NTU Scholars

~dspace7導論與文件資料處理實務~

台大圖書館系統資訊組

江玉婷

2022/8/25

TAIR團隊

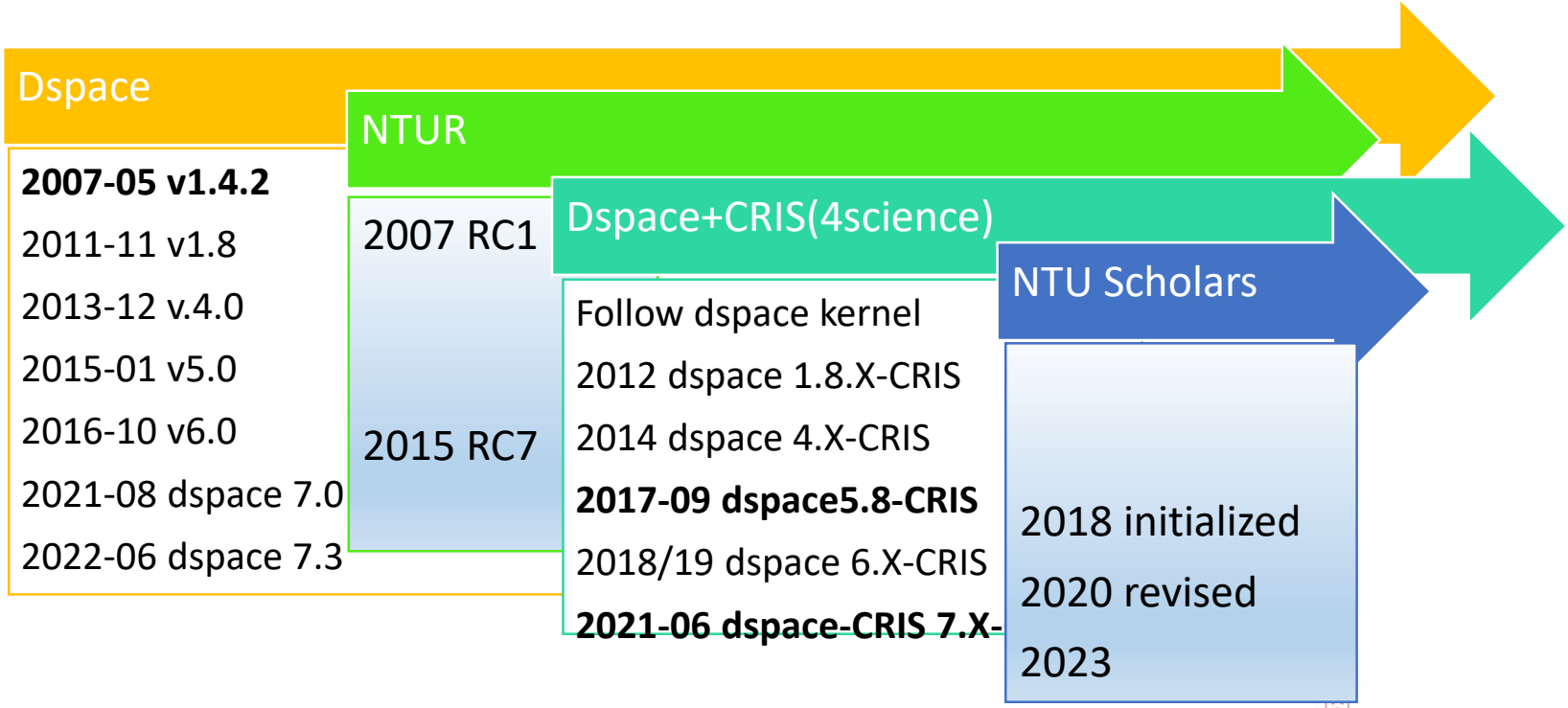
陳光華館長、陳慧華組長

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發展進程



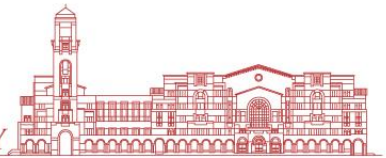
NATIONAL TAIWAN UNIVERSITY LIBRARY

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)

