需要執行多項圖資處理工作所產生之資訊負載。

本文章第一部份為前言，第二部份係簡介一般集中式 WebGIS 架構方式，第三部份則說明目前中央研究院計算中心空間資訊技術小組所採行之 WebGIS 架構方式，第四部份主要針對集中式架構與改良架構進行實驗比較並分析其差異，第五部份為結論與未來展望。

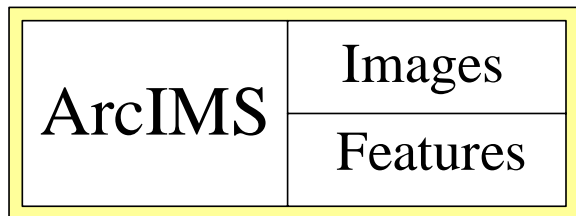
貳、集中式 WebGIS 架構方式

傳統上，一般對於數位典藏成果的展示，是透過網頁、資料庫、WebGIS 等作法，將典藏成果呈現在使用者面前，如【圖一】所示：



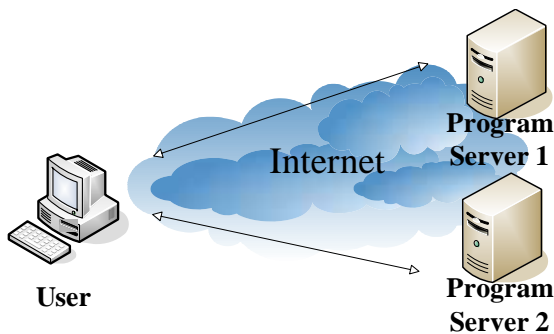
圖一 傳統數位典藏成果的展示方式

而 WebGIS 的方式，則是採用 ESRI 公司所設計之 ArcIMS[1]軟體，其中包含了處理影像(Images)及圖徵(Features)等功能，架構如【圖二】所示：



圖二 WebGIS 呈現的方式

在過去，對於不同的計畫或是研究主題，即便是彼此間有使用到相同的圖資，在一般的 WebGIS 系統架構中，仍是無可避免地必須在每一個伺服器主機裡重複建置相同的資訊，其關係圖如【圖三】所示。此一作法對於資訊的重複使用性而言，不僅無法適時發揮典藏的效益，且顯然是對資訊資源造成相當大的浪費，殊不經濟。



圖三 傳統的 WebGIS 系統架構

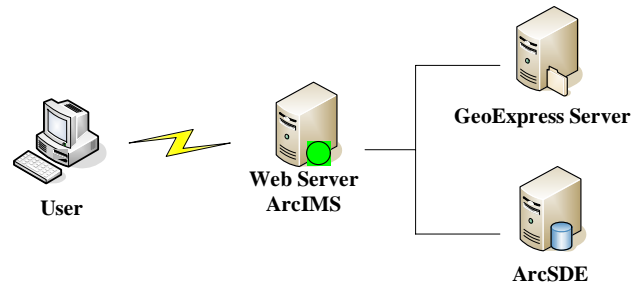
除此之外，由於是將所有的運算皆放置於單一主機之上，故而，一旦當所需呈現的典藏內容資料量相當龐大時，主機的運轉與效能則可能因此大打折扣。換言之，主機運轉本身即需要耗費相當大的運算能力，如此一來，連帶地就會影響主機處理使用者要求

的速率，可以想見其結果自然是大幅降低整體服務的效果。

為有效改善此一缺失，以功能為導向之分散式 WebGIS 架構則可謂是一套相當優異的解決方案。以下茲解釋分析之。

參、功能為導向之分散式 WebGIS 架構

所謂功能導向之分散式 WebGIS 架構，即是針對前述【圖二】中 ArcIMS 的部份做出一些修改；也就是將原本處理 Images 及 Features 的部份獨立出來，交由其他的伺服器作處理，其作法如【圖四】所示：



圖四 功能導向之分散式 WebGIS 架構圖

使用此種分散式的架構，不但可以輕易達到資料的重複利用，同時，對於主機負載，也能有效地分散到其他伺服器上，大幅提升機器的處理速率。

3.1 伺服系統簡介

3.1.1 Lizardtech GeoExpress Server[2]

此一工具為 Lizardtech 公司所設計之空間伺服器軟體。由於該軟體符合 Open Geospatial Consortium(以下簡稱 OGC¹)[3]的規範，對於影像資訊也能有效處理，因此，只要您的主機或裝有 GeoExpress Server 的主機透過 Web Map Service (WMS) 的方式，即可於其上進行所需的資訊交換。

