

# NTU Scholars

## 文件資料處理

台大圖書館系統資訊組

江玉婷

2019/03/26

TAIR團隊

陳光華館長、陳慧華組長

吳瑟量、江玉婷、林柏源、蔡玉紋、邱莉媛



SINCE 1928



 國立臺灣大學圖書館

# 大綱

---

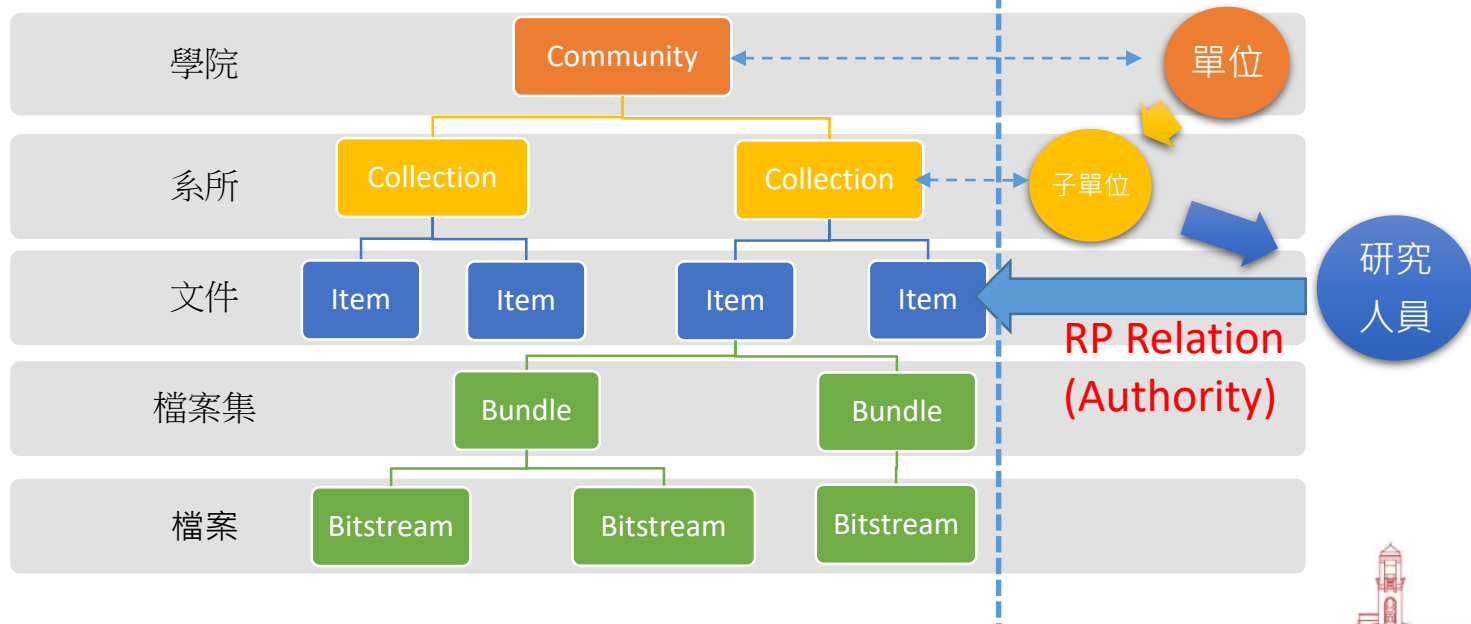
- 系統資料結構
- 資料移轉
- 相關設定



# 資料架構

文件視角(機構典藏)

研究人員視角(學術庫)



# 資料架構



# 資料結構

- Metadata內容種類

- Descriptive Metadata

- 描述物件相關資訊，可以提供檢索
    - 預設為Dublin Core

- Administrative Metadata

- 管理用資訊，如版權，存取權限，版本控制等

- Structural Metadata

- 描述欄位與其他欄位的關係

- Preservation Metadata

- 物件建立及變更等歷史資訊

NTU Scholars / 管理

後設資料欄位註冊 框架(sche

注意：新增一個後設資料欄位並不會在上傳表單中新增一個對應的輸入框，您可以在文件新增後，再使用修改後設資料的方式新增。

| ID / 欄位 / Qualifier / 欄位描述 | 欄位          | 欄位描述        | 操作                                |          |
|----------------------------|-------------|-------------|-----------------------------------|----------|
| 4                          | contributor | advisor     | Use primarily for thesis advisor. | 更新 刪除... |
| 5                          | contributor | author      |                                   | 更新 刪除... |
| 6                          | contributor | editor      |                                   | 更新 刪除... |
| 7                          | contributor | illustrator |                                   | 更新 刪除... |



# 資料結構

- 所有物件都有Metadata
  - Communities, Collections, Items, EPeople, Groups
  - 使用預設DC格式
    - Metadatavalue table
    - Collection name: dc.title

# 資料結構

- 彈性的Metadata 資料欄位 (table: metadatavalue)

- Text-Field : 欄位內容
- Authority : 權威資料代碼參照
- Confidence : 可信度
- Place : 出現順序
- Lang : 語言
- Resource\_type

| title         | author | date |
|---------------|--------|------|
| 臺灣原住民基本法之回... | 王泰升    | 1997 |

| resource_type | resource_id | metadata_field | metadata_value |
|---------------|-------------|----------------|----------------|
| Item          | 55          | title          | 臺灣原住民基本法之回...  |
| Item          | 55          | author         | 王泰升            |
| item          | 55          | date           | 1997           |

- 新增其他物件如 : eperson group(6), eperson(7)...