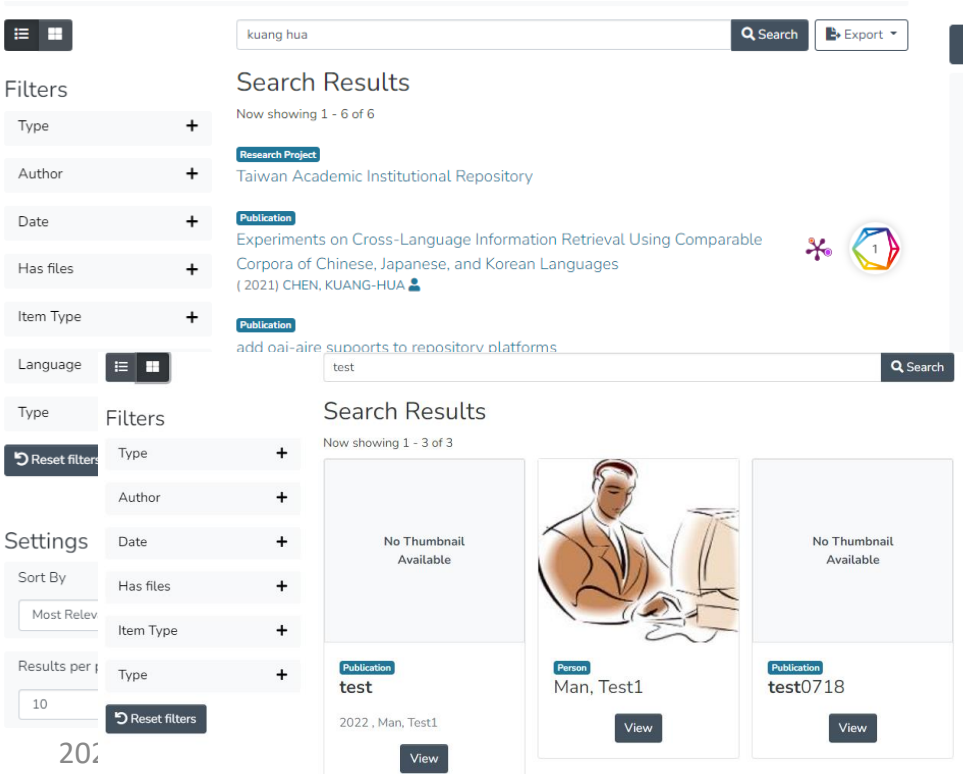


介面變革

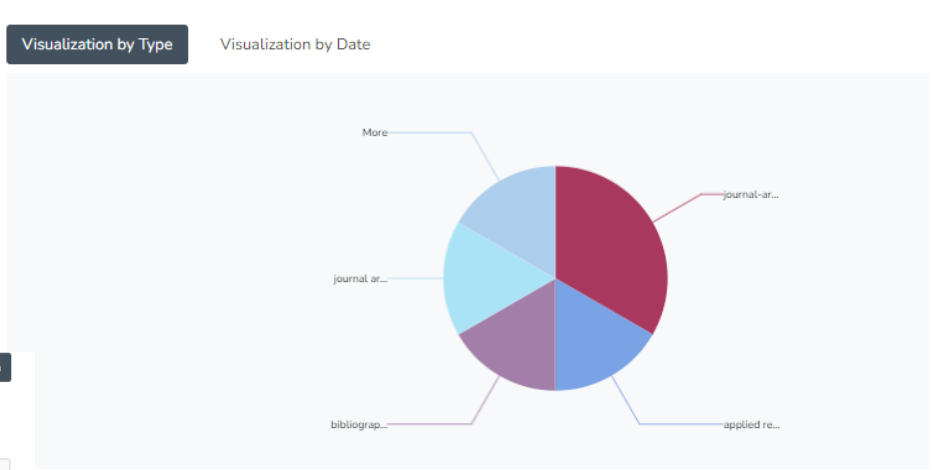
- 網站主題
  - Bootstrap(V4.X)
    src/themes/dspace
  - 客製化
    src/themes/custom
- UI體驗
  - drag-and-drop
  - smart reloading
  - 快取機制