GeoExpress Server 所提供的服務與 WebGIS、ArcIMS 之間溝通，主要是藉由 Lizardtech 所開發的 Express Server Plugin for ArcIMS 軟體來進行所需影像檔案之交換。

3.1.2 ESRI ArcSDE

ArcSDE(Spatial Database Engine)為 ESRI 公司所生產之關連性資料庫管理系統。一般在數化圖資時，可能有時所產生的檔案高達數百甚至上千個百萬位元組

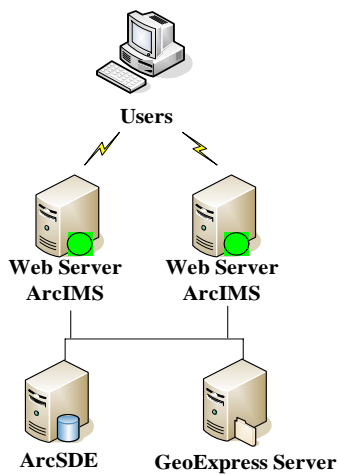
¹ OGC 是業界所支持的「開放式空間資訊協會」，該協會藉由制訂一系列的標準將地理資訊系統與空間資料整合，搭配網際網路與開放式的協定而達成跨平台之地理資訊系統。

(Mega Byte)。如此大量的資料如果直接載入記憶體執行，往往會耗費掉整個系統資源，甚至會因為圖資量過大，造成系統不穩定或當機之後果；而使用類似 ArcSDE 之類的管理系統來進行處理相同的資訊，我們可以先針對資料量大的圖資做前處理，之後欲呈現時，再局部載入使用者所感興趣的部份，如此，則不會造成主機系統整個被大批資料『綁架』的窘境，進而達到有效提升系統使用效率的目的。

而 ArcSDE 所提供的服務與 WebGIS、ArcIMS 之間的溝通，則是直接透過 ArcIMS 配合特定之帳號密碼，即可方便地將使用者所需之圖資在兩者之間相互傳遞。

3.2 圖資之重複使用

在採行分散式之架構後，對於不同的主題或是計畫，使用者可以將圖資分別儲存於 GeoExpress Server 與 ArcSDE 等伺服器中，如需要使用時，再由主題網頁的伺服器向上述的兩種伺服器進行存取即可，其方式如【圖五】所示：



圖五 分散式架構之存取方式

舉例來說，在臺灣原住民數位典藏計畫中，原住民文化地圖的底圖是採用台灣地區的縣市界（請參見【圖六】）[4]；而台灣貝類地理分佈查詢系統的底圖，同樣地也是採用台灣地區的縣市界（請詳見【圖七】）[5]。



圖六 臺灣原住民數位典藏計畫所採用之台灣地區縣市界底圖



圖七 台灣貝類地理分佈查詢系統同樣採用台灣地區的縣市界為底圖

在此，我們可以很明顯地發現，如依本文主旨所闡釋的作法，則一個相同的圖資僅需在 ArcSDE 中建置一套，當不同的主題或計畫需要使用到該圖資時，即可透過連結伺服器的方式以取得的該資料，藉此達到圖資之重複使用的目的。

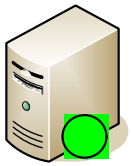
為進一步說明其間的差異，以下茲以一個簡單的實驗暨其所獲得之數據分析，清楚呈現兩者不同的結果。

肆、實驗與結果

為比較集中式架構與分散式架構的差異性，以下針對不同的資料(分別包含 Images 與 Features)進行實驗 ESRI ArcSDE 測試如下：

傳統集中式的架構，即是指在同一部機器上負責 Web Server 與 ArcIMS 的執行，以 WebGIS 的方式來呈現圖資的成果；而分散式的架構，則是將原本全部由 ArcIMS 負責處理 Images 與 Features 的部份區隔出來，由其他的機器搭配 GeoExpress 與 ArcSDE 來負責處理。