# 資料移轉

- Community & Collection
- 文件及檔案
- 帳號
- 權限
- 姓名權威資料



# 注意事項

由於新舊系統架構差異，  
請勿直接於資料庫表格中新增 / 編輯資料，  
均需透過系統後台或command line指令工具



# Community & Collection

- 資料架構
  - Collection：Dspace裝資料的容器，item一定要放在collection下
  - Community：collection的類別，可以有一層或多層
- 建議
  - 以異動性低的最小實體單位為collection
  - 不再以資料類型作collection
- 建立方式
  - 單筆重新建立
  - 以xml檔匯入



# Community & Collection

- xml匯入資料格式

```
<community identifier="123456789/1">
  <name>Community Name</name>
  <description>Descriptive text</description>
  <intro>Introductory text</intro>
  <copyright>Special copyright notice</copyright>
  <sidebar>Sidebar text</sidebar>
  <community identifier="123456789/2">
    <name>Sub CommunityName</name>
    <community identifier="123456789/3"> ...[ad infinitum]...
  </community>
  <collection identifier="123456789/4">
    <name>Collection Name</name>
    <description>Descriptive text</description>
    .....
    <license>Special licence</license>
    <provenance>Provenance information</provenance>
  </collection>
</community>
</import_structure>
```

- 匯入指令

```
[dspace]/bin/dspace structure-builder -f /path/to/source.xml -o
path/to/output.xml -e admin@user.com
```



# Community & Collection

- Community架構設定
  - 建立解除關聯

```
[dspace]/bin/dspace community-filiator --set --parent=parentID --child=childID
```

```
[dspace]/bin/dspace community-filiator --remove --parent=parentID -child=childID
```



# OAI資料同步

- 安裝NTUR patch程式以支援OAI-ORE
  - Open Archives Initiative Object Exchange and Reuse
  - 設定更新

- config/dspace.cfg

```
ore.authoritative.source = oai
plugin.named.org.dspace.content.crosswalk.IngestionCrosswalk = \
    .... org.dspace.content.crosswalk.OREIngestionCrosswalk = ORE
plugin.selfnamed.org.dspace.content.crosswalk.IngestionCrosswalk = .....
    org.dspace.content.crosswalk.OREIngestionCrosswalk
```

- config/oaicat.properties

```
Crosswalks.ore=org.dspace.app.oai.PluginCrosswalk
```

- 程式更新

```
org.dspace.app.oai.DSpaceRecordCatalog
org.dspace.app.oai.PluginCrosswalk
org.dspace.content.crosswalk.OREDisseminationCrosswalk
org.dspace.content.crosswalk.OREIngestionCrosswalk
org.dspace.content.crosswalk.XHTMLHeadDisseminationCrosswalk
```



# OAI資料同步

## (Open Archives Initiative Object Exchange and Reuse)

- 欄位轉換及對應
  - 自動產生所需欄位
    - dc.date.issued
    - dc.contributor.author (dc.creator)
    - dc.relation.journalVolume, dc.relation.journalIssue, dc.relation.pages (dc.relation)
  - 多值欄位切分：dc.subject及dc.contributor.author
    - 以利權威代碼資料對應
- 隱藏資料無法移轉



# OAI資料同步

- 設定collection OAI來源

- 後台設定
- 資料表直接設定
  - harvested\_collection

- 排程：定時與來源資料端同步

```
$DSPACE/bin/dspace oai import
```

- 手動同步單一collection

```
dspace dsrun  
org.dspace.app.harvest.Harvest -r -e  
admin@ntu.edu.tw -c [collection  
internal id]
```

## 獲取設定

內容來源

- 這是一個標準的集合
- 這個類別自動從外部服務獲取內容

OAI 提供者

<http://ntur.lib.ntu.edu.tw/ir-oai/>

OAI 資料集 ID

hdl\_246246\_4835

後設資料格式

Dublin Core

內容已被獲取

- 只獲取後設資料
- 獲取後設資料並對應到檔案 ( 必須支援 ORE )
- 獲取後設資料和檔案 ( 必須支援 ORE )