查詢結果頁面

- List/grid



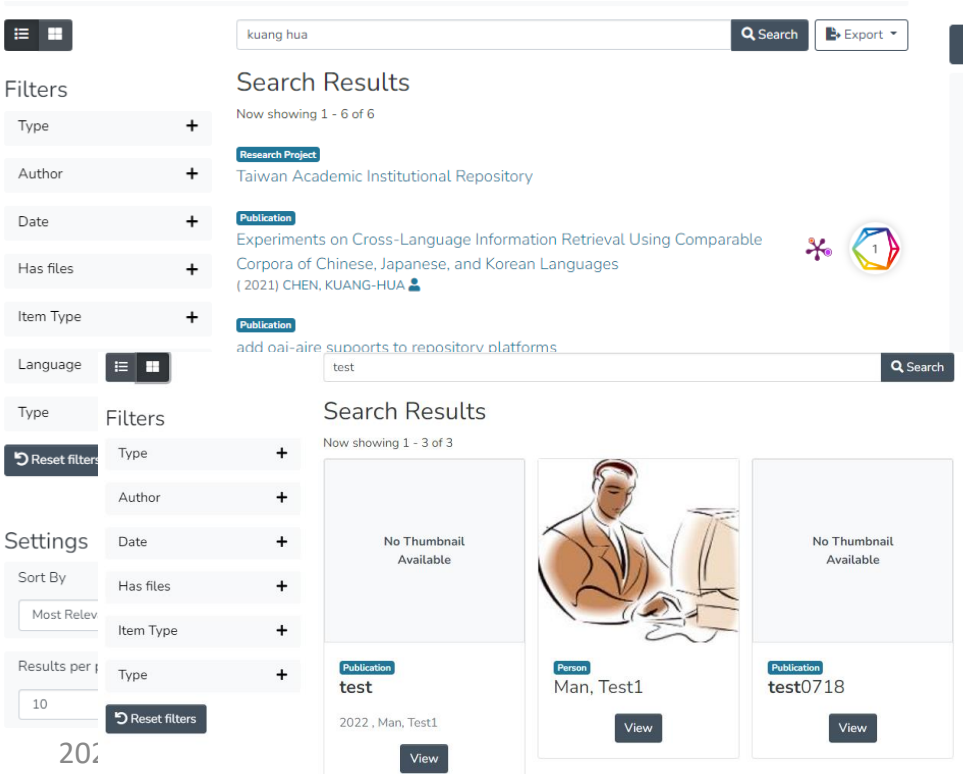
- 視覺化後分類圖表



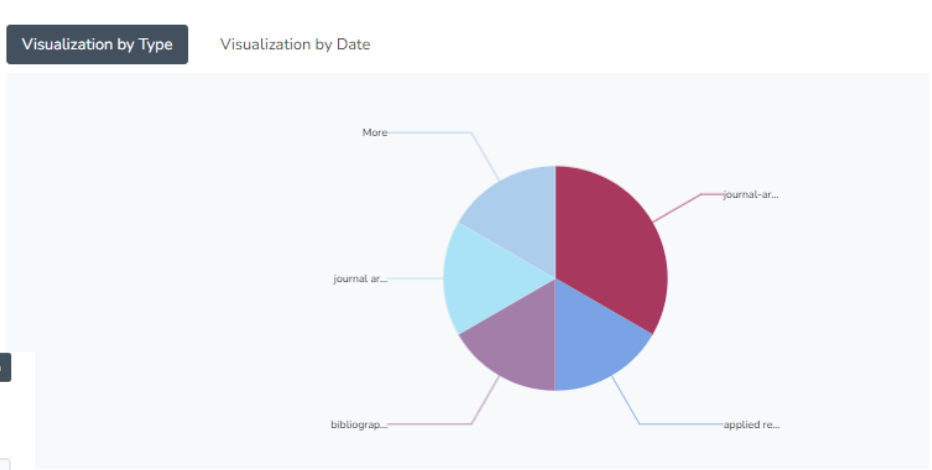
TAIWAN UNIVERSITY LIBRARY

查詢結果頁面

- List/grid



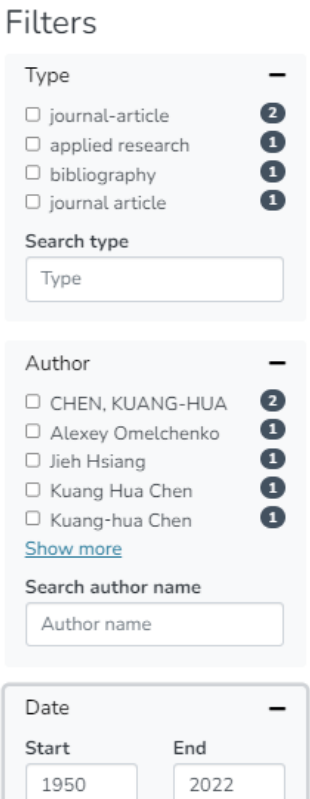
- 視覺化後分類圖表

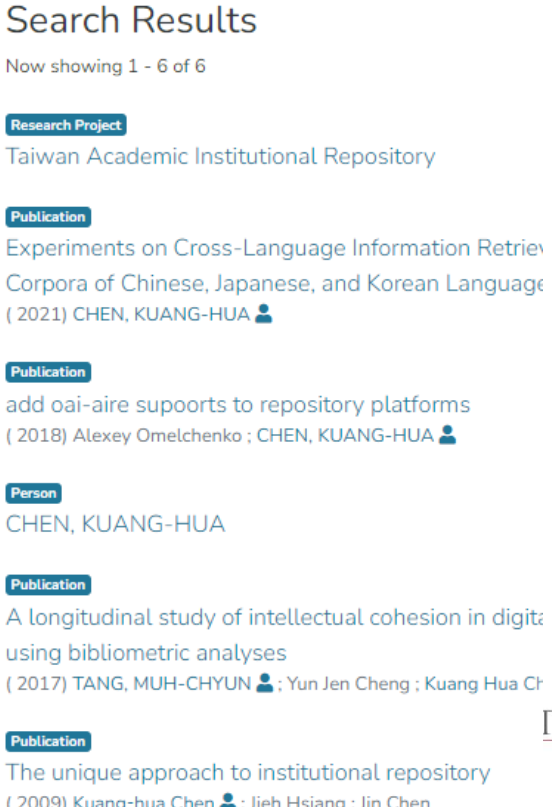


TAIWAN UNIVERSITY LIBRARY

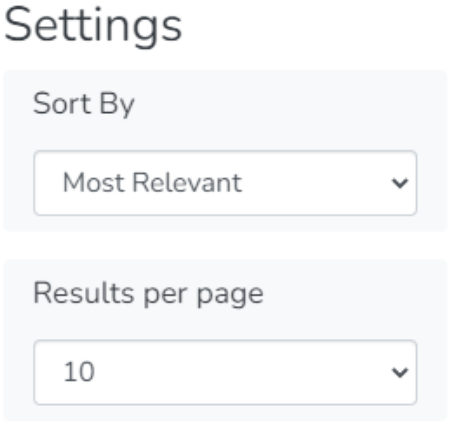
