

NTU Scholars

~dspace7導論與文件資料處理實務~



SINCE 1928



 國立臺灣大學圖書館

台大圖書館系統資訊組

江玉婷

2022/8/25

TAIR團隊

陳光華館長、陳慧華組長

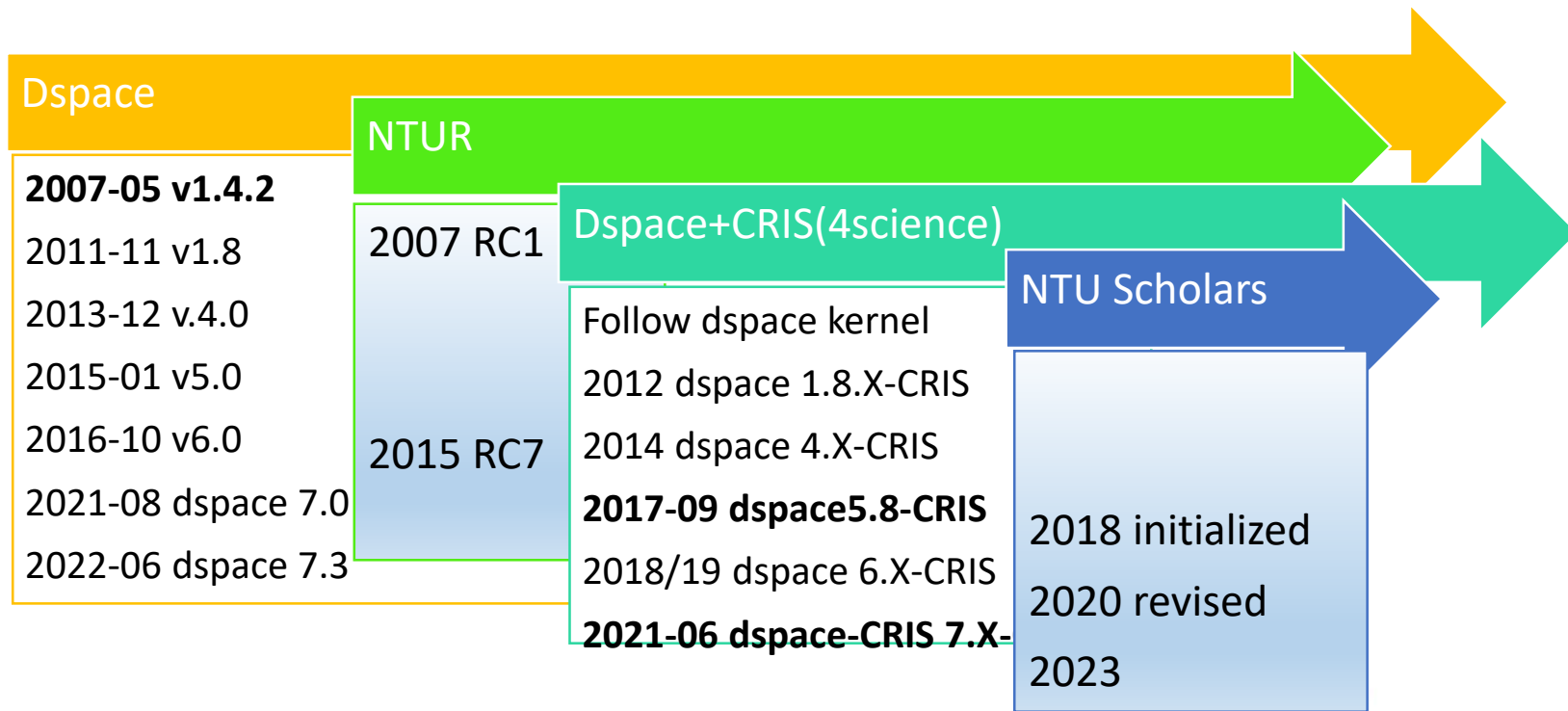
吳瑟量、蔡玉紋、邱莉媛、江玉婷、林柏源

大綱

- Dspace+CRIS發展現況
- 新版功能特色
- 文件資料管理與移轉



Dspace, Dspace-CRIS與NTU Scholars發展歷程



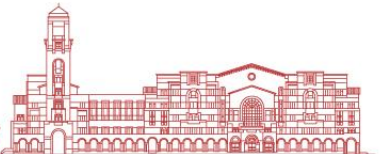
新版功能特色

- 介面技術
- 前端瀏覽及查詢
- 資料編輯與上傳
- 管理功能
- 後端：REST API



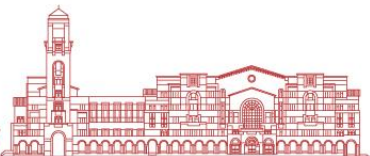
新的前端技術

- Angular UI : Javascript-based Type Script
- 前端網頁組成
 - A *.component.ts (TypeScript)
 - A *.component.html (HTML)—HTML5
 - Component tag: HTML-like tags (e.g. <ds-header>, <ds-footer>).
 - A *.component.scss (Sass / CSS)



介面變革

- 網站主題
 - Bootstrap(V4.X)
src/themes/dspace
 - 客製化
src/themes/custom
- UI體驗
 - drag-and-drop
 - smart reloading
 - 快取機制



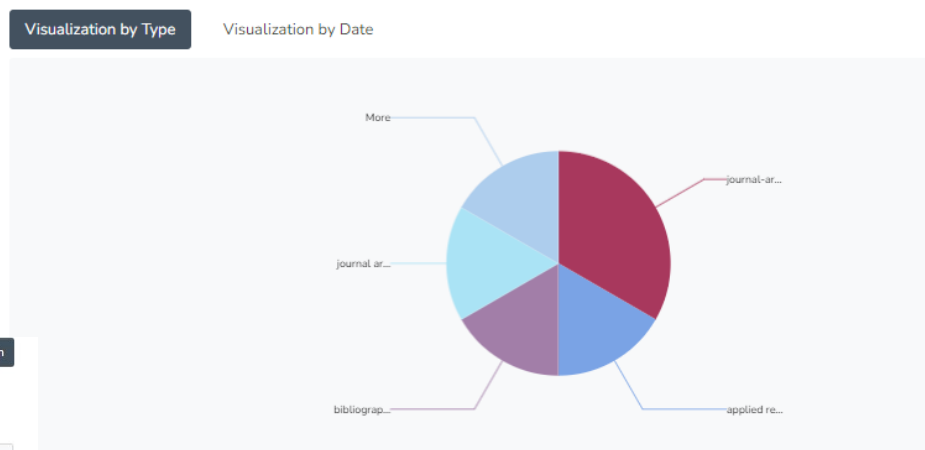
查詢結果頁面

- List/grid

A screenshot of a search results page. At the top, there is a search bar containing the text "kuang hua" and buttons for "Search" and "Export". Below the search bar, the page is titled "Search Results" and indicates "Now showing 1 - 6 of 6". On the left side, there is a "Filters" sidebar with expandable sections for "Type", "Author", "Date", "Has files", "Item Type", and "Language". The main content area shows a search result for a "Research Project" titled "Taiwan Academic Institutional Repository". Below this, there is a "Publication" result titled "Experiments on Cross-Language Information Retrieval Using Comparable Corpora of Chinese, Japanese, and Korean Languages (2021) CHEN, KUANG-HUA". To the right of the text are two small icons: a purple flower and a colorful cube with the number "1". At the bottom of the search bar area, there is a secondary search bar with the text "test" and a "Search" button.