最後的獲取結果

Harvest from  
<http://ntur.lib.ntu.edu.tw/ir-oai/request>  
successful



# 整批匯入

- 將舊系統資料匯出，整理成CRIS系統所需資料格式
- 需選擇特定collection
- 內容格式
  - DSpace Simple Archive Format
  - CSV Metadata：管理介面匯出資料

# 整批匯入—SAF

- DSpace Simple Archive Format (SAF)

```
archive_directory/  
  item_000/  
    dublin_core.xml      -- qualified Dublin Core metadata for metadata fields  
belonging to the dc schema  
    metadata_[prefix].xml -- metadata in another schema, the prefix is the name of  
the schema as registered with the metadata registry  
    contents             -- text file containing one line per filename  
    collections         -- text file that contains the handles of the collections  
the item will belong two. Optional. Each handle in a row.  
                        -- Collection in first line will be the owning collection  
    file_1.doc          -- files to be added as bitstreams to the item  
    file_2.pdf  
  item_001/  
    dublin_core.xml  
    contents  
    file_1.png  
    ...
```



# 整批匯入—SAF

- NTUR SAF匯出

```
[dspace]/bin/dspace export --type=COLLECTION --id=collectionID_or_handle -  
-dest=/path/to/destination --number=seq_num
```

- 執行匯入

- 管理→整批匯入(有檔案大小限制)

- 伺服器執行指令

```
[dspace]/bin/dspace import  
--add --eperson=joe@user.com  
--collection=CollectionID  
--source=items_dir  
--zip=filename.zip --mapfile=mapfile
```

NTU Scholars / 管理者

## 批次上傳

Select the type of the input data

Simple Archive Format (zip file via upload)

Select data file to upload

瀏覽... 未選擇檔案。

選擇文件的所屬集合 (選項)

此欄位為選項，如果不提供資訊，您應該要在檔案中包含文件所屬集合

選擇集合

上傳

# 整批匯入— MetadataOnly

- 匯出現有清單
  - 後台→文件管理→瀏覽單一系統後匯出
- 整理成所需格式
  - 轉換為utf-8 csv檔
  - 內容結構
    - 標題列：id, collection, action, [metadata\_column]
    - 新增：在id欄位寫+



# 整批匯入— MetadataOnly

- 匯入執行方式

- 管理介面

- 有筆數限制，可設定

- ```
${dspace.dir}/config/module/bulkedit.cfg
```

- 可一次匯入多個collection

- 管理→內容→匯入後設資料

- 執行指令

- ```
${dspace.dir}/bin/metadata-import -f [file full path] -e [eperson email]
```



# 整批匯入— MetadataOnly

🏠 首頁 內容 ▾ CRIS 模組

NTU Scholars

研究

- 研究人
- 計畫
- 單位
- Mana
- 匯出
- 匯入
- Web s

機構典藏

- 文件
- 去除重複
- 審核流程
- 監看者
- 皮用工作
- 已隱藏文件
- 私人文件
- 匯入後設資料**
- 整批匯入
- 權威管理

## 匯入後設資料

瀏覽... 0912.csv

上傳

新的文件:

|                                   |                        |
|-----------------------------------|------------------------|
| 加到新的從屬集合                          | (123456789/64095): 醫學系 |
| 新增 (dc.contributor.author[en_US]) | YANG, PAN-CHYR         |
| 新增 (dc.date.issued)               | 1991/4/1               |
| 新增 (dc.title[en_US])              | thisistest             |