查詢結果頁面

- 後分類篩選機制





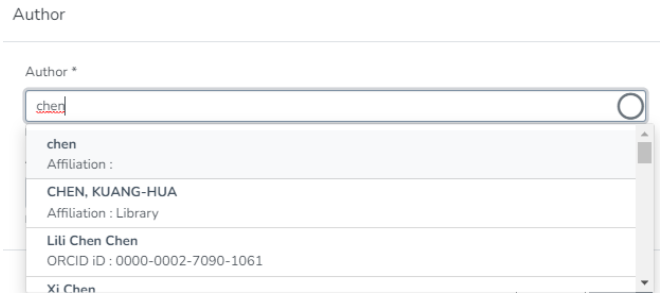
- 排序與呈現筆數設定

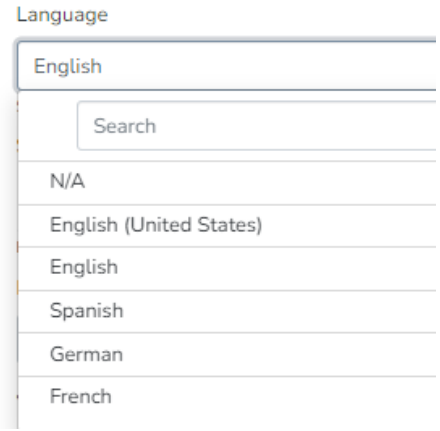


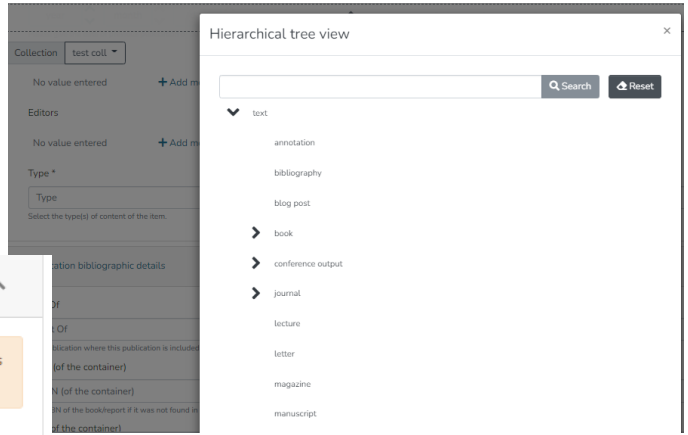
TAIWAN UNIVERSITY LIBRARY

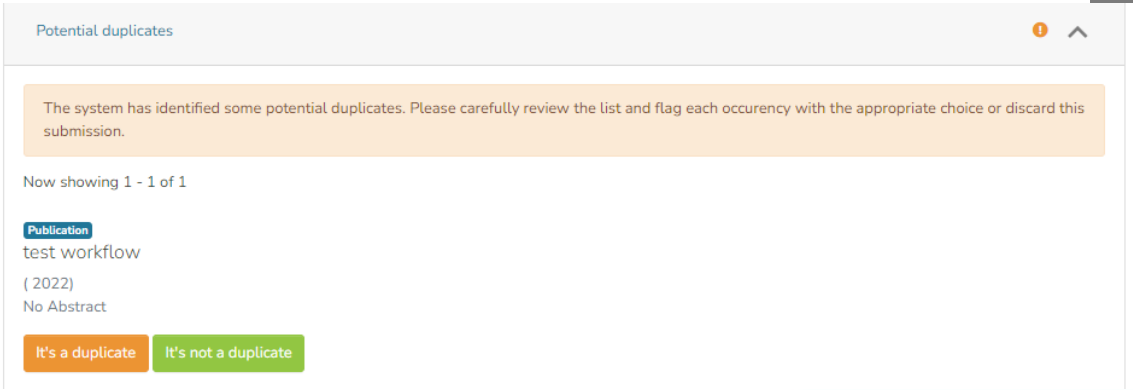
Metadata資料編輯

- 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