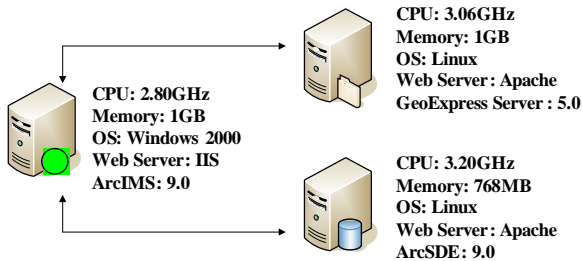
在實際比較不同的架構時，我們將實驗分成兩組，其中【實驗一】為傳統集中式的架構，系統的配備如【圖八】所示；此一架構乃是將所有 WebGIS 的功能全部集中於同一部伺服器，由此一伺服器同時負責對使用者的網頁與 WebGIS 功能做回應。



CPU: 2.80GHz
Memory: 1GB
OS: Windows 2000
Web Server: IIS
ArcIMS: 9.0

圖八 集中式系統配備

而【實驗二】則是採行分散式的架構，系統的配備如【圖九】所示。在此架構中，我們將原本由同一伺服器負責處理 Features 與 Images 的部份獨立出來，將該兩部份分別交由另外兩部裝有 ArcSDE 與 Express Server 的伺服器做處理；也就是如【圖九】中，右半部兩台伺服器所肩负之功能。

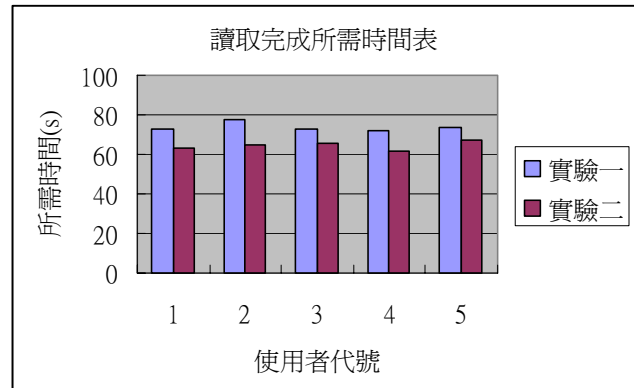


圖九 分散式系統配備

4.1 ESRI ArcSDE 測試

此部份我們針對全台灣的道路 Features 檔案作一測試，這個檔案為全台灣包含國道、省道、縣道、一般

道路、...等所有的道路路線共 266,045 筆資料，我們利用不同方式比較讀取完所有道路路線所需的時間，比較的結果如【表一】所示：



表一 ESRI ArcSDE 測試結果

在【表一】中的【實驗一】，是表示採用一般集中式的方法在 WebGIS 上進行呈現；而【表一】中【實驗二】，則是表示採用分散的方式，將 Features 的部份交由 ArcSDE 處理。

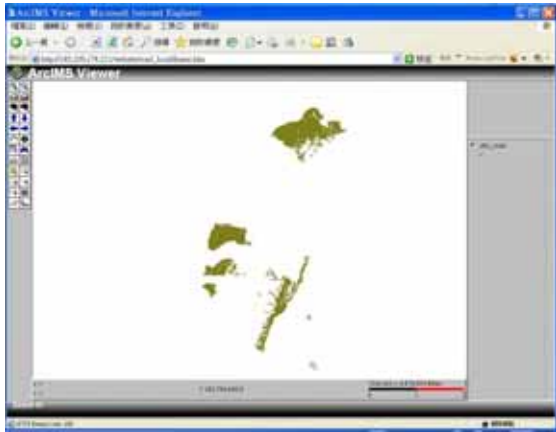
由以上的實驗結果比較，我們可以很清楚地看到，傳統集中式的方法在呈現運算結果時，伺服器端不但需要負責 Web Server 的工作，此外也必須負擔圖資處理的動作；此一情形如果只是在少數使用者連線的情況下，這樣的架構也許還足敷因應；但是假設同時有大量的使用者提出連線要求時，可想而知，此一架構在伺服器端會因為負載過重，造成伺服器回應使用者要求的時間相對變長。如此一來，將置使用者於冗長的等待中，造成使用者時間的浪費；然而此一情況在分散式的架構中將不致發生，乃是由於分散式架構讓單獨的主機處理單獨的工作，降低單一主機的資訊負載量，使得效能可以有效提升。

以本實驗為例，當使用者對伺服器提出瀏覽全台灣道路的要求時，前端的 Web Server 與 ArcIMS 可以在接受到此一請求時，將此一請求轉而由後端的 ArcSDE 進行處理；當 ArcSDE 得到結果時，再藉由 ArcIMS 呈現出其得到的結果。此一方式對前端的伺服器而言，主要的工作即是在接受與呈現使用者所提出的要求，所以同時所能承受的連線人數必定會高於傳統使用的集中式架構；再者，又因為 ArcSDE 本身就是專門負責圖資的儲存與搜尋，在處理使用者的要求時，其效能又會比單獨使用 ArcIMS 為好。