套用修改

取消



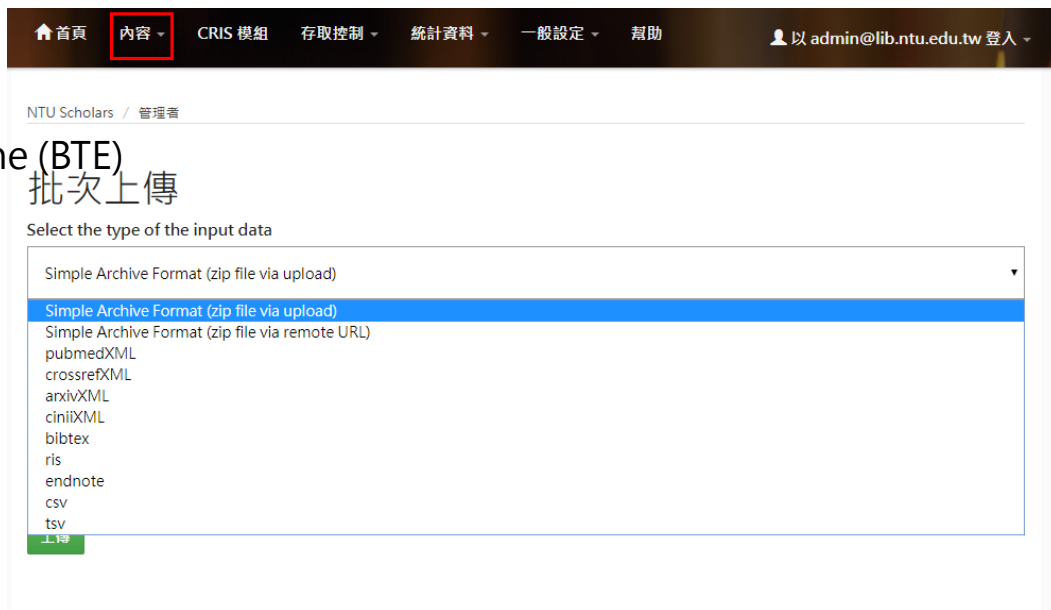
# 整批匯入—MetadataOnly

- 其他匯入格式

- 管理→內容→整批匯入

- Biblio-Transformation-Engine (BTE)

- arXiv
- BibTeX (\*.bib)
- CiNii
- CrossRef
- CSV (\*.csv)
- EndNote
- OAI-PMH
- PubMed
- RIS (\*.ris)
- TSV (\*.tsv)



The screenshot shows the NTU Scholars system interface. At the top, there is a navigation bar with links for '首頁', '內容', 'CRIS 模組', '存取控制', '統計資料', '一般設定', and '幫助'. The '內容' link is highlighted with a red box. Below the navigation bar, the user is logged in as 'admin@lib.ntu.edu.tw'. The main content area displays 'NTU Scholars / 管理' and '批次上傳'. A dropdown menu is open, showing the following options: 'Simple Archive Format (zip file via upload)', 'Simple Archive Format (zip file via remote URL)', 'pubmedXML', 'crossrefXML', 'arxivXML', 'cinniXML', 'bibtex', 'ris', 'endnote', 'csv', 'tsv', and '上傳' (highlighted in green).

```
[dSPACE]/bin/dSPACE import -b -m mapFile -e [email] -c 123456789/1 -s path-to-my-  
arxiv-file -i arxivXML
```



# 資料取得介面(API) — 資料物件

- REST API <http://<dSPACE-webapp>/rest/>
- 資料物件處理

| 功能            | Community | Collection | Item | Bitstream |
|---------------|-----------|------------|------|-----------|
| 查詢所有列表        | ✓         | ✓          | ✓    | ✓         |
| 查詢子物件         | ✓         | ✓          | ✓    |           |
| 查詢單一物件資訊      | ✓         | ✓          | ✓    | ✓         |
| 建立物件          | ✓         | ✓          | ✓    | ✓         |
| 更新物件及Metadata | ✓         | ✓          | ✓    | ✓         |
| 刪除物件          | ✓         | ✓          | ✓    | ✓         |
| 查詢/更新存取政策     |           |            | ✓    | ✓         |

- 登入/登出



# 資料取得介面(API)—資料物件

- 執行方式

- GET : 查詢

- GET /communities/{communityId}/collections - Returns an array of collections
    - GET /items/{item id}/metadata - Return metadata of the specified item.

- POST : 新增、修改、刪除、登入

- POST /items/{item id}/bitstreams - Add a bitstream to the specified item.
    - DELETE /communities/{communityId}/collections/{collectionId}
    - POST /login - Method for logging into the DSpace RESTful API.  
{"email":"test@dspace","password":"pass"}. Returns a token which must be included in future requests in the "rest-dspace-token" header.

- 回傳物件: JSON OR XML

Item Object

```
{ "id":14301,"name":"2015 Annual  
Report","handle":"123456789/13470","type":"item","link":"/rest/items/14301","lastMod  
fied":"2015-01-12 15:44:12", "parentCollection":null, "parentCommunityList":null,  
"bitstreams":null, "archived":"true","withdrawn":"false"}
```



# 登入授權機制(Authentication)

- 本地端密碼登入
- SSO標準協定
  - Shibboleth
  - LDAP
  - X.509
- IP授權

```
${dspace.dir}/config/dspace.cfg
```

```
plugin.sequence.org.dspace.authenticate.AuthenticationMethod =  
    org.dspace.authenticate.IPAuthentication
```

```
${dspace.dir}/config/modules/authentication-ip.cfg
```

```
ip.GROUPNAME = iprange[, iprange ...]
```

- 外部登入
  - ORCID
  - 客製化 : NTUAuthentication



# 登入授權機制(Authentication)

```
${dspace.dir}/config/modules/authentication.cfg
```

```
plugin.sequence.org.dspace.authenticate.AuthenticationMethod =  
org.dspace.authenticate.OAuthAuthenticationMethod,
```

```
org.dspace.authenticate.PasswordAuthentication, org.dspace.authenticate.NTUAuthentication
```

```
${dspace.dir}/config/modules/authentication-[method].cfg
```

Log In to DSpace

New user? [Click here to register.](#)

Please enter your e-mail address and password into the form below.