A screenshot of a search results page in grid view. The search bar at the top contains "test" and a "Search" button. The page is titled "Search Results" and indicates "Now showing 1 - 3 of 3". On the left side, there is a "Filters" sidebar with a "Reset filters" button and expandable sections for "Type", "Author", "Date", "Has files", "Item Type", and "Type". Below the filters is a "Settings" section with "Sort By" set to "Most Relev" and "Results per page" set to "10". The main content area displays three search results in a grid. The first result is a "Publication" titled "test" with the text "2022, Man, Test1" and a "View" button. The second result is a "Person" titled "Man, Test1" with a "View" button and a small illustration of a man in a suit. The third result is a "Publication" titled "test0718" with a "View" button. The first and third results have a placeholder image that says "No Thumbnail Available".

- 視覺化後分類圖表



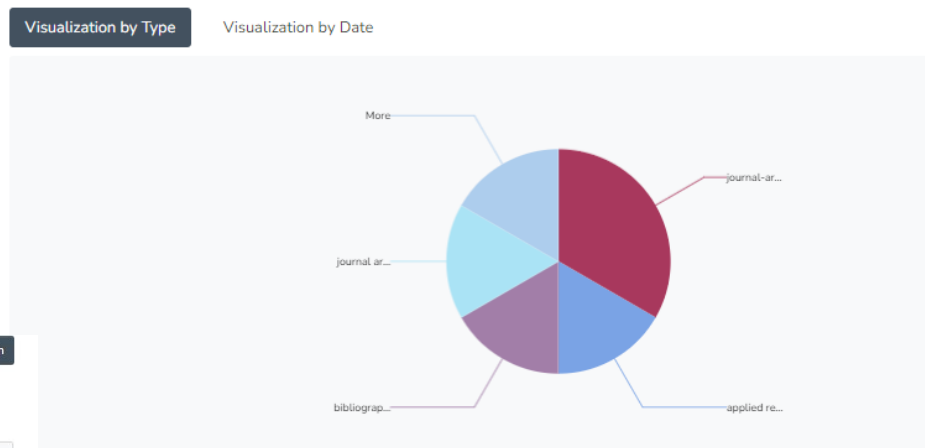
查詢結果頁面

- List/grid

A screenshot of a search results page. At the top, there is a search bar containing the text "kuang hua" and buttons for "Search" and "Export". Below the search bar, the page is titled "Search Results" and indicates "Now showing 1 - 6 of 6". On the left side, there is a "Filters" sidebar with expandable sections for "Type", "Author", "Date", "Has files", "Item Type", and "Language". The main content area shows a search result for a "Research Project" titled "Taiwan Academic Institutional Repository". Below this, there is a "Publication" result titled "Experiments on Cross-Language Information Retrieval Using Comparable Corpora of Chinese, Japanese, and Korean Languages (2021) CHEN, KUANG-HUA". To the right of the text are two small icons: a purple flower and a colorful cube with the number "1". At the bottom of the search bar area, there is a secondary search bar with the text "test" and a "Search" button.

A screenshot of a search results page in grid view. The search bar at the top contains "test" and has a "Search" button. The page is titled "Search Results" and shows "Now showing 1 - 3 of 3". On the left, there is a "Filters" sidebar with "Reset filters" and expandable sections for "Type", "Author", "Date", "Has files", and "Item Type". Below the filters is a "Settings" section with "Sort By" set to "Most Relev" and "Results per page" set to "10". The main content area displays three grid items. The first item has a placeholder "No Thumbnail Available". The second item is a "Person" result for "Man, Test1" with a thumbnail illustration of a man in a suit and a "View" button. The third item is a "Publication" result for "test0718" with a placeholder "No Thumbnail Available" and a "View" button.

- 視覺化後分類圖表



查詢結果頁面

- 後分類篩選機制

Filters

Type —

- journal-article 2
- applied research 1
- bibliography 1
- journal article 1

Search type

Author —

- CHEN, KUANG-HUA 2
- Alexey Omelchenko 1
- Jieh Hsiang 1
- Kuang Hua Chen 1
- Kuang-hua Chen 1

[Show more](#)

Search author name

Date —

Start	End
2020	2022

Search Results

Now showing 1 - 6 of 6

Research Project
Taiwan Academic Institutional Repository

Publication
Experiments on Cross-Language Information Retrieval Corpora of Chinese, Japanese, and Korean Language (2021) CHEN, KUANG-HUA