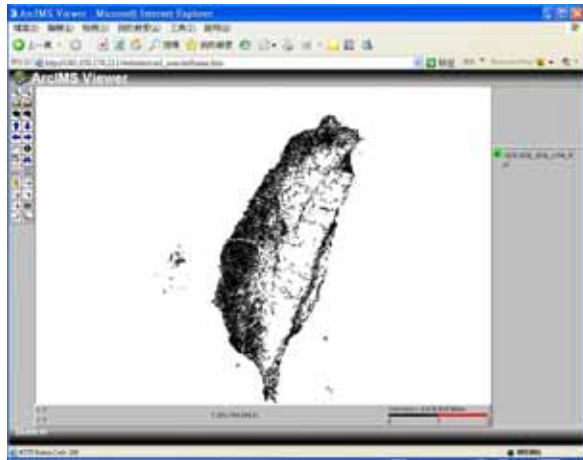
至此，根據上述的結果顯見，不論是由理論推導或是從實驗數據，都可以得到相同的結果，亦即：分散式的架構會比集中式的架構效能還要好。

除此之外，採用分散式的架構尚有傳統集中式架構所無法達成的效果。在下列的【圖十】與【圖十

一】中所呈現的，是在網頁上呈現圖資內容載入後第 30 秒載入的情況：



圖十 傳統集中式架構中程式載入圖資第 30 秒之載入情況



圖十一 分散式架構中程式載入圖資第 30 秒之載入情況

其中，【圖十】是一般傳統使用的集中式架構。在此架構中，由於 ArcIMS 不能針對圖資進行整體的分析，所以在顯示內容時，僅能局部地緩慢載入圖資。換言之，如果想要得到圖資內容概略的情況是無法做到，一定要等到全部圖資載入後，才可得到結果；而【圖十一】則是將處理圖資的部份交由 ArcSDE 負責處理，前端則是專司接受使用者的請求與網頁的呈現。由於 ArcSDE 對於圖資優異的處理效能，針對使用者的請求，可以先針對要求的內容進行處理，所以能夠在短時間就先載入大部份的圖資。此一優點，對於一般使用者來說可謂相當便捷，因為在短時間即可得知圖資大略的情況，而不需花費冗長時間等待全部的圖資載入後，才可以得到完整結果的呈現。

4.2 GeoExpress Server 測試

此部份實驗主要是針對 WebGIS 在影像檔案呈現的時候，伺服器對於圖形處理的負載程度；至於所處理的影像檔案則是 SPOT 40m 解析度的全台灣衛星影像。

實驗方式同樣分為兩組：一組是集中式的架構，也就是將 Web Server、ArcIMS 放置於單一伺服器，由此伺服器處理從使用者發出的請求；換句話說，也就是將使用者在 WebGIS 系統上對於顯示 Images 的要求直接由 ArcIMS 單獨處理。另一組則是將使用者在 WebGIS 系統上對於 Images 的服務交由裝有 GeoExpress Server 的主機處理。換言之，當使用者對於 WebGIS 上 Images 的部份感興趣進而提出瀏覽的請求時，前端裝有 Web Server 與 ArcIMS 的主機在接收到此一要求，就會將要求轉交由後端裝有 GeoExpress Server 的主機負責處理使用者的請求；當得到結果時，再將其結果送至前端的主機，由前端的主機進行結果的呈現；而前後端兩部伺服器之間，則是透過之前提到符合 OGC 規範的 WMS 協定的方式連結。

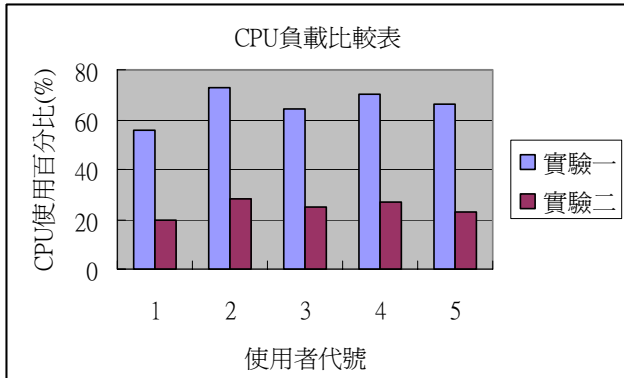
本次實驗的取樣方式，是由五組不同的機器分別操作該分散式架構的 WebGIS 系統五分鐘，於五分鐘內取樣 CPU 負載的高峰點並記錄之，將其結果整理可以得到如【表二】的結果。