E-mail Address:

Password:

[Have you forgotten your password?](#)

Use your **ORCID** or create a new one

ORCID provides a persistent digital identifier that distinguishes you from every other researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between you and your professional activities ensuring that your work is recognized.

Click on the ORCID logo to sign in with your existent record or create a new one.



NTUR

Log In  計算機及資訊網路中心帳號認證系統



# 帳號

- 主機本地端帳號：Dspace EPerson
- 新系統管理介面無整批建立帳號功能
- command line使用者操作
  - 新增，修改，刪除

```
[dspace]/bin/dspace user --add --email jquser@example.com -g John -s User --password hiddensec  
[dspace]/bin/dspace user --modify -m george@example.com  
[dspace]/bin/dspace user --delete -n martha
```

- 整批處理
  - 刪除久未登入的使用者

```
[DSpace]/bin/dsrun org.dspace.eperson.Groomer -a -b 07/20/1969 -d  
[DSpace]/bin/dsrun org.dspace.eperson.Groomer -a -b 07/20/1969
```



# 帳號

- ORCiD

- 不存在的帳號自動建立

```
config/modules/authentication-oauth.cfg  
autoregister=true
```

- 單一簽入(台大規格)

- 提供single sign on 之介接API (web service)

- 需有技術人員開發界接程式
- 運作概念
  - 遠端認證介面 (各校提供url)
  - 回傳認證結果 (各校提供 session)
  - 若認證通過，以session number取回使用者資料 (各校提供API)
  - 若使用者第一次登入，則新建帳號 (NTUR)
  - 依使用者所屬系所自動授予上傳權限給名稱相同的社群 (NTUR)



# 權限—系統功能使用權限

- 功能使用權限：EpersonGroup
  - 讀取、上傳、編輯、刪除、審核、整批上傳、類別管理
- 系統無提供整批設定工具
- 資料庫處理
  - 取得新舊資料表之collection\_id, eperson\_id, epersongroup\_id對應後，以程式整批處理
    - epersongroup：群組主表
      - 到metadatavalue table查詢群組名稱：resource\_type\_id=6
      - 命名規則：COLLECTION\_[collection\_id]\_ACTION\_NAME
    - Epersongroup2eperson：群組與使用者關聯
    - Group2group：群組間關聯



# 資料取用控制(Authorization)

- 資料主體
  - Community
  - Collection
  - Item
  - Bitstream
- 授權對象
  - Eperson
  - Eperson Group
    - Anonymous
    - Administrator
    - COLLECTION\_XXX\_SUBMIT
    - COLLECTION\_XXX\_WORKFLOW
    - COLLECTION\_XXX\_ADMIN

- 存取權限

| POLICY NAME            | 說明     | 適用物件                                     |
|------------------------|--------|--|
| READ                   | 讀取     | Community<br>Collection<br>ItemBitstream |
| WRITE                  | 寫入/編輯  |  |
| ADD/REMOVE             | 新增/刪除  |  |
| WORKFLOW XX            | 審核     |  |
| ADMIN                  | 管理     |  |
| DEFAULT_ITEM_READ      | 預設讀取文件 | Collection                               |
| DEFAULT_BITSTREAM_READ | 預設讀取檔案 |  |
| WITHDRAWN_READ         | 讀者隱藏文件 |  |

- 適用日期



# 資料取用控制(Authorization)

- 一般使用者權限設定
  - Collection上傳權限：COLLECTION\_XXXX\_SUBMIT
  - 上傳後之item並無自動給予WRITE/REMOVE權限
    - 需具有COLLECTION\_ADMIN以上角色的使用者(群組)才能修改和刪除
    - 以版本控制取代文件編輯：上傳者可建立版本

## 歷史版本

\* 選擇版本

| 版本                       | 文件 | 編輯器               | 日期           |                        |
|--------------------------|----|-------------------|--------------|------------------------|
| <input type="checkbox"/> | 2  | 123456789/25142*  | admin, admin | 2018-09-21 05:18:37.62 |
| <input type="checkbox"/> | 1  | 123456789/25142.1 | admin, admin | 2015-03-17 09:22:18.0  |

刪除版本

返回

## 管理工具

編輯(上傳模式)

編輯(管理模式)

匯出文件備份

匯出便於移轉的文件

匯出後設資料

為此文件建立新版本

顯示歷史版本

## 版本

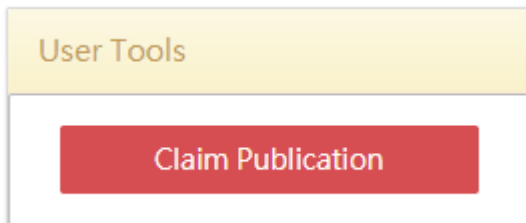
建立文件的新版本：25118

建立新版本的理由



# 資料取用控制(Authorization)

- 資料認領 (claim publication)