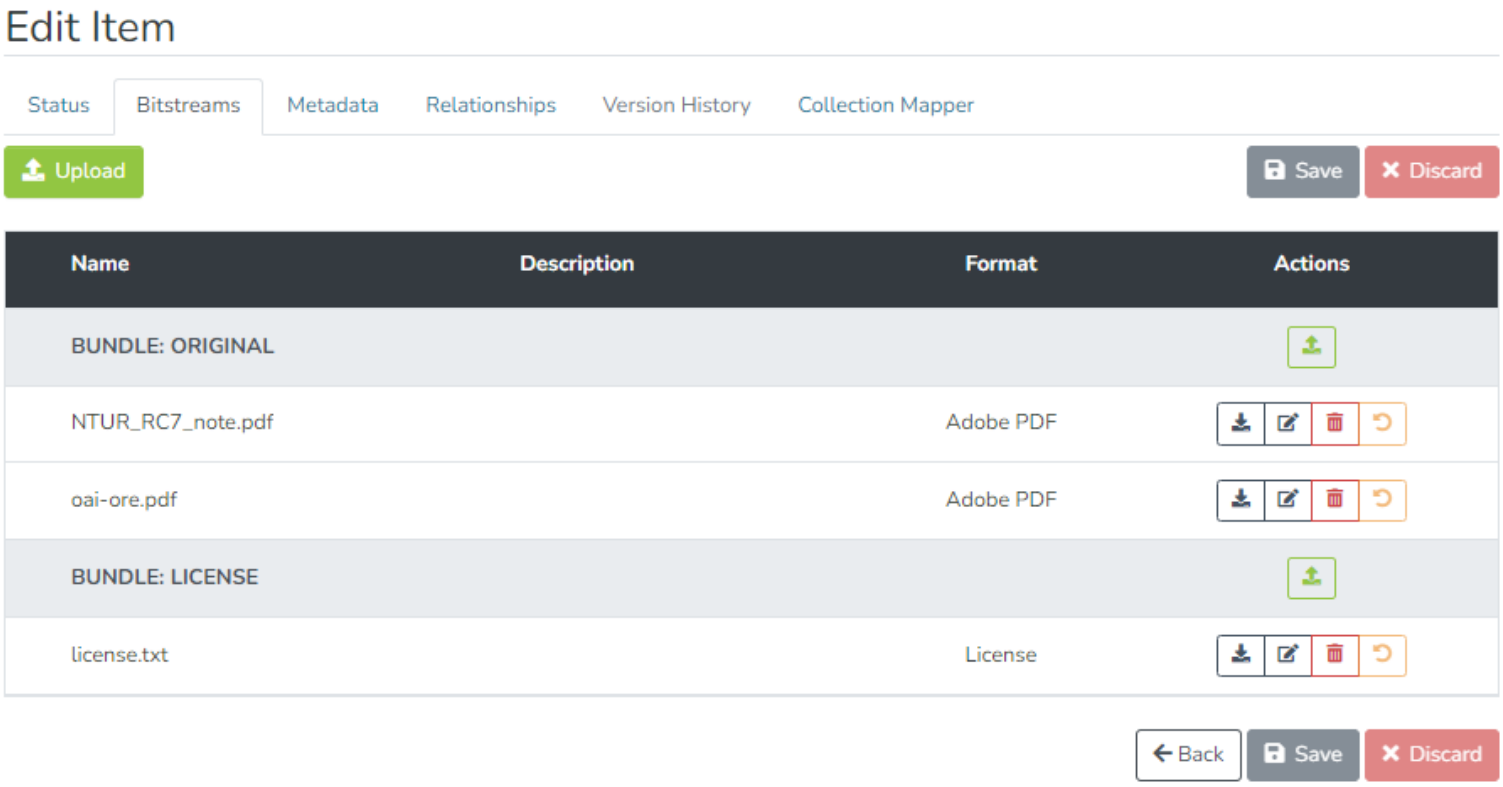






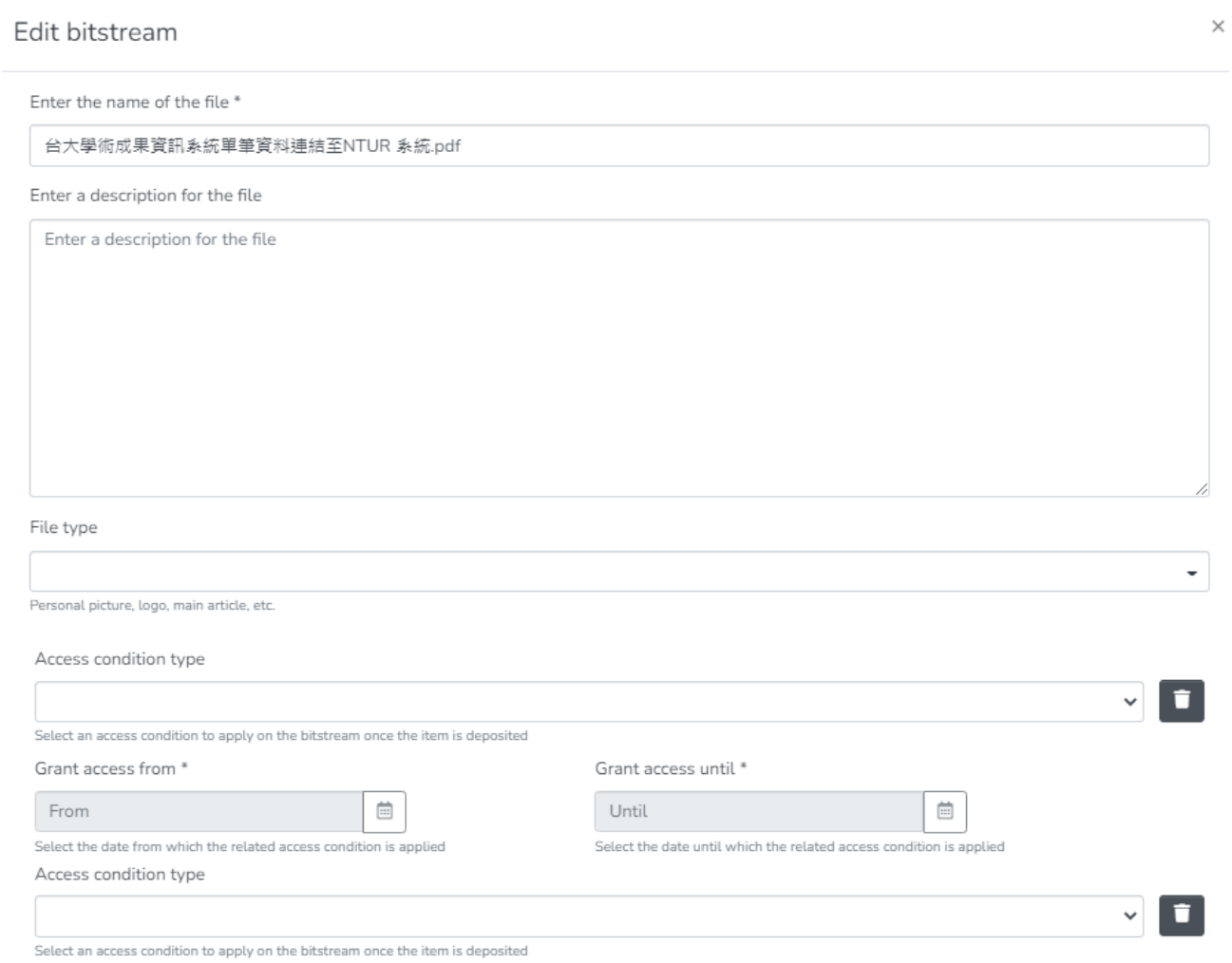
JNIVERSITY LIBRARY

Metadata資料編輯



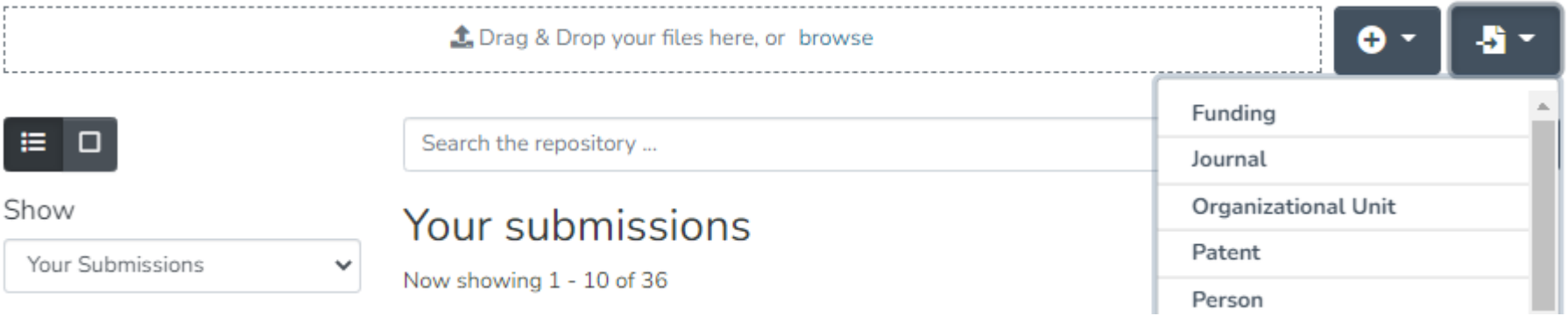
檔案編輯

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



MyDSpace

- 上傳功能列

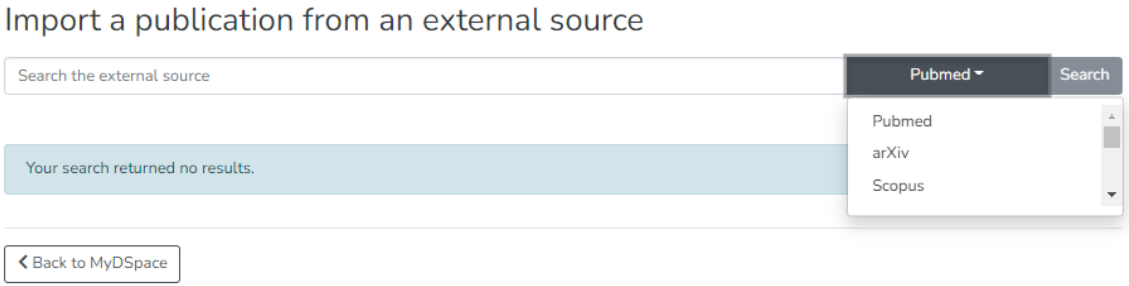


- 文件審核工作流程