Publication
add oai-aire supports to repository platforms (2018) Alexey Omelchenko ; CHEN, KUANG-HUA

Person
CHEN, KUANG-HUA

Publication
A longitudinal study of intellectual cohesion in digital libraries using bibliometric analyses (2017) TANG, MUH-CHYUN ; Yun Jen Cheng ; Kuang Hua Chen

Publication
The unique approach to institutional repository (2009) Kuang-hua Chen ; Jieh Hsiang ; Jin Chen

- 排序與呈現筆數設定

Settings

Sort By

Most Relevant

Results per page

10



Metadata資料編輯

- 控制詞彙建議
- 欄位內容自動補完
- 樹狀結構與多值
- 檢查重複

Language

English

Search

N/A

English (United States)

English

Spanish

German

French

Author

Author *

chen

chen

Affiliation :

CHEN, KUANG-HUA

Affiliation : Library

Lili Chen Chen

ORCID iD : 0000-0002-7090-1061

Xi Chen

Hierarchical tree view

Collection test coll

No value entered

+ Add m

Editors

No value entered

+ Add m

Type *

Type

Select the type(s) of content of the item.

text

annotation

bibliography

blog post

book

conference output

journal

lecture

letter

magazine

manuscript

Potential duplicates

The system has identified some potential duplicates. Please carefully review the list and flag each occurrence with the appropriate choice or discard this submission.

Now showing 1 - 1 of 1

Publication

test workflow

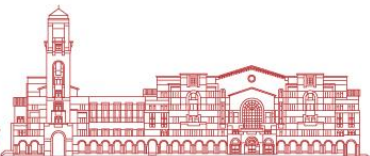
(2022)

No Abstract

It's a duplicate

It's not a duplicate

UNIVERSITY LIBRARY



Metadata資料編輯

Edit Item

Status

Bitstreams

Metadata

Relationships















Version History

Collection Mapper

 Upload

 Save

 Discard

Name	Description	Format	Actions
BUNDLE: ORIGINAL			
NTUR_RC7_note.pdf		Adobe PDF	   
oai-ore.pdf		Adobe PDF	   
BUNDLE: LICENSE			
license.txt		License	   

 Back

 Save

 Discard

檔案編輯

- 拖拉式檔案上傳
- 拖拉式資料排序
- 檔案Metadata
- 檔案存取政策

Edit bitstream ✕

Enter the name of the file *

台大學術成果資訊系統單筆資料連結至NTUR 系統.pdf

Enter a description for the file

Enter a description for the file

File type

Personal picture, logo, main article, etc.

Access condition type

Select an access condition to apply on the bitstream once the item is deposited

Grant access from *

From



Select the date from which the related access condition is applied

Grant access until *

Until



Select the date until which the related access condition is applied

Access condition type

Select an access condition to apply on the bitstream once the item is deposited

MyDSpace

- 上傳功能列

The screenshot displays the MyDSpace interface. At the top, there is a dashed box containing the text "Drag & Drop your files here, or browse" with a small upload icon. To the right of this box are two dark blue buttons: one with a white plus sign and a dropdown arrow, and another with a white document icon and a dropdown arrow. Below the upload area, on the left, is a "Show" dropdown menu currently set to "Your Submissions". In the center, there is a search bar with the placeholder text "Search the repository ...". To the right of the search bar, the heading "Your submissions" is displayed, followed by the text "Now showing 1 - 10 of 36". On the far right, a vertical dropdown menu is open, listing categories: "Funding", "Journal", "Organizational Unit", "Patent", and "Person".

- 文件審核工作流程

文件匯入/匯出

- 單筆資料編輯
- 外部資料上傳
- 多種格式支援
 - SAF：支援物件關聯編輯
 - CSV
 -
- 上傳流程(workflow)
- 指定文件/檔案存取政策

Import a publication from an external source

Search the external source

Your search returned no results.