- 資料存取申請 (request a copy)
  - 檔案不開放下載時，向原上傳者申請資料存取權

Items in This Item:

| Description              | Size  | Format  |           |
|--------------------------|---|---------|-----------|
| ous-Research-Article.pdf | Accepted version of my precious scientific manuscript | 8.85 kB | Adobe PDF |

[View/Open](#) [Request a copy](#)

## Request a document copy: Test item with closed access document

Requester name:

Requester e-mail:

Files:

all files (of this document) in restricted access

the file(s) you requested

Message:

# 權限—資料存取

- 權限設定元件

- 設定資料主體：社群、類別、文件、檔案
- 授權對象：使用者或群組
- 動作：讀取、新增、編輯、刪除、審核、類別管理員、整批上傳...
- 存取日期限制：正向標示可讀取日期

- 整批設定

- 管理→存取控制→授權→進階/文件大量政策管理工具
- 無法整批設定開放存取日期

- 以IP群組設定方式取代舊有各文件瀏覽限制設定

```
$ {dspace.dir}/config/modules/authentication-ip.cfg  
ip.GROUPNAME = iprange[, iprange ...]
```

The screenshot shows the DSpace administration interface. At the top, there is a navigation bar with links for '首頁', '內容', 'CRIS 模組', '存取控制', '統計資料', '一般設定', and '幫助'. Below this, the breadcrumb path is 'NTU Scholars / 管理者'. A dropdown menu is open under '存取控制', showing options for '使用者', '群組', and '授權', with '授權' selected. The main content area is titled '管理授權政策' and includes a help icon. Below the title, it says '選擇一個物件並管理其政策：'. There is a table with three rows: '管理社群政策', '管理集合政策', and '管理文件政策'. The last row is highlighted with a red box, and a link '進階/文件大量政策管理工具' is visible below it. Below the table, there is a section titled '進階政策管理' with a help icon. It contains a warning message: '這功能可以讓您大量新增或移除一個集合內所有文件或檔案的權限。警告：如果移除讀取權限會導致...'. There are two dropdown menus: '集合' and '群組'. The '集合' dropdown is open, showing a list of departments and research centers, with '中國文學系' selected. The '群組' dropdown is also open, showing a list of groups, with 'Administrator' selected. At the bottom, there is a section for '執行動作' with 'WRITE' selected.

# 姓名權威資料

- NTUR做法：
  - 於後台建立作者姓名變異表
  - 以作者查詢文件時，系統自動做詞彙擴展，增加搜尋廣度
  - 無法解決同名同姓問題
- NTU Scholars做法：
  - 於研究人員(RP)主檔中建立別名
  - 查詢研究人員時，會依作者各種姓名的建檔查詢可能符合資料，再選擇特定研究人員進入取得其所屬文件
- 匯入研究人員主檔時，可一併匯入姓名權威資料



# 相關設定

- 權威資料及下拉式選單
- 外部資料來源參照
- 各種key值設定
- SOLR索引
- 背景排程工作



# 權威資料管理 (Authority Management)

- 權威代碼 (Authority Key)
- 欄位權威參照設定

```
`${dspace.dir}/config/dspace.cfg
```

- 研究物件權威資料定義 (ChoiceAuthority plugin)

```
plugin.named.org.dspace.content.authority.ChoiceAuthority =  
org.dspace.app.cris.integration.RPAuthority = RPAuthority,  
org.dspace.app.cris.integration.OUAuthority = OUAuthority,
```
- 欄位關聯設定

```
choices.plugin.dc.contributor.author = RPAuthority  
choices.presentation.dc.contributor.author = suggest  
authority.controlled.dc.contributor.author = true
```
- 新增文件資料時
  - 自動建立新的權威物件
  - 自動建立文件與權威物件連結



# 權威資料管理 (Authority Management)

- 設定item與權威物件之關聯

The screenshot displays a web interface for authority management. At the top, there are tabs for 'dc', 'relation', and 'ispartof'. A search box contains the text 'Government Information Quarterly' and is highlighted with a red box. Below the search box, a dropdown menu shows 'journals00183' with a magnifying glass icon next to it, also highlighted with a red box. A red arrow points from the search box to the dropdown. Below the dropdown, a list of results is shown, with the first item 'Government Information Quarterly(journals00183)' highlighted. A second red arrow points from this item to the dropdown. The interface also shows a language dropdown set to 'en' and a search button with a magnifying glass icon.

