



# Open Data 推廣/經驗分享

顏嗣鈞 主任  
臺大計算機及資訊網路中心

# 碩博士招生明細



## 執行現況：

- 104~105學年碩博士招生資料共1300餘筆，計每年會有：碩士班甄試、碩士班招生、碩士班在職專班、EMBA、GMBA、博士甄試、博士招生等7種招生資訊。

## 遭遇問題：

- 招生名額會有流用的機制，所以不同時間去蒐集資料數字不盡相同。
- 分組招生名額共用，無法分列各組名額。
- 目前除統計數字公開資料為各考試之簡章，不易收集考科等詳細資料。

# 碩博士招生明細



## 解決方式：

- 因目前是追溯過去學年度資料，故以最後的各考試結束最後流用的數額呈現最準確的數字。
- 為維持招生總人數數字正確，分組招生人數記錄在個組別時，分組統計；分組名額為0者，取消記錄該分組資料。
- 從各招生檢查以人工或程式取出各考科資料；或結合招生系統資料庫找出各考科資料。

## 後續規劃及工作項目：

- 通常是先甄試、一般考試、再來碩士專班，資料統計時間最好是當年度碩士博士班招生都結束統計。
- 建議有統一的做法，或只統計到各研究所為單位。
- 等蒐集到各學校的資料後，宜訂定統一的格式給各學校參考使用。

# 圖書相關資料



## 執行現況：

- 已蒐集95年~105年間，相關資料共60萬筆，其中104年~105年共123956筆。進行ISBN及語言兩欄位資料值檢查及格式對應整理。

## 遭遇問題：

- 語言欄目前資料為英文縮寫表示，與教育部規劃有落差。
- 限館內參閱、推薦單位、經費來源、總借閱冊次…等4項必填欄位圖書館未提供。
- ISBN轉出資料格式數字文字半參，顯示易有錯誤。
- 出版年份資料非單純西元年(日本年號，民國年號…等)。
- 借閱統計記錄部分圖書館有其困難。

# 圖書相關資料



## 解決方式：

- 語言先以目前資料做人工對應填寫。
- 未來與圖書館討論是否能提供短少的欄位值。
- 人工逐一調整ISBN顯示格式有誤之資料。
- 因出版年份非必填欄位，先以圖書館提供資料填入。
- 與借閱統計有關資料本次先不填。

## 後續規劃及工作項目：

- 現有資料的語言類別有217種，希望教育部能針對現有語言類別所需資料訂定統一標準，以利後續資料整理。
- 必填欄位短少部分，建議安排時間與圖書館討論能否提供。

# 校園可供外借空間



## 執行現況：

- 本校可供外借的空間計有教室、會議室、宿舍、戶外展演場、戶外大型場地等類別，目前依照場館空間資訊網所整理之連結，手動整理資料，已完成總務處、體育館、文理工農各院等較多場地的單位共計100多筆資料。

## 遭遇問題：

- 由於本校外借空間並無統一管理單位，無法有現有資料庫取得資料。
- 場地描述並無統一的規範及欄位，雖收集的欄位眾多，只有幾個欄位是所有資料都有的。收費的項目種類也眾多不適合以資料庫整理。
- 場地資料沒有歷史資料，目前並沒有辦法查詢 104~105學年的場地外借情形。



# 校園可供外借空間



## 解決方式：

- 依法場地收費都會訂定收費辦法，只能從網頁或收費辦法人工整理出各可供外借場地的資料。
- 依照填表說明如果有其他收費項目或方法或規定可填在備註欄，造成每筆資料的備註欄都很大。
- 統計目前的場地資料為105學年度資料，並無104學年資料。

## 後續規劃及工作項目：

- 訂定合適的欄位後下放到各單位填寫，應該會收錄到比較多且即時的資料。但缺乏誘因讓校內各單位執行此動作。
- 場地資料雖大多公開在網路上，但是否適合彙整成公開資料？會不會有引發其他問題的疑慮，如學校為非營利單位場地對外收費稅務的問題，學校兼營旅館僅能提供學術交流活動投宿，否則會衍生牌照的問題會一一浮現。

# 校內(外)賃居資訊



## 執行現況:

- 已確認校內學生宿舍範圍(共21棟)、體育館、文理工農各院等較多場地的單位共計100多筆資料，正進行各欄位判讀對應分析。

## 遭遇問題:

- 各宿舍僅統計房間數與單一房間床位數量，無明確宿舍總床數。
- 宿舍冷氣卡及用水電資料僅於各宿舍簡介中描述，無統一呈現方式。
- 缺乏校外租屋資訊
  - 系統資料主要是提供暑假短期住宿的資訊。
  - 學務處是在暑期宿舍不夠分配時，才會派人進行前往調查蒐集資料。
  - 資料僅為當學年暑假使用，屬特定期間才有效的資料。(住宿組表示目前系統資料已經有一段時間都沒有更新)。



# 校內(外)賃居資訊



## 解決方式:

- 各別人工整理出宿舍每層樓及每棟房間總數，配合其床位數逐一計算。
- 電話或個別瀏覽宿舍網頁說明，逐一記載冷氣及水電狀況。
- 目前無法提供校外賃居資料。

## 後續規劃及工作項目:

- 若後續資料提供為常態，需與住宿組討論相關欄位是否可照教育部格式提供。

# 課程表



## 執行現況：

- 已蒐集104學年~105學年共計52871筆課程資料，正進行各欄位判讀對應分析

## 遭遇問題：

- 星期、開始時間、結束時間，教務處資料儲存方式不同，無法直接對應。
- 通識課程資料存於其他獨立資料庫。
- 授課語言目前僅在備註欄描述。

# 課程表



## 解決方式：

- 授課星期與時間目前以人工對應方式逐一進行補齊資料。
- 人工查詢通識課程分類，並對現有資料進行處理。
- 語言先以預設值填寫，同時提供備註欄內容。

## 後續規劃及工作項目：

- 若資料提供為常態，需與教務處資訊組研議並請其提供。
- 教育部應參考各校資料內容，將資料格式調整更符合各校現行狀況。

# 校行事曆



## 執行現況：

- 已整理台大104學年~105學年校行事曆資料。

## 遭遇問題：

- 現有校行事曆只提供活動時間、辦理事項兩項欄位，不易篩選資料內容。

# 校行事曆



## 解決方式：

- 以人工整理並新增：**性質、類別、對象**欄位，以及對應之分類。例如：
  - **性質**：「行政」、「學籍」、「考試」、「課程」、「成績」、「活動」、「假日」、「招生」...等
  - **類別**：
    - 行政：「學期開始/結束」、「上課開始/結束」、「補上班/補課」、「寒假/暑假」、「會議」...等
    - 學籍：「註冊(含繳費/退費)」、「休學/退學」、「學位證書申請」...等
    - ...
  - **對象**：「教職員生」、「教師」、「師生」、「學生」、「學士班」、「碩/博班」、「國際學生/僑生/陸生」、「新生」、「教學助理」、「校友」...等

## 後續規劃及工作項目：

- 若資料提供為常態，需與教務處研議並請其補充資料。