[← Back to MyDSpace](#)

Pubmed ▾

Search

Pubmed ▲

arXiv

Scopus ▼



MyDSpace

- 可查詢之文件列表
 - 歷史上傳文件
 - 待審核文件

The screenshot displays the MyDSpace repository interface. At the top, there is a search bar with the text "Search the repository ..." and buttons for "Search" and "Export". Below the search bar, the "Your submissions" section is visible, showing "Now showing 1 - 10 of 27".

The "Filters" section is highlighted with a red box. It includes a "Status" filter with a minus sign, showing "Workspace" (15) and "Archived" (12). There is also a "Search status" input field with the text "Status". Below the filters, there are sections for "Type" and "Date", both with plus signs, and a "Reset filters" button.

The "Settings" section is also visible, showing "Sort By" set to "Last modified Descendi" and "Results per page" set to "10".

The main content area displays a list of submissions. The first submission is titled "The unique approach to institutional repository" by Kuang-hua Chen and Jieh Hsiang Jin Chen. It is marked as "Archived" and "Publication". Below the title, there is a "View" button. The second submission is titled "Social Communication Questionnaire-Chinese Version" by (PsycTESTS Dataset, 2011) GAU, SUSAN SHUR-FEN. It is marked as "Workspace" and "Item". Below the title, there are "Edit" and "Delete" buttons. The third submission is titled "add oai-aire supports to repository platforms" by (2018) Alexey Omelchenko CHEN, KUANG-HUA. It is marked as "Archived" and "Item". Below the title, there is a "View" button. The fourth submission is titled "No title" and is marked as "Workspace" and "Item". Below the title, there are no buttons.

多媒體物件

- 嵌入式數位檔案瀏覽框架
 - 支援IIIF (International Image Interoperability Framework)
 - Mirador 3.0
- 串流式內容播放 (streaming)



統計機制

- Solr 統計
 - 使用統計
 - 下載
 - 查詢
 - 工作流程
- 各階層物件/文件統計
 - Community
 - Collection
 - Item

Statistics by Category



統計機制

Experiments on Cross-Language Information Retrieval Using Comparable Corpora of Chinese, Japanese, and Korean Languages