# 權威資料管理 (Authority Management)

- 權威資料參照統計：管理→內容→權威管理

權威使用統計 RPAuthority

| Stat   | # of ...   |
|--|--|
| # of prelim matched items to RPs by confidence | <div style="display: flex; justify-content: space-between; align-items: center;"><div style="display: flex; gap: 10px;"><span>🔍</span><span>🗑️</span><span>⚠️</span><span>✖️</span><span>❓</span><span>⚙️</span><span>✅</span></div><div style="display: flex; gap: 10px;"><span>0</span><span>1</span><span>0</span><span>0</span><span>76</span><span>1</span><span>295389</span></div></div> <div style="text-align: center; margin-top: 5px;"><span style="border: 1px solid gray; padding: 2px 5px;">Accept/Reject</span></div> |
| # of authors total in the IR (RP & non-RP)     | 786551   |
| # of in use RPs                                | 2551   |
| # of authority keys pending approval           | 3  |
| # of items (prelim & accepted) on RPs          | 183564   |

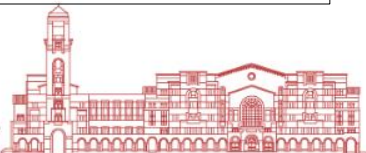


# 資料選擇管理 (Choice Management)

- 應用時機
  - 單筆資料上傳
  - 單筆資料編輯
- Metadata輸入之選項參照
  - Input-form.xml檔內簡單參照
  - 階層性控制詞彙
    - OWL或RDF相容

```
`${dspace.dir}/config/input-forms.xml
<field>
  <dc-schema>dc</dc-schema>
  <dc-element>type</dc-element>
  <dc-qualifier></dc-qualifier>
  <repeatable>>true</repeatable>
  <label>Type</label>
  <input-type value-pairs-
name="common_types">dropdown</input-type>
  <hint></hint>
  <required></required>
  <vocabulary closed="false">abc</vocabulary>
</field>
...
config/controlled-vocabularies/abc.xml
<select name="common_types">
  <option VALUE="article">Article</option>
  <option VALUE="book">Book</option>
.....
</select>
```

```
choices.plugin.dc.type = common_types
choices.presentation.dc.type = select
```



# 外部資料來源參照欄位對應

- Biblio-Transformation-Engine (BTE)

- 整批匯入指令

```
[dspace]/bin/dspace import -b -m mapFile -e example@email.com -c 123456789/1 -s path-to-my-xxxxx-file -i [xxxxXML]
```

- 欄位對應設定 [dspace]/config/spring/api/bte.xml

```
<bean name="wosInputMap" class="java.util.HashMap" scope="prototype">
  <constructor-arg>
    <map key-type="java.lang.String" value-type="java.lang.String">
      <entry key="isild" value="isild" />
      <entry key="doi" value="doi" />
      <entry key="issn" value="jissn" />
      <entry key="journalTitle" value="journal" />
      <entry key="title" value="title" />
      <entry key="year" value="issued" />
      <entry key="volume" value="volume" />
      <entry key="issue" value="issue" />
      <entry key="startPage" value="firstpage" />
      <entry key="endPage" value="lastpage" />
    
```

```
<bean name="outputMap" class="java.util.HashMap" scope="prototype">
  <constructor-arg>
    <map key-type="java.lang.String" value-type="java.lang.String">
      <entry value="jissn" key="dc.identifier.issn" />
      <entry value="jeissn" key="dc.identifier.eissn" />
      <entry value="pisbn" key="dc.identifier.isbn" />
      <entry value="eisbn" key="dc.identifier.eisbn" />
      <entry value="journal" key="dc.relation.ispartof" />
      <entry value="proceedingsPaper.journal" key="dc.relation.ispartofbook" />
    
```



# 外部資料來源參照欄位對應

- crosswalks

[dspace]/config/crosswalks/

- oaidc.properties
- orcid-work-metadata.properties
- bibtex.template
- citation.template
- endnote.template

```
# <schema>.<element>.<qualifier|*> = element oai_dc
# you can map more then one metadata to the same oai_dc element,
# see for example (dc.contributor.editor, dc.contributor.advisor, etc.)
dc.title.* = title
dc.contributor.author = creator
dc.contributor = contributor
dc.contributor.editor = contributor
dc.contributor.advisor = contributor
dc.contributor.illustrator = contributor
dc.contributor.other = contributor
dc.date.* = date
dc.description = description
dc.description.abstract = description
dc.description.sponsorship = description
dc.description.statementofresponsibility = description
```



# 外部資料來源參照欄位對應

- Google索引優化

- Google Scholar Metadata對應

- Highwire press tag (in HTML HEAD tag)

- Prefix: citation\_ 

```
<meta name="citation_keywords" content="article" />
<meta name="citation_title" content="Examining the socio-technical determinants in
<meta name="citation_issn" content="0740-624X" />
<meta name="citation_language" content="aaa" />
<meta name="citation_author" content="TUNG-MOU YANG" />
<meta name="citation_date" content="2016" />
<meta name="citation_abstract_html_url" content="https://scholars.lib.ntu.edu.tw/h
```

- 系統設定

- 開啟highwire press tag功能

```
${dspace.dir}/config/dspace.cfg
```

```
google-metadata.enable = true
```