由【表二】的【實驗一】可以清楚看到，單一伺服器需要擔負 WebGIS、ArcIMS 及處理影像的工作，這對於伺服器本身的負載是相當高的。除此之外，單獨以 ArcIMS 處理影像資料，該程式處理的模式是一次載入全部的影像再進行呈現，假設在使用者只有少數人，而且所需處理顯示的影像資料不大的情況下，伺服器對於資料的負載也許還可以應付；如果使用人數一旦增多或者影像的資料非常龐大，則所耗費的系統資源是可以想見。

以本次實驗的檔案為例，全台灣衛星影像檔案大約是將近 200MB 的 SID 格式檔案，當使用者想要查詢時，程式是一次全部載入約 200MB 的檔案；假設同時有十個使用者在線上進行同步的查詢檢索，那麼，系統的效能就會因為資源的耗盡，進而拖慢服務的反應時間。

而針對分散式的系統架構，由於前端伺服器專司接受使用者提出的請求與網頁結果的呈現，真正進行運算的是由後端裝載有 GeoExpress Server 的軟體，所以同時接受大量使用者的請求是有問題；除此之外，GeoExpress Server 對於處理影像的資料有其特別之處，在處理影像檔案時，並非一次就載入全部的資訊，而是針對使用者感興趣的部份進行呈現。再以全台灣衛星影像為例，也許使用者僅對於台北地區約

10MB 至 20MB 的資訊有興趣，使用者藉由圈選的方式選取台北地區，則系統僅會載入被圈選台北地區約 10MB 至 20MB 的影像資料做呈現，而並不會全部載入全台灣地區的影像資料，如此一來，既不會拖慢系統服務的反應時間，也可以同時服務到更多的使用者。



表二 GeoExpress Server 測試結果

根據【表二】呈現的數據所示，對於兩個實驗比較的結果可以很清楚地看出，【實驗一】採用以傳統集中式 WebGIS 架構建構出整個系統，使用者在操作系統時，系統的負載相當容易升高。此一情形在單一使用者的使用時就已如此，其將更不利於提供多人同時連線使用，原因在於多人同時進行連線時，將會使的系統始終處於負載偏高的狀態，而導致服務品質大幅降低。反觀【實驗二】，在採用分散式 WebGIS 架構來建構整個系統後，使用者在操作系統時，伺服器的負載明顯低於傳統集中式架構的負載，在負載降低的情況下，伺服器即可以提供更多的連線，以同時服務更多的使用者。

伍、結論與未來展望

透過上述的兩項實驗的進行，我們可以清楚獲知兩者結果的迥異，亦能明白昔日傳統集中式的作法，對於架構 WebGIS 系統往往只注重資料建立或處理過程與作法的方便性緣故，而將所有的圖資放置於同一部伺服器內，如今實驗證明，這其實並不是一種最為明智的數典資訊處理機制與方法；尤其對於需要長期展示典藏成果之主機而言，更應該要有整體而全面性的規劃，如此，才可以達到資源的重複使用，同時也可以有效降低主機的負載，進而服務更多的使用者之典藏系統建置目標；此一部份，未來仍有待大家攜手共同努力。

放眼今日，Open GIS 的系統發展有愈來愈蓬勃的趨勢，當我們在使用商業化的解決方案之時，且不論其檔案型態、乃至於資料的交換格式等，亦會參照標準的規格來進行處理，如此，將可使得不同 WebGIS 的系統其資料得以順利完成交換、達到充份地應用，以發揮資料與資訊的實際效益與邊際效能；而此一結果亦方能大致貼近並符合數位典藏計畫所追求的宏遠目標。

陸、網路參考文獻

- [1] ArcIMS
<http://support.esri.com/index.cfm?fa=software.filteredGateway&PID=16>
- [2] GeoExpress Server
<http://www.lizardtech.com/products/exp/>
- [3] OGC Specifications <http://www.opengeospatial.org/>
- [4] 臺灣原住民數位典藏計畫中原住民文化地圖
<http://gis2.sinica.edu.tw/aborigines/>
- [5] 台灣貝類地理分佈查詢系統
<http://gis2.sinica.edu.tw/website/shell/>