Date Issued
Author(s)
DOI

2021
CHEN, KUANG-HUA 
10.1007/978-981-15-5554-1_2



Altmetric



Dimensions



Google-Scholar



Views

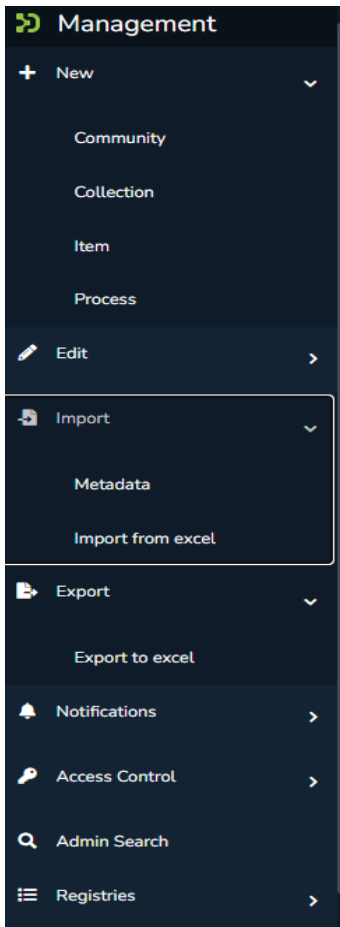


Downloads

- Google analytics
 - `google.analytics.key = UA-XXXXXX-X`



管理功能



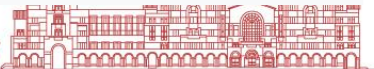
- ← 固定式管理功能列

Home • Processes Overview

Processes Overview

Now showing 1 - 20 of 54 + New

Process ID	Name	User	Start time (UTC)	Finish time (UTC)	Status	
54	curate	admin ntu	2022-08-23 03:09:15	2022-08-23 03:09:15	COMPLETED	
53	item-export	admin ntu	2022-07-20 08:13:41	2022-07-20 08:13:41	COMPLETED	
52	harvest	admin ntu	2022-07-20 07:31:11	2022-07-20 07:31:14	COMPLETED	
51	harvest	admin ntu	2022-07-20 07:24:37	2022-07-20 07:24:40	COMPLETED	
50	harvest	admin ntu	2022-07-20 07:21:18	2022-07-20 07:21:21	COMPLETED	
49	bulk-import	admin ntu	2022-07-18 07:01:14	2022-07-18 07:01:14	COMPLETED	
48	bulk-import	admin ntu	2022-07-18 07:00:27	2022-07-18 07:00:27	COMPLETED	
47	bulk-import	admin ntu	2022-07-18 06:59:09	2022-07-18 06:59:09	FAILED	



管理功能

- 文件查詢
- Curation Tasks
- 管理者代登入

Home • Administrative Search

Search the repository ...

Administrative Search

Now showing 1 - 10 of 15

Organizational Unit
Library

Organizational Unit
National Taiwan University

Publication
test
(2022) Man, Test

Person
chen

Person

Filters

- Private +
- Withdrawn +
- Author +
- Date +
- Has files +
- Item Type +

Settings

Sort By
Most Relevant ▾

背景程序執行與管理

Create a new process

Script

metadata-deletion

- Choose a script...
- process-cleaner
- bulk-import
- collection-export
- metadata-deletion
- metadata-export
- metadata-import
- harvest
- filter-media
- import-nbevents
- orcid-bulk-pull
- orcid-bulk-push
- import-oaire-suggestions
- update-item-references
- item-enhancer
- bulk-item-export
- item-export
- curate
- index-discovery
- cris-layout-tool

metadata-deletion

Delete all the values of the specified metadata field

`-m --metadata <value> metadata field name`

`-l --list` lists the metadata fields that can be deleted

[Home](#) • [Processes Overview](#) • [Create a new process](#)

Create a new process

Script

item-export

Parameters

--id

--format

Add a parameter...

Cancel

Save

後端：整合式資料存取平台

- REST API
- OAI-PMH
- OpenAIRE4
- SWORD (v1 or v2)
- RDF (linked open data)

REST API

- For Angular
- 外部系統整合
- HAL browser

The HAL Browser [Go To Entry Point](#) [About The HAL Browser](#) [Login](#)

Explorer

Custom Request Headers

Properties

```
{
  "dSPACEUI": "https://demo7.dspace.org",
  "dSPACEName": "DSpace7 Demo provided by 4Science",
  "dSPACEServer": "https://api7.dspace.org/server",
  "dSPACEVersion": "DSpace 7.4-SNAPSHOT",
  "type": "root"
}
```

Links

rel	title	name / index	docs	GET	NON-GET
actuator					
authn					
authorizations					
bitstreamformats					
bitstreams					
browses					

Inspector

Response Headers

```
200 200
access-control-expose-headers: Authorization, expires, Location, Content-Disposition
cache-control: no-cache, no-store, max-age=0, must-revalidate
connection: Keep-Alive
content-language: en
content-type: application/hal+json;charset=UTF-8
date: Tue, 23 Aug 2022 03:16:22 GMT
expires: 0
keep-alive: timeout=5, max=100
pragma: no-cache
server: Apache
strict-transport-security: max-age=31536000 ; includeSubDomains, max-age=31536000
transfer-encoding: chunked
vary: Origin,Access-Control-Request-Method,Access-Control-Request-Headers
x-content-type-options: nosniff
x-frame-options: SAMEORIGIN, DENY
x-xss-protection: 1; mode=block
```

Response Body

```
{
  "dSPACEUI": "https://demo7.dspace.org",
  "dSPACEName": "DSpace7 Demo provided by 4Science",
  "dSPACEServer": "https://api7.dspace.org/server",
  "dSPACEVersion": "DSpace 7.4-SNAPSHOT",
  "type": "root",
  "_links": {
    "actuator": {
      "href": "https://api7.dspace.org/server/actuator"
    }
  }
}
```