- 設定欄位對應

```
${dspace.dir}/config/crosswalks/google-metadata.properties
```

```
google.citation_title = dc.title
```

```
google.citation_publisher = dc.publisher
```

```
google.citation_author = dc.author | dc.contributor.author | dc.creator
```



# 各種key值設定

- **google** : [dspace]/config/dspace.cfg

```
jspui.google.analytics.key=  
key.googleapi.maps =
```
- **查詢外部資料庫API** [dspace]/config/dspace.cfg

```
#### Lookup submission provider ####  
submission.lookup.scopus.apikey =  
submission.lookup.scivalcontent.apikey =  
submission.lookup.webofknowledge.ip.authentication = true  
submission.lookup.webofknowledge.user =  
submission.lookup.webofknowledge.password =
```



# Solr索引

- 索引指令

```
[dspace]/bin/dspace index-discovery [-cbhf[r <item handle>]]
```

- 索引設定

- [dspace]/config/modules/discovery.cfg

- discovery.search.server**

- discovery.search.server=[http://localhost:8080/solr/search]

- discovery.index.authority.ignore[.field]**

- discovery.index.authority.ignore=true

- discovery.index.authority.ignore.dc.contributor.author=false

- discovery.index.authority.ignore-prefered[.field]**

- discovery.index.authority.ignore-prefered=true

- discovery.index.authority.ignore-prefered.dc.contributor.author=false

- discovery.index.authority.ignore-variants[.field]**

- discovery.index.authority.ignore-variants=true

- discovery.index.authority.ignore-variants.dc.contributor.author=false



# Solr索引設定

- [dspace]/config/spring/api/discovery.xml
  - Sidebar facets : **searchFilterAuthor** , **searchFilterSubject** , **searchFilterIssued**
  - Search filters : **searchFilterTitle** , **searchFilterAuthor** , **searchFilterSubject** , **searchFilterIssued**
  - Sort fields : **sortTitle** , **sortDateIssued**
  - defaultFilterQueries
  - Recent Submissions.
  - Hit highlighting
    - The fields dc.title, dc.contributor.author & dc.subject can contain hit highlighting.
    - The dc.description.abstract & full text field are used to render search snippets.
  - Non indexed metadata fields
    - **Community/Collections**: dc.rights (copyright text)
    - **Items**: dc.description.provenance



# 背景排程工作

- 本機索引
  - \$DSPACE/bin/dspace index-discovery > /dev/null
  - \$DSPACE/bin/dspace index-authority > /dev/null
  - \$DSPACE/bin/network-builder : 研究人員網路
  - \$DSPACE/bin/dspace generate-sitemaps
- 統計
  - \$DSPACE/bin/view-and-download-retrieve : 瀏覽及下載統計
  - \$DSPACE/bin/dspace stats-util -l
- 批次工作
  - \$DSPACE/bin/subscribe-\* : 訂閱紀錄發送
  - \$DSPACE/bin/dspace checker -l -p
  - \$DSPACE/bin/dspace cleanup > /dev/null



# 背景排程工作

- 資料同步
  - OAI : `$DSPACE/bin/dspace oai import > /dev/null`
  - ORCID : `$DSPACE/bin/dspace dsrun org.dspace.app.cris.batch.ScriptPushOrcid`
  - Scopus: `./dspace dsrun org.dspace.app.cris.batch.ScopusFeed -p <submitter-email> -c collectionID`
  - Web of Knowledge: `./dspace dsrun org.dspace.app.cris.batch.WosFeed -p <submitter-email> -c collectionID`
  - Pubmed Europe: `./dspace dsrun org.dspace.app.cris.batch.PMCEuropeFeed -p <submitter-email> -c collectionID`
- 引文資訊查詢
  - Scopus: `$DSPACE/bin/scopus-retrieve`
  - Web of Knowledge: `$DSPACE/bin/wos-retrieve`
  - PMC: `$DSPACE/bin/pubmed-retrieve-citation-second`



# 備份

- 系統檔案：dspace主目錄
  - 文件檔案：[dspace]/assetstore
  - 參數設定檔：[dspace]/config
- 資料庫
  - postgresql
- 歷史紀錄(無法重建)
  - [dspace]/log/\*, [dspace]/reports/\*, ....等
- 系統程式(可重建)
  - [tomcat]/webapps/\*
  - [dspace]/bin/\*, [dspace]/lib/\*
- 索引檔(可重建)
  - [dspace]/search/\*, [dspace]/solr/\*, [dspace]/sitemaps/\*,



# 障礙排除

- 查看log
  - [dspace]/log/solr.log : 索引類錯誤
  - [dspace]/log/dspace.log.[date] : 內部系統錯誤
  - [tomcat]/log/catalina.\* , localhost.\* : 其他錯誤(如畫面一片空白)
- 設定log模式為debug模式
  - 修改/opt/dspace/config/log4j.properties
  - 把INFO改為DEBUG



感謝聆聽 敬請指教



SINCE 1928

