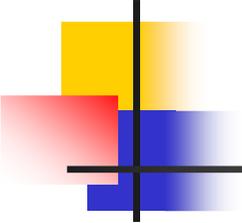


# 以XML資料庫系統建置後設資料儲存庫

---

王祥安、黃建中、林彥君  
國立台灣科技大學資訊工程系

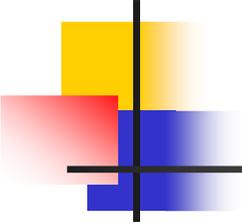
報告者 黃建中



# 大綱

---

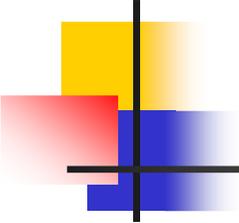
- 簡介
  - 後設資料
  - 後設資料規格書
  - 後設資料儲存庫系統
  - XML
- 以XML呈現後設資料規格
- XML文件的儲存方式
- 後設資料儲存庫系統
- 結論



# 後設資料(Metadata)

---

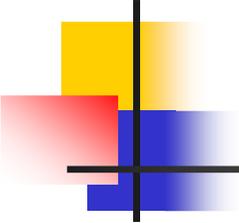
- 描述資料的資料(Data about data)。
- 描述典藏文物之內容及特徵，有助於檢索及管理典藏文物。
- 後設資料的分析為開發數位典藏系統之基礎資訊建設工作，其分析結果會影響典藏資料的完整性，及後續系統的建置。



## 後設資料規格書 (一)

---

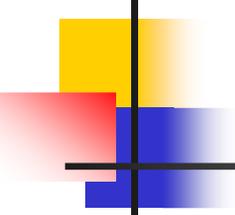
- 用來呈現後設資料的分析結果，可供典藏單位確認其後設資料需求。
- 作為典藏單位、後設資料分析人員、系統開發人員三方溝通的橋樑。



## 後設資料規格書 (二)

---

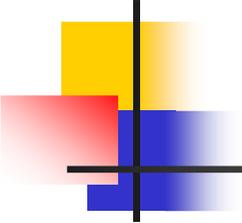
- 在目前的後設資料規格分析過程中，大多以人工的方式處理，並沒有太多軟體工具可以輔助。製作一份規格書需花費大量時間與人力。
- 製作完成的規格書多以紙本或電子檔案的方式儲存，對於後續查詢與再利用會受到限制。



# 後設資料儲存庫系統

---

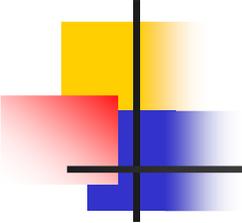
- 儲存過去已發展的後設資料規格及國際後設資料標準。
- 提供後設資料規格的查詢與製作機制，可快速地查詢與建立後設資料規格。
- 預期可減少後設資料規格書之製作時間與人力。



# XML

---

- 由W3C所制定的標記語言。
- 具有可自定標籤(tag)、可擴充性、易於程式處理等特性，常用來呈現結構複雜的資料。



# 以XML呈現後設資料規格

---

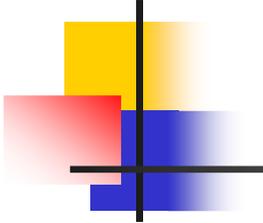
- 參考中研院後設資料工作組所製作的後設資料需求規格書，以XML呈現典藏計畫之後設資料規格。
- 參考中研院後設資料工作組對Dublin Core、CDWA 的翻譯，將其內容以XML呈現。

# 漢代簡牘後設資料 需求規格書部份內容

| 項目名稱 | 英文名稱  | 資料型態       | 大小                |         |     |
|------|-------|------------|-------------------|---------|-----|
| 基本資料 |       |            |                   |         |     |
| 識別資料 | 遺物登錄號 | Identifier | Registered Number | Varchar | 30  |
|      | 簡號    |            | Pek. Coll.        | Varchar | 200 |

| 項目名稱 | 必填    | 多值 | 屬性 | 提供者 |
|------|-------|----|----|-----|
| 基本資料 |       |    |    |     |
| 識別資料 | 遺物登錄號 | *  | 唯一 | 填表者 |
|      | 簡號    |    | 唯一 | 填表者 |

| 項目名稱 | Dublin Core Elements |            |
|------|----------------------|------------|
| 基本資料 |                      |            |
| 識別資料 | 遺物登錄號                | Identifier |
|      | 簡號                   | Identifier |



