XLS Import and Export tool

The purpose of this section is to describe the Export/Import (Excel 97/2003–xls) web functionalities and the format of Excel document necessary for ingesting (also used for export) in DSpace-CRIS.

Export

The export functionality is available on Administrative section of DSpace-CRIS at:

<dspace-url>/cris/administrator/index.htm

The export page shows two boxes to set the query and the entity type of the entities used for export. You can choose only one entity type of DSpace-CRIS per export (e.g. you can export ResearcherPage in the single export but not ResearcherPage and OrganizationUnit together; To export the full repository needs to export one by one the entities type).

The export is available also via batch script:

```
dspace dsrun org.dspace.app.cris.batch.ScriptCrisBulkExport -e <rp|pj|ou|events|journals|..> -f
<export_destination_file>
```

Import

The import functionality is available on Administrative section of DSpace-CRIS (like export) and via Batch Script:

```
dspace dsrun org.dspace.app.cris.batch.ScriptCrisBulkChanges -e <rp|pj|ou|events|journals|..> -f
<export_source_file>
```

The web page shows two boxes to setup the import with the entity type and the file to import. The batch script also uses the same options, launch it with **–h** for other information.

Excel (excel 2007) structure

The document (both for export and import) contains two sheets:

- main_entities
- nested_entities

Main entities Sheet:
The first four columns are the same in all exported type of entities (**CRISID**, **UUID**, **SOURCEREF**, **SOURCEID**) while the other columns depend on the specific entity configuration.

Nested entities Sheet:

The *nested_entities* sheet contains the nested object owned by the object defined in *main_entities* sheet. In fact the first three columns are **CRISID_PARENT**, **SOURCEREF_PARENT**, **SOURCEID_PARENT** that are related to the entity contained in the *main_entities* sheet. The system during the ingest process will process first the main_entities sheet and then the nested_entities sheet, in order to make available the parent object when importing the nested objects. The other three columns **UUID, SOURCEREF, SOURCEID** are mandatory for the file structure. The other columns depend on the specific nested entity configuration.

Only import needs an **ACTION** and a **NONE** column to manipulate the entities. Below is the list of possible actions:

CREATE, UPDATE, DELETE, HIDE, SHOW

Note that **ACTION** column **MUST** be the first column of the **main_entities** sheet and **NONE** column the last, the **NONE** column must be filled with '#' for all entities .

- CREATE build a brand new entity, if the object is founded in the repository works as UPDATE
- UPDATE* try to update the existent entity founded on database. The entity is retrieved using the **CRISID**, **UUID**, **SOURCEREF**, **SOURCEID**, if no entity is found on the database OR there are no values in **CRISID**, **UUID**, **SOURCEREF**, **SOURCEID** cells then it works as CREATE action
- DELETE***: remove the entity and its nested object from repository
- HIDE**: set the public visibility to hide, if the object is not found works as CREATE
- SHOW**: set the visibility to public , if the object is not found works as CREATE

*Please note that if you don't want modify a metadata you can remove the column from the excel file for import, but please make attention because do not recognize the column with a new type of metadata not present in the DSpace-CRIS at moment of the import

**Please note that this action works in append mode this means that if you export an entity and add the action column with HIDE/SHOW it will double every metadata (except for the related nested object that works different). If you want just to change the visibility of the entity you should blank the cells and make action with HIDE/SHOW value

***Please note that to remove only nested you have to empties all cells related to the nested row except identifiers column from (column from A to F in the Nested Entities Sheet).

KNOWN ISSUE: https://github.com/4Science/DSpace/issues/14 and https://github.com/4Science/DSpace/issues/13

Below there are the requirements for the values in the cells:

- visibility: PUBLIC or HIDE; is an attribute used in every metadata/cell (not Mandatory, PUBLIC is default)

LINK

```
[visibility=<ATTRIBUTE> URL=<ATTRIBUTE>]<VALUE>
```

ATTRIBUTE:

- URL: is the value link (in the example below "URL= http://www.personalpage.com /~mypage/")

VALUE:

- <VALUE > in the example below "Personal Page" is the link description saved into database

Example:

```
[visibility=PUBLIC URL=http://www.personalpage.com /~mypage/]Personal Page
```

POINTER and CLASSIFICATION

```
[visibility=<ATTRIBUTE> CRISID=<ATTRIBUTE> SOURCEID= <ATTRIBUTE> SOURCEREF=<ATTRIBUTE> UUID=<ATTRIBUTE>]<VALUE>
```

ATTRIBUTE:

- CRISID: identifier for the pointer to a DSpace-CRIS entity
- UUID: Universally Unique Identifier for the pointer to a DSpace-CRIS entity
- SOURCEID and SOURCEREF: external identifier for the pointer to a DSpace-CRIS entity

NOTE that the UUID or CRISID or the couple SOURCEID/SOURCEREF are mandatory in order to create the relationship.

VALUE:

- <VALUE> is not used in import.

Example:

```
[visibility=PUBLIC CRISID=ou00024 SOURCEID=1 SOURCEREF=external-database]Department of Logistics and Maritime
Studies
```

FILE

```
[visibility=<ATTRIBUTE>LOCAL=<ATTRIBUTE>]<VALUE>
```

ATTRIBUTE:

- LOCAL: true/false; false if the file is not on the DSpace-CRIS file system, true otherwise*
- DELETE: true for remove the file related the property definition defined

VALUE:

- <VALUE> the absolute path on the local machine to the file**

*Please note that if local attribute is false then the file will be retrieved from the local absolute path on server machine

**Please note that the path MUST be visible to the user that launch the batch script or the Tomcat users

Example:

[visibility=PUBLIC LOCAL=false]/path_to_file /image.jpg

TEXT and other types

```
[visibility=<ATTRIBUTE>]<VALUE>
```

VALUE:

- <VALUE> The value saved into database