NATIONAL TAIWAN UNIVERSITY LIBRARY

文件匯入/匯出

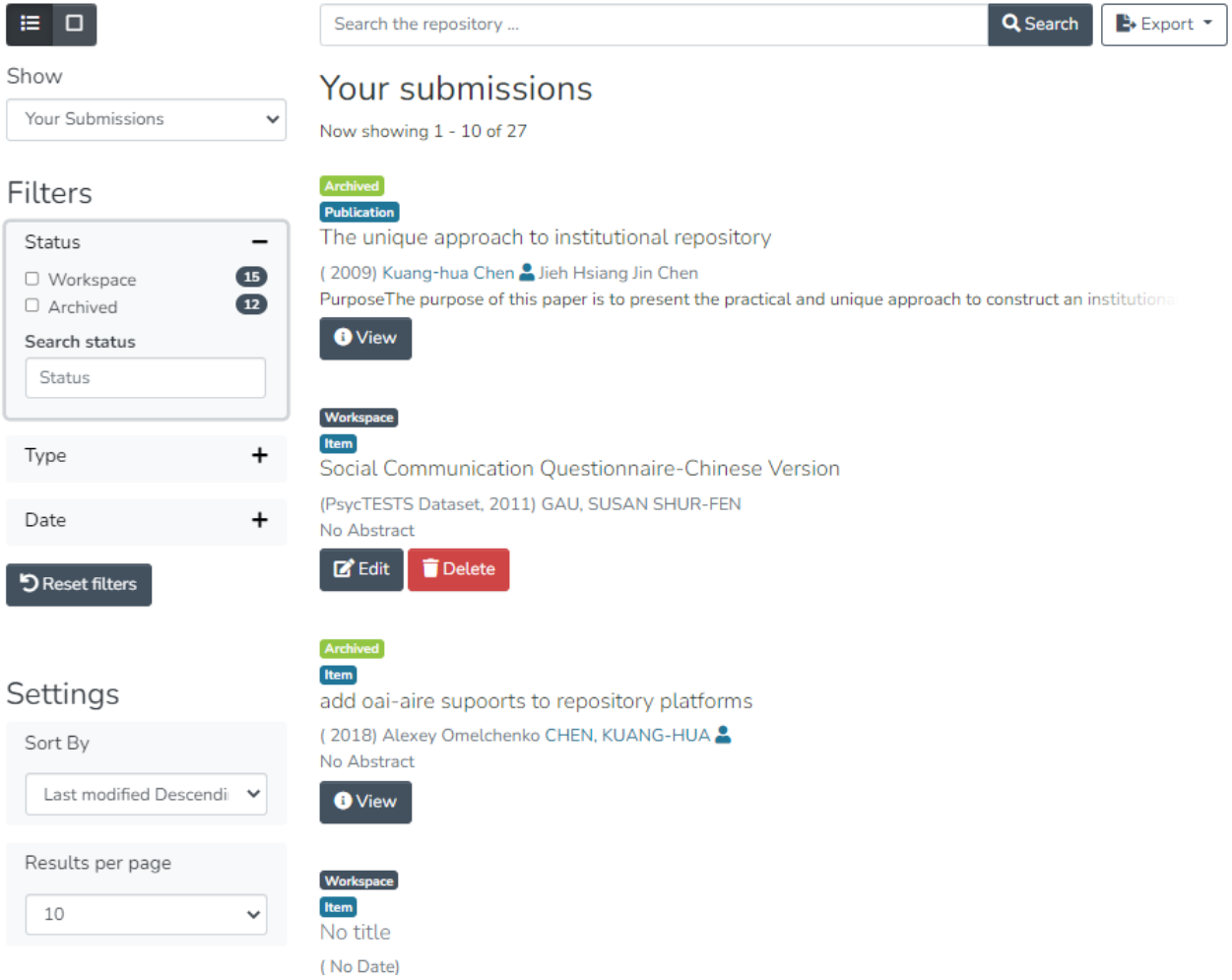
- 單筆資料編輯
- 外部資料上傳
- 多種格式支援
  - SAF：支援物件關聯編輯
  - CSV
  - ....
- 上傳流程(workflow)
- 指定文件/檔案存取政策



NATIONAL TAIWAN UNIVERSITY LIBRARY

MyDSpace

- 可查詢之文件列表
  - 歷史上傳文件
  - 待審核文件



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)