# 以XML呈現 漢代簡牘後設資料規格

```
<?xml version="1.0" encoding="BIG5" ?>
```

```
- <計畫名稱 name="漢代簡牘">
```

```
- <基本資料>
```

```
- <識別資料 英文名稱="Identifier" 必填="*" 多值=" ">
```

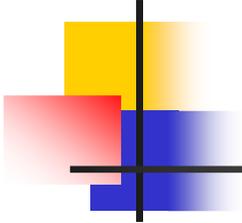
```
  <遺物登錄號 英文名稱="Registered Number" 資料型態="Varchar" 大小="30" 必填=" " 多值=" " 屬性="唯一" 提供者="填表者" DC-Element="Identifier"> </遺物登錄號>
```

```
  <簡號 英文名稱="Pek. Coll." 資料型態="Varchar" 大小="200" 必填=" " 多值=" " 屬性="唯一" 提供者="填表者" DC-Element="Identifier"> </簡號>
```

```
  </識別資料>
```

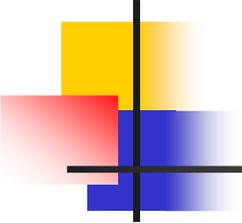
```
  </基本資料>
```

```
</計畫名稱>
```



# 以XML呈現 Dublin Core基本元素

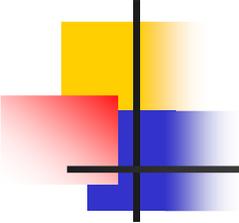
```
<?xml version="1.0" encoding="big5" ?>
- <標準名稱 name="Dublin Core">
- <DC基本元素>
  <標題 英文名稱="Title" />
  <著作者 英文名稱="Creator" />
  <主題-關鍵字 英文名稱="Subject and Keywords" />
  <描述 英文名稱="Description" />
  <出版者 英文名稱="Publisher" />
  <貢獻者 英文名稱="Contributor" />
  <日期 英文名稱="Date" />
  <資料類型 英文名稱="Resource Type" />
  <格式 英文名稱="Format" />
  <資料識別 英文名稱="Resource Identifier" />
  <來源 英文名稱="Source" />
  <語言 英文名稱="Language" />
  <關連 英文名稱="Relation" />
  <範圍 英文名稱="Coverage" />
  <管理權 英文名稱="Rights Management" />
</DC基本元素>
</標準名稱>
```



# XML文件的儲存方式

---

- 目前儲存XML文件的資料庫系統，大體上可分為兩大類：
  - 以關聯式資料庫系統儲存。
  - 以原生型XML資料庫系統(native XML database system)儲存。