其他技術議題

- 相關底層技術版本更新
 - Upgrade UI to Angular v10
 - Upgrade UI to Node v12 or v14 support
 - Upgrade Backend to Solr v8 support
 - Upgrade to ORCID v3 support
 - Upgrade to SHERPA v2 support
- 資訊安全更新
 - LGTM.com 資安議題回報及修正
 - 管理介面資安機制補強
 - REST API資安機制補強



資料移轉

- OAI資料同步
- 整批匯出/匯入



OAI資料同步

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

- 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
```

[http://\[OAI URL\]/request?verb=ListRecords&metadataPrefix=ore&from=](http://[OAI URL]/request?verb=ListRecords&metadataPrefix=ore&from=)

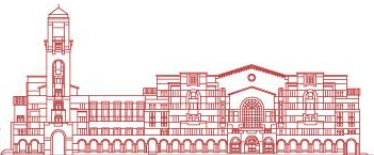
<http://ntur.lib.ntu.edu.tw/ir-oai/request?verb=ListRecords&metadataPrefix=ore&from=2018-01-01T10:45:23Z&until=2019-06-31T00:51:11Z>

OAI資料同步

- 欄位轉換及對應
 - 自動產生所需欄位

新系統欄位	舊系統欄位
dc.date.issued	dc.date
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
 - table: 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	<input type="radio"/> 這是一個標準的集合 <input checked="" type="radio"/> 這個類別自動從外部服務獲取內容
OAI 提供者	<input type="text" value="http://ntur.lib.ntu.edu.tw/ir-oa"/>
OAI 資料集 ID	<input type="text" value="hdl_246246_4835"/>
後設資料格式	<input type="text" value="Dublin Core"/>
內容已被獲取	<input type="radio"/> 只獲取後設資料 <input type="radio"/> 獲取後設資料並對應到檔案 (必須支援 ORE) <input checked="" type="radio"/> 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dspace-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 DC metadata for metadata fields belonging to the dc sche  
    metadata_[prefix].xml  -- metadata in another schema  
    contents         -- text file containing one line per filename  
    collections      -- contains the handles of the collections the item will belong to.  
                    -- 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

SAF轉檔工具--SAFBuilder

- 執行環境
 - Command line / terminal
 - Java JDK
 - Git
 - Maven
- 下載安裝及執行

```
git clone https://github.com/Dspace-Labs/SAFBuilder.git
cd SAFBuilder
./safbuilder.sh -c /AAA_batch-metadata.csv -z
```

- 輸入：csv+file

	A	B	C	D	E	F
1	filename	dc.title	dc.contributor	dc.date.issued	dc.description	dc.format.extent
2	Alabama.pdf	Alabama	robert r. riley	12/14/1819	yellowhammer state	4447100
3	Alaska.pdf	Alaska	Sean Parnell	12/3/59	the last frontier	626,932
4	American Samoa.pdf	American Samoa	Togiola Tulafono		None	73.97 years
5	Arizona.pdf	Arizona	Jan Brewer	2/14/12	grand canyon state	5130632
6	Arkansas.pdf	Arkansas	Mike Beebe	6/15/1836	the natural state	2,834,797
7	California.pdf	California	arnold schwarzeneg	9/9/1850	golden state	33871648
8	Colorado.pdf	Colorado	bill ritter	8/1/1876	centennial state	4,861,515

- 輸出：SAF格式



整批匯入— MetadataOnly

- 匯出現有清單
 - 後台→文件管理→瀏覽單一系所後匯出
- 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+



	ACTION	ID必要欄位	其他資料欄位
新增	留空白	在id欄位寫「+」	直接填寫 最多十八欄
更新	留空白	以匯出資料為準，必填	最多十八欄 不更新者不列
刪除	EXPUNGE	以匯出資料為準，必填	可直接刪除

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
```

```
gui-item-limit =
```