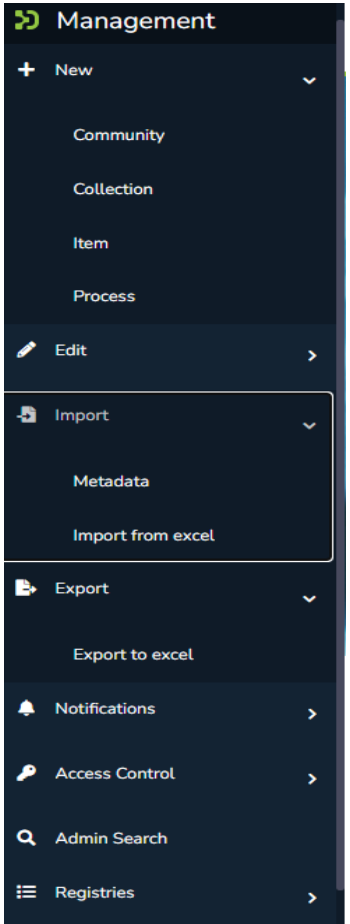
統計機制

- 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

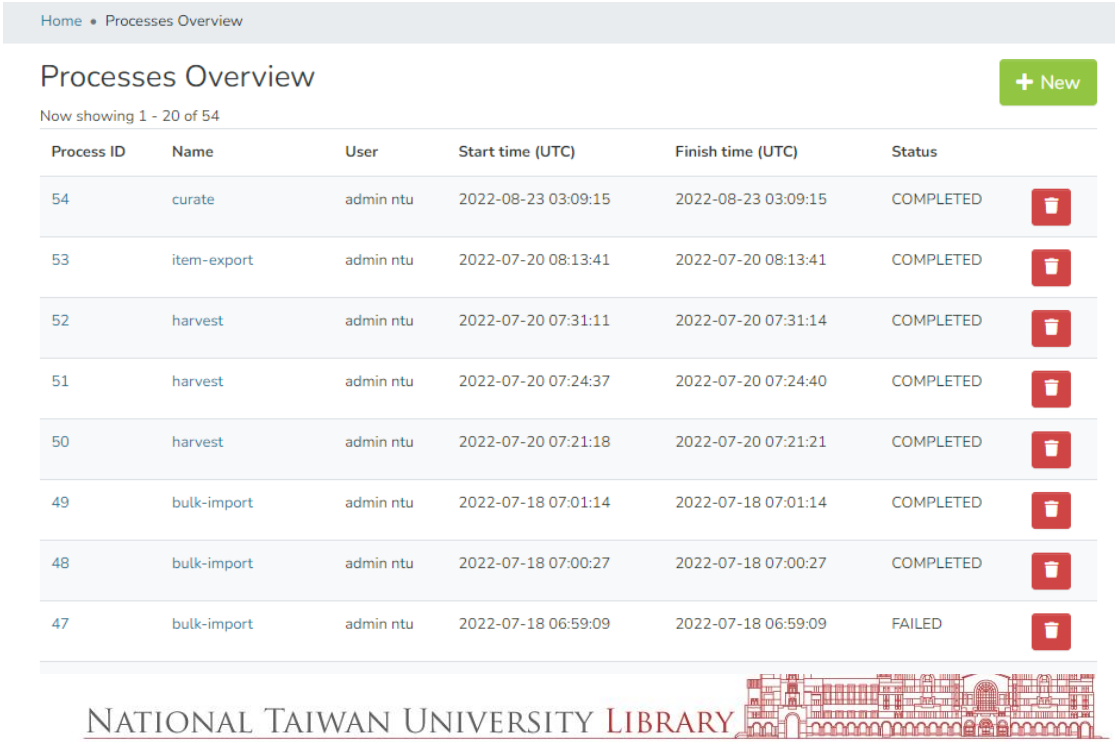
統計機制

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

管理功能

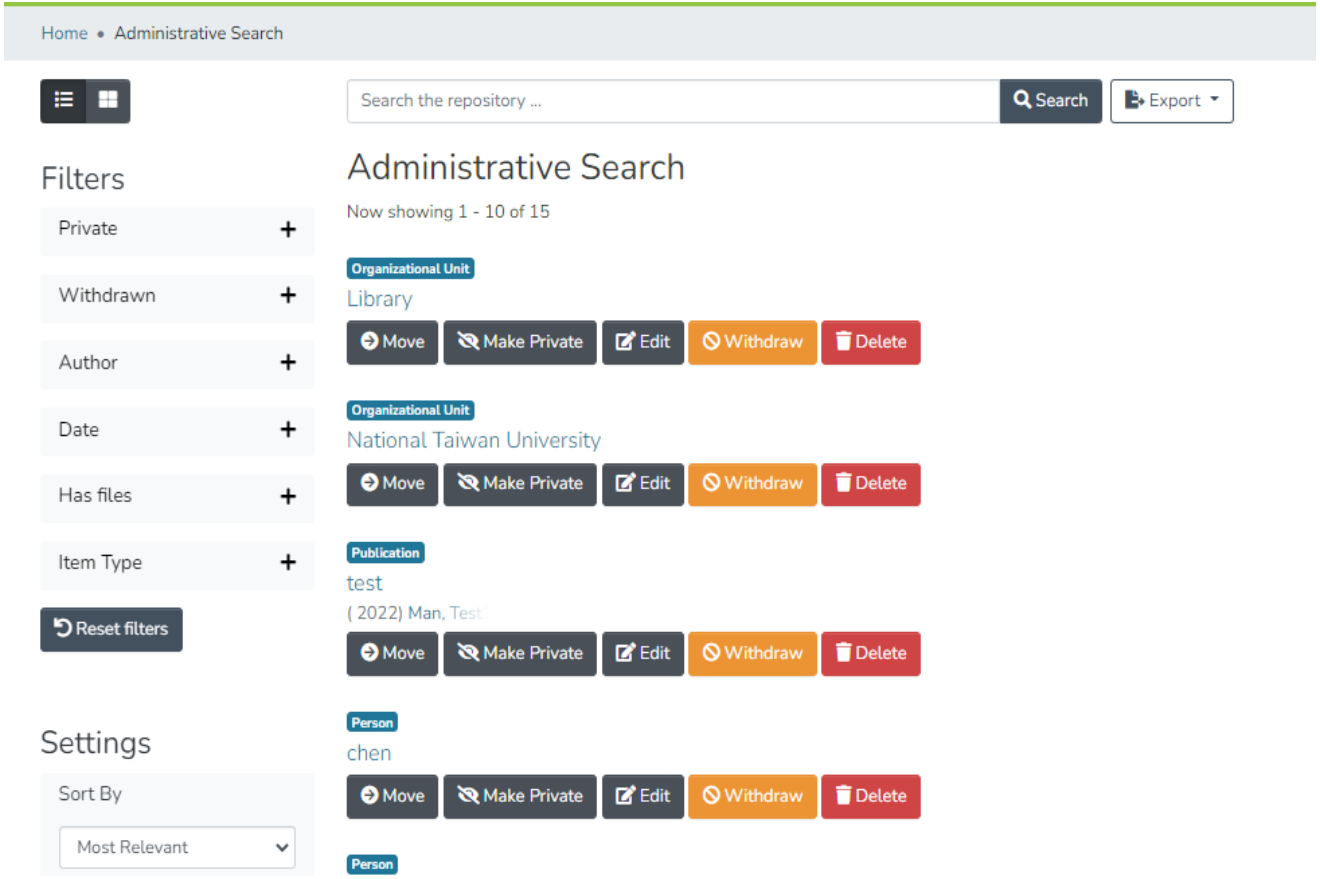


- ←固定式管理功能列



管理功能

- 文件查詢
- Curation Tasks
- 管理者代登入



背景程序執行與管理

Create a new process

Script

metadata-deletion

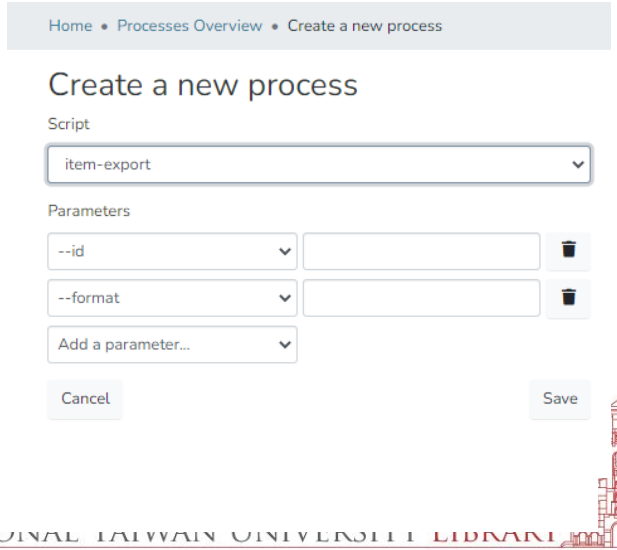
- Choose a script...
- process-cleaner
- bulk-import
- collection-export
- metadata-deletion
- metadata-export
- metadata-import
- harvest
- filter-media
- import-nbevents
- orcid-bulk-pull
- orcid-bulk-push
- import-oaire-suggestions
- update-item-references
- item-enhancer
- bulk-item-export
- item-export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