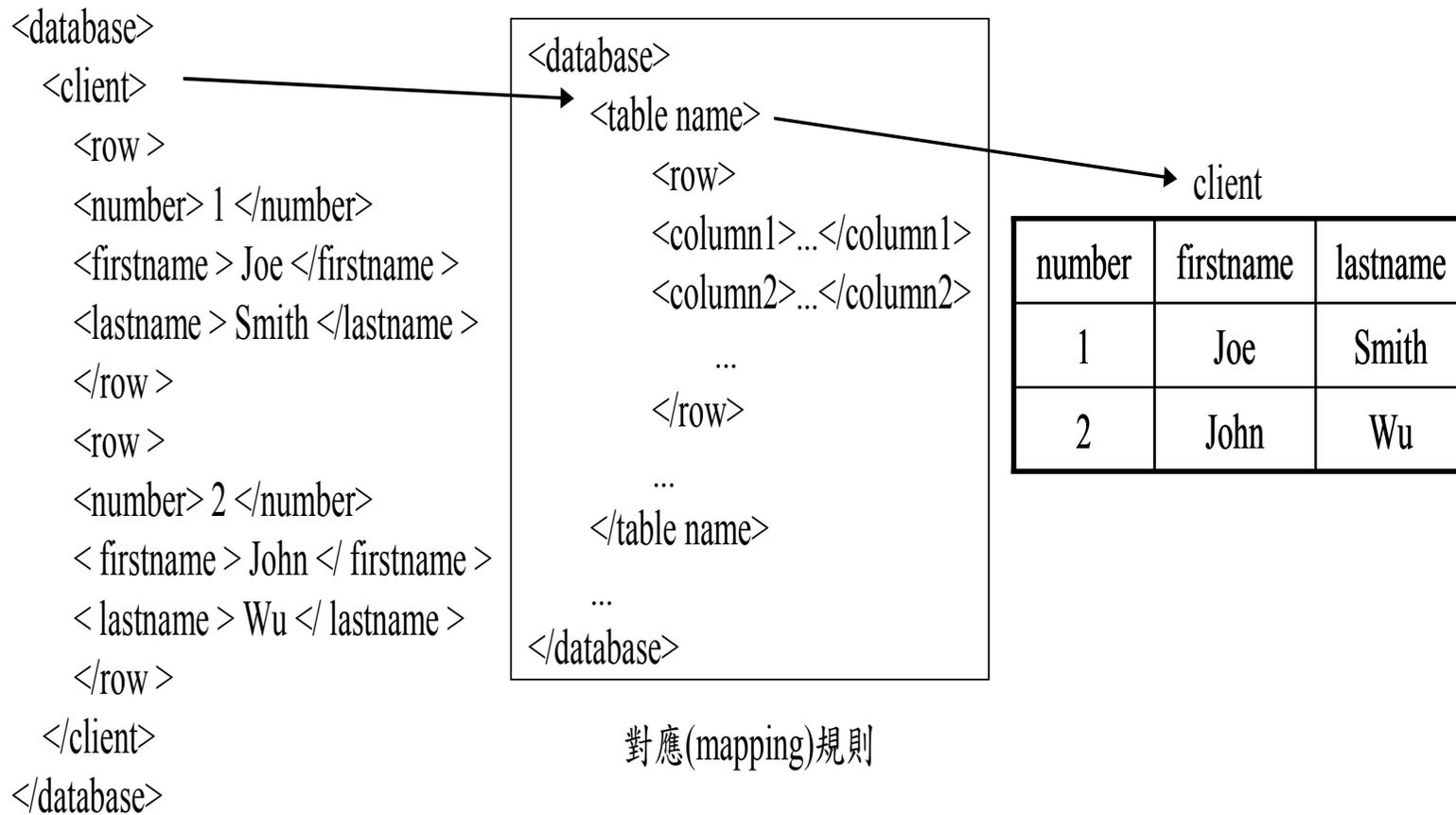
# 關聯式資料庫系統

## 儲存XML文件的方式

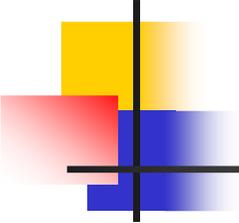
---

- 主要有兩種方式：
  - 透過事先定義的對應規則，將XML文件拆解成多個部份，存放在不同的表格(table)。
  - 將XML文件存入表格中的一個大型欄位(large object, LOB)。

# 透過對應規則將XML文件 拆解成表格來儲存(一)



XML文件



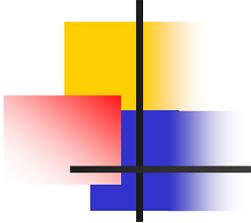
# 透過對應規則將XML文件 拆解成表格來儲存 (二)

- 優點：

- 可快速地處理結構固定的XML文件。
- 可直接套用在關聯式資料庫系統之上。

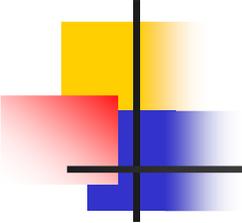
- 缺點：

- 當處理結構複雜的XML文件時，會產生大量的表格，需透過複雜的資料結合(join)，才能組合成完整的XML文件。
- 若XML文件結構需更動時，可能需重新定義對應規則。無法彈性地處理XML文件。



# 將XML文件存入LOB (I)

| id | date       | XML (LOB)   |
|----|------------|---|
| 1  | 2005-05-05 | <pre>&lt;purchase&gt;   &lt;customer Id="12345"&gt;     &lt;name&gt; Fred Allen &lt;/Name&gt;     &lt;address&gt;       &lt;street&gt;123 3rd Ave.&lt;/street&gt;       &lt;city&gt;Taipei&lt;/city&gt;     &lt;/address&gt;   &lt;/customer&gt;   &lt;order&gt;     &lt;item number="aa12345678" quantity="1"/&gt;     &lt;item number="aa98765432" quantity="2"/&gt;   &lt;/order&gt; &lt;/purchase&gt;</pre> |
| 2  | 2005-05-15 | <pre>&lt;purchase&gt;   &lt;customer Id="98765"&gt;     &lt;name&gt; John Smith &lt;/Name&gt;     &lt;address&gt;       &lt;street&gt;456 2nd Ave.&lt;/street&gt;       &lt;city&gt;Taipei&lt;/city&gt;     &lt;/address&gt;   &lt;/customer&gt;   &lt;order&gt;     &lt;item number="bb12345678" quantity="3"/&gt;     &lt;item number="bb98765432" quantity="4"/&gt;   &lt;/order&gt; &lt;/purchase&gt;</pre> |