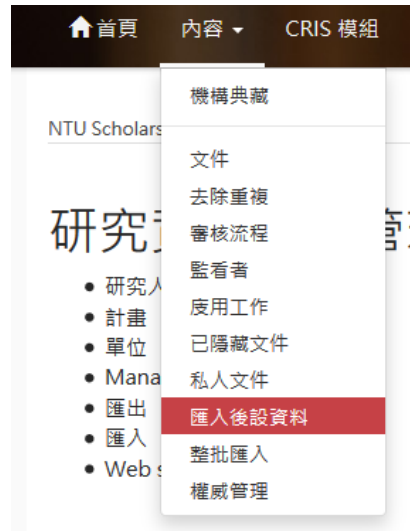
```
allowexpunge = true
```

- 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]
```



匯入後設資料

瀏覽... 0912.csv

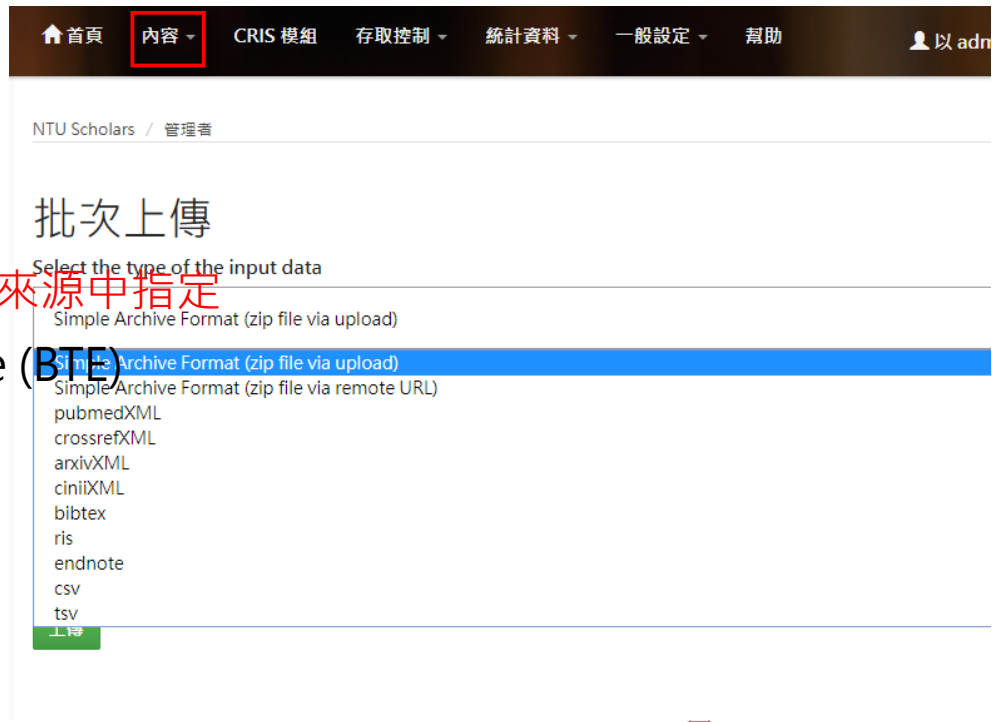
上傳



整批匯入

- 其他匯入格式
 - 管理→內容→整批匯入
 - 若不指定collection，需於資料來源中指定
 - Biblio-Transformation-Engine (BTE)
 - arXiv BibTeX (*.bib)
 - CiNii CrossRef
 - CSV (*.csv) EndNote
 - OAI-PMH PubMed
 -

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



NTU Scholars / 管理

批次上傳

Select the type of the input data

- Simple Archive Format (zip file via upload)
- Simple Archive Format (zip file via upload)**
- Simple Archive Format (zip file via remote URL)
- pubmedXML
- crossrefXML
- arxivXML
- ciniiXML
- bibtex
- ris
- endnote
- csv
- tsv
- 上傳



感謝聆聽 敬請指教



SINCE 1928