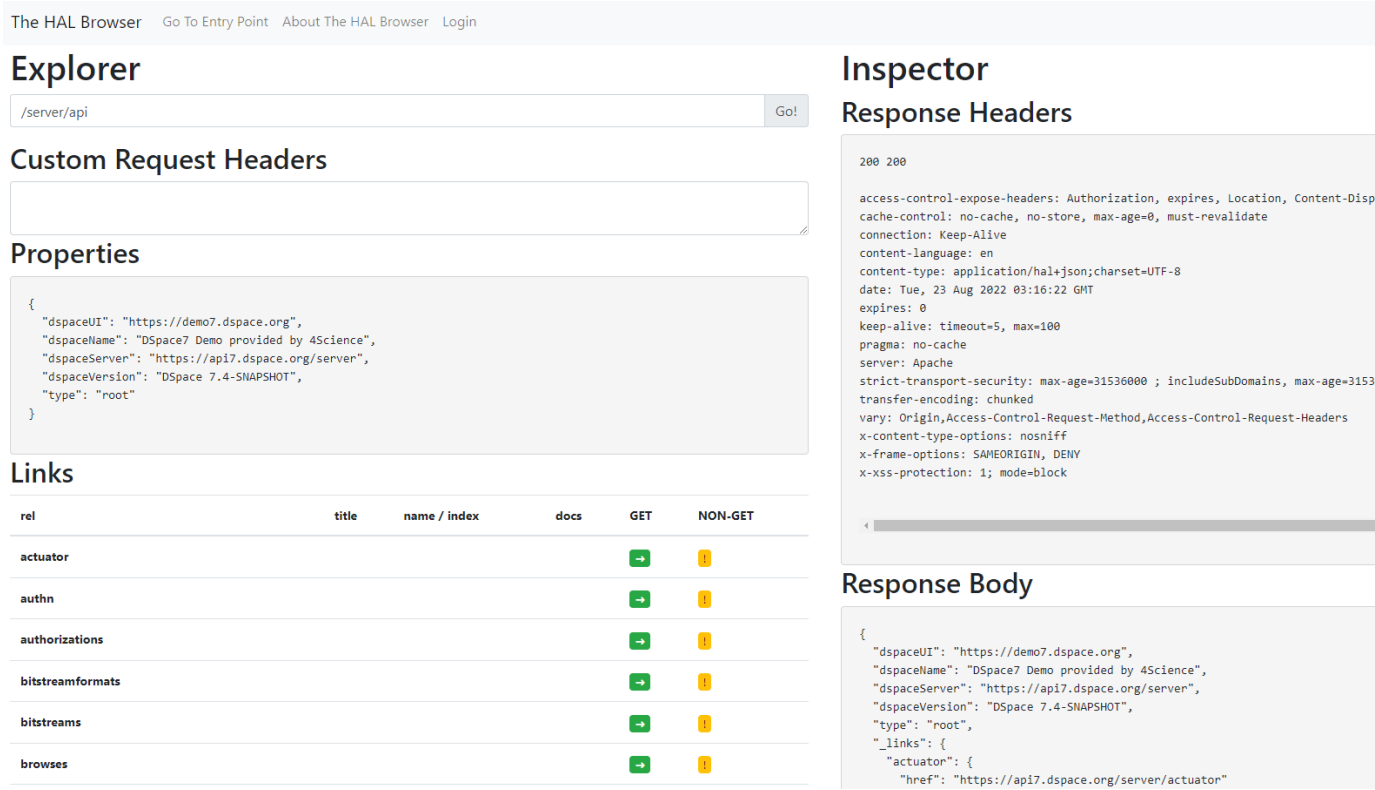


後端：整合式資料存取平台

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

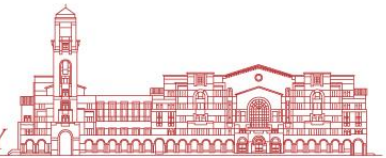
REST API

- For Angular
- 外部系統整合
- HAL browser



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

NATIONAL TAIWAN UNIVERSITY LIBRARY

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

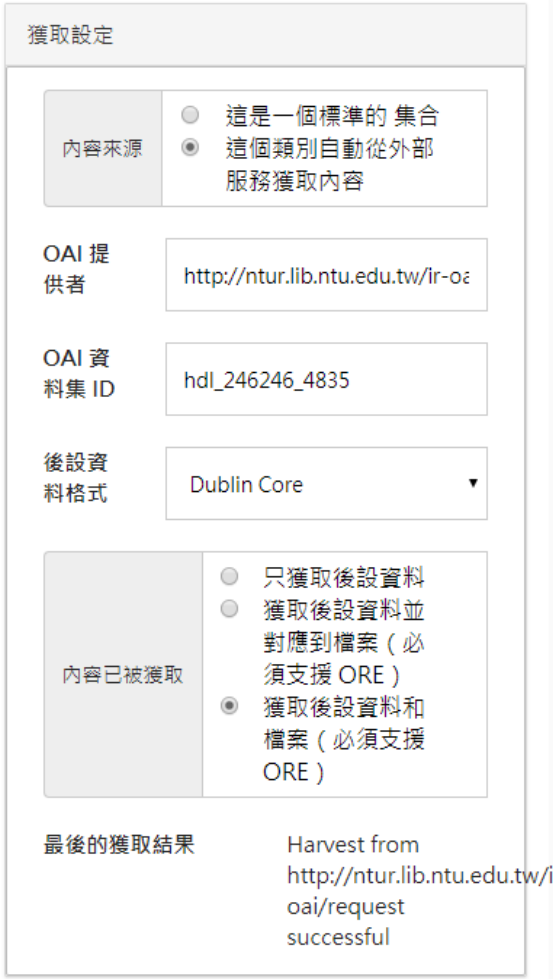
OAI資料同步

- 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]
```



NATIONAL TAIV

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
```

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 schwarzenegg	9/9/1850	golden state	33871648
8	Colorado.pdf	Colorado	bill ritter	8/1/1876	centennial state	4,861,515

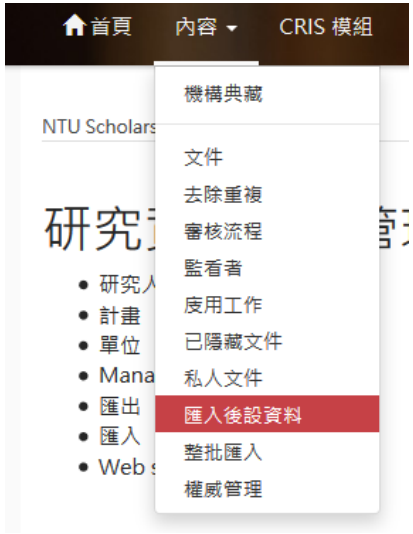
- 輸出：SAF格式

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]
```

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
```

感謝聆聽　敬請指教

NATIONAL TAIWAN UNIVERSITY LIBRARY

SINCE 1928

國立臺灣大學圖書館