# 將XML文件存入LOB (II)

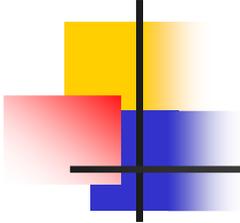
---

- 優點：

- 設計簡單，不需複雜的表格拆解動作即可儲存XML文件。亦可直接套用在關聯式資料庫系統之上。

- 缺點：

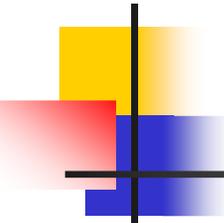
- 當僅需處理特定元素(element)之資料時，仍需輸出整份XML文件，再進行資料剖析(parsing)，才能正確地處理特定元素。



# 原生型XML資料庫系統 (Native XML Database System)

---

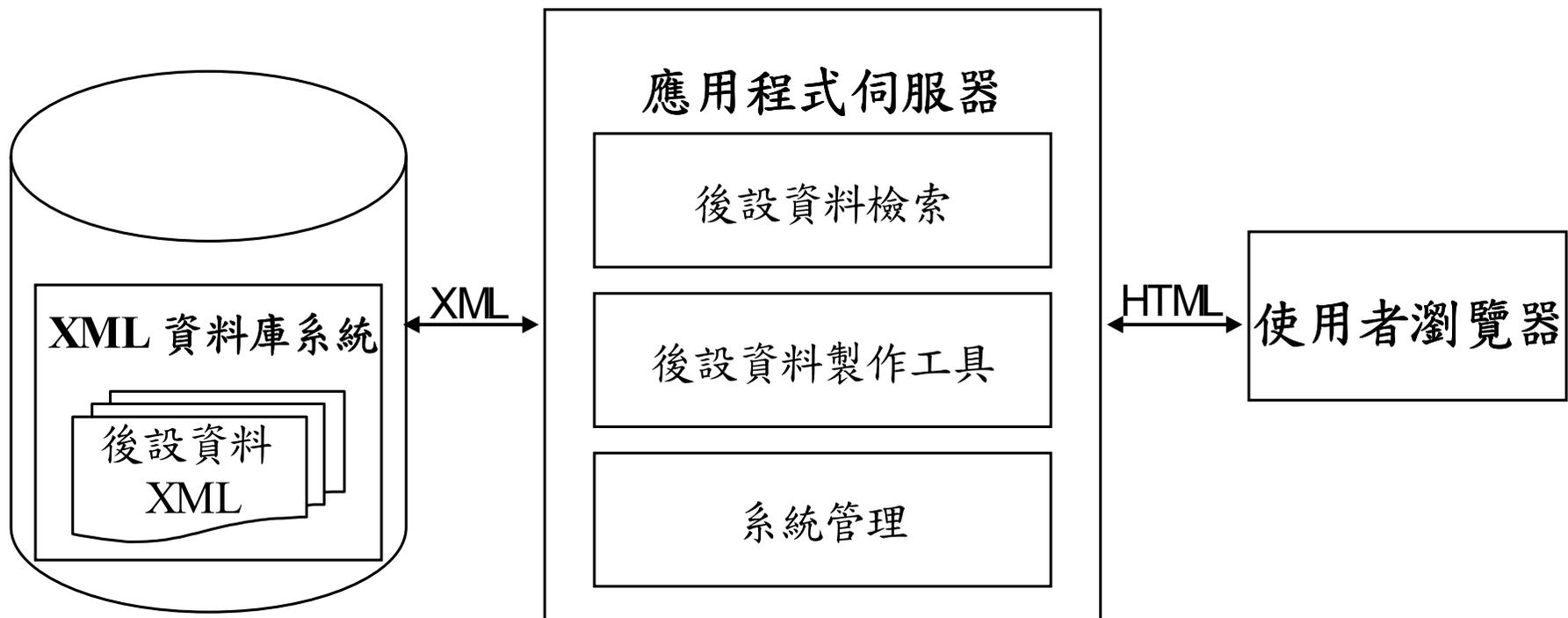
- 以XML文件為基本儲存單位，不需進行文件拆解動作，或預先設定儲存欄位的資料型態。
- 支援XPath、XQuery、XUpdate等查詢或更新XML文件內容的機制。
- 可容易地針對XML文件進行後續程式開發與應用。

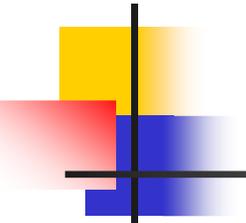


# 選用原生型XML資料庫的原因

- 與關聯式資料庫系統相比，原生型XML資料庫系統對XML文件的處理較完整，且有一些免費的資料庫系統可以使用。
- 數位典藏領域後設資料結構與內容複雜，且資料需求經常改變，以XML文件為基本儲存單位，對於後續更改XML文件有較高的彈性，所以我們以原生型XML資料庫系統儲存XML文件。

# 後設資料儲存庫系統架構





## 將XML文件轉換成HTML

---

- 運用JAXP的DOM API撰寫一程式，負責將XML文件中的內容轉換成HTML。
- 依據後設資料規格書中的表格形式，此程式會計算出XML文件中，各element的階層數及子element個數，並取出各element的名稱及attribute的值，以產生正確的HTML。

後設資料檢索功能

後設資料製作工具

資料庫管理

功能區

| 項目名稱                   | 英文名稱                                  | 資料型態          | 大小      | 必填 | 多值 | 屬性                 | 提供者 |
|------------------------|---------------------------------------|---------------|---------|----|----|--------------------|-----|
| <b>★全宗層次</b>           |                                       |               |         |    |    |                    |     |
| <a href="#">類型</a>     | Type                                  | Varchar       | 4       | *  |    | 固定值：檔案系統自動產生       | 系統  |
| <a href="#">典藏地</a>    | Geographic Location                   | Varchar       | 22      | *  |    | 預設值：中華民國臺灣省南投縣     | 填表者 |
| <a href="#">機關代碼</a>   | Institution Code                      | Varchar       | 6       | *  |    | 預設值：TH             | 填表者 |
| <a href="#">典藏機關</a>   | Administrator Unit                    | Varchar       | 16      | *  |    | 固定值：國史館臺灣文獻館系統自動產生 | 系統  |
| <a href="#">全宗號</a>    | Group Name                            | Varchar       | 3       | *  |    | 預設值：000            | 填表者 |
| <a href="#">全宗名</a>    | Group Number                          | Varchar       | 22      | *  |    | 預設值：臺灣總督府          | 填表者 |
| <a href="#">檔案附屬層級</a> | Description of Subordinate Components | Subgroup Name | Varchar | 40 | *  | 下拉式選單              | 填表者 |
| <b>系列層次</b>            |                                       |               |         |    |    |                    |     |
| <a href="#">副全宗名</a>   | Subgroup Name                         | Varchar       | 40      | *  |    | 下拉式選單              | 填表者 |

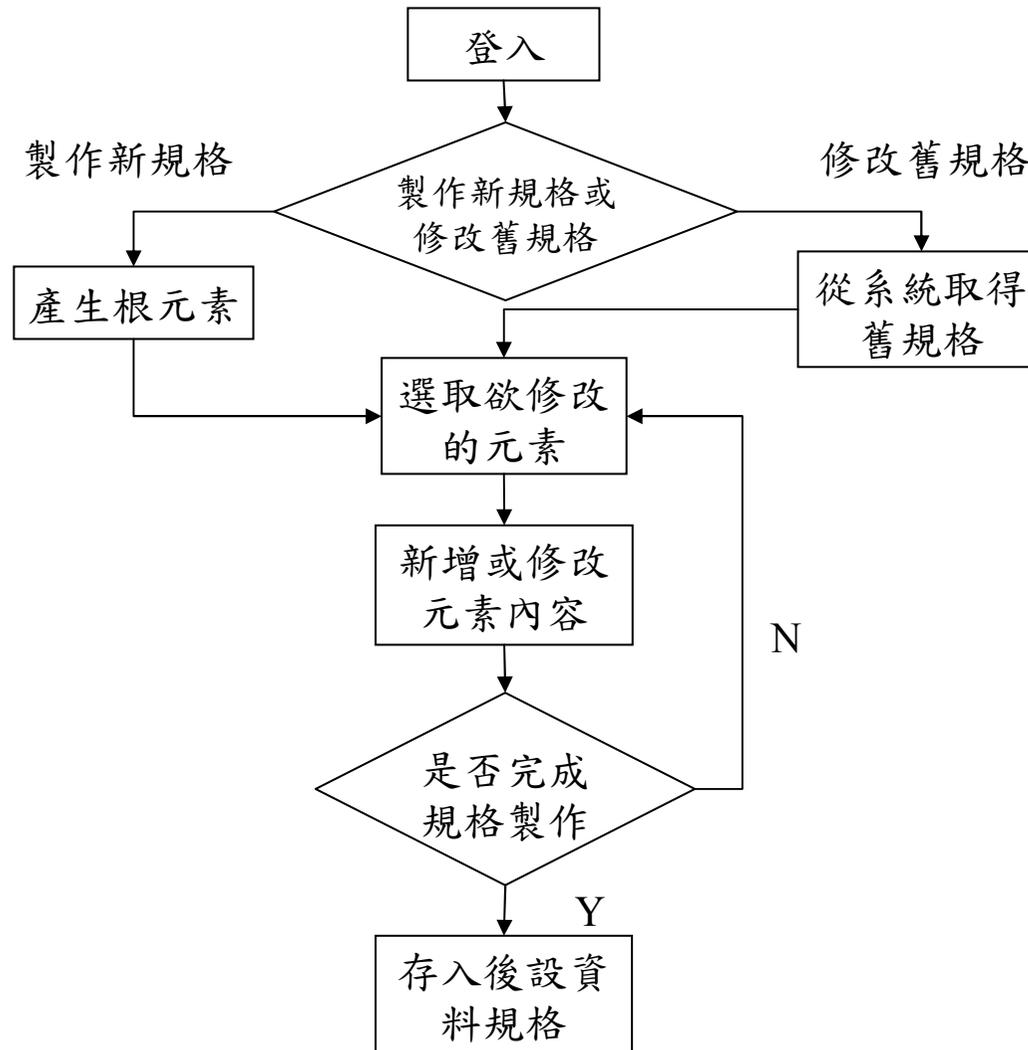
後設資料區

搜尋  為

之後設資料

操作區

# 後設資料製作工具運作流程

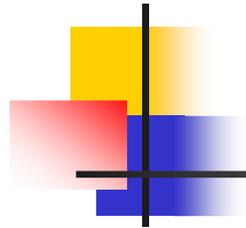


# 後設資料建置畫面

| 項目名稱        | 英文名稱               | 資料型態       | 大小                | 必填      | 多值      | 屬性 | 提供者 |       |     |
|-------------|--------------------|------------|-------------------|---------|---------|----|-----|-------|-----|
| <b>基本資料</b> |                    |            |                   |         |         |    |     |       |     |
| 識別資料        | ★遺物登錄號<br>新增 修改 刪除 | Identifier | Registered Number | Varchar | 30      | *  | 唯一  | 填表者   |     |
|             | 簡號                 |            | Pek. Coll.        | Varchar | 200     |    | 唯一  | 填表者   |     |
| 品名          | 主要名稱               | Title      | Main              | Varchar | 50      | *  |     | 填表者   |     |
|             | 其他                 |            | 類型                | Type    | Varchar | 20 |     | 下拉式選單 | 填表者 |
|             |                    |            | 名稱                | Name    | Varchar | 50 |     | ◎     | 填表者 |
|             | 備註                 |            | Remarks           | Varchar | 50      |    | 填表者 |       |     |

# 後設資料製作工具

## 修改元素內容

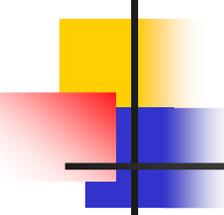


請輸入各欄位資訊來修改

| 項目名稱  | 英文名稱              | 資料型態    | 大小 |
|-------|-------------------|---------|----|
| 遺物登錄號 | Registered Number | Varchar | 30 |

| 必填                       | 多值                       | 屬性 | 提供者 |
|--------------------------|--------------------------|----|-----|
| <input type="checkbox"/> | <input type="checkbox"/> | 唯一 | 填表者 |

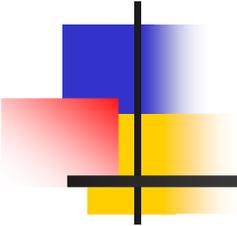
|            |
|------------|
| DC-Element |
| Identifier |



## 結論

---

- 後設資料儲存庫系統可讓數位典藏領域的不同計畫與機構，透過網路分享各自的後設資料規格，其提供後設資料規格的查詢與建立機制，期望讓整個典藏系統的建置更迅速與完整。



報告完畢

---

Q&A

謝謝您的指教