EJ Website content migration

Common approach

There will be a specific process/script to retrieve the following web assets:

- Documents
- Press releases
- News items

The process shall only retrieve the last published version of the web assets. Past versions or draft versions shall not be exported.

The export process shall include an option to filter by a range of dates.

The export process will export the documents/images to the local file system and will generate a csv file with all the metadata.

The csv file shall include additional columns to specify the path where the documents and images are stored after exporting data.

The process will generate a log file with the export operations: exported items or unexpected events when processing a particular document or page.

Document library migration

This particular script will go through the libraries under the "Document library" and retrieve documents (files) and the associated metadata.

The export process shall retrieve all documents, regardless of the date, even though there is the possibility to filter by date range.

The process will extract all the documents from all the sites and subsites in the "Document Library" area.

- · File Name
- · Title (Document Set content type)
- · Date Created
- · ~~Created by (user)~~ > To exclude this field to ensure data protection compliance.
- · Date Last Modified
- · Last Modified (user)
- · Url (full document path)
- · Document Language (Eurojust Document Content-Type)
- · ~~Document Description (Document Set Content-Type)~~ > This field might not be useful for the new website.

For the document title, exported from the document set title, it should in all cases include the full title, including the date in brackets for lots of documents.

Eurojust framework

Casework publications

- Questionnaire CJEU judgments regarding EAW - Executive Summary (pdf, 154 Kb)
  bg cs da de el en es et fi fr ga hr hu it lt lv mt nl pl pt ro sk sl sv

- Questionnaire CJEU judgments regarding EAW (March 2020) (docx, 6124 Kb)
  en

- Second report of the observatory function on encryption (joint Europol-Eurojust report - January 2020) (pdf, 1093 Kb)
  en

- Questionnaire on the impact of the EAW ruling (November 2019) (MS Word version) (docx, 4877 Kb)
  en

- Guidelines for deciding on competing requests for surrender and extradition (October 2019) (pdf, 219 Kb)
  bg cs da de el en es et fi fr ga hr hu it lt lv mt nl pl pt ro sk sl sv

---

EJ: We will need to process this of course, but beyond the mechanic last updated / created dates, we will need to consider this as the document date

Note the date information included in the title could go with other information) so it should be parsed and processed (either in the export or in the import process) if you want to automate this.

To simplify the process, we propose you to use the date created field (as it is automatically created) and if needed, it could be manually updated.

The import process shall be done in a second phase. The documents shall be created as web contents in Drupal. Based on this, whenever a legacy document is uploaded, we can specify the language version in which is stored.

Press releases / News Items

There will be a separate script to export the press releases and another for the news items, retrieving the following fields:

- Url

- · Name
- · Title
- · Page Image (Image)
- · Page Content (HTML with images)
- · Article Date
- · Short Page Content

There will be a CSV file including the above fields plus the images (actually videos, infographics, etc.) referenced in the page content.

Images will be extracted in the same process, and extracted as

The process shall retrieve only the past three years (i.e. 2018-2020). Previous contents shall be archived in a third party solution.

Howevert, the script shall consider the possibility to export data between a given period.

Sitemap.xml

There will be a specific script (based on the existing webpart to build the sitemap page) to retrieve a the sitemap as an XML file.

This file shall summarise per section, a list of pages including the title and the url.

And the sitemap.xml would be similar to this:

```
<Sitemap>
    <Section title="About us">
        <Subsection title="Background">
            <page title="History" url="http://eurojust.europa.eu/about/background/Pages/history.aspx"/>
            <page title="Mission and tasks" url="http://eurojust.europa.eu/about/background/Pages/history.aspx"/>
            <page title="Eurojust case business" url="http://eurojust.europa.eu/about/background/Pages/eurojust-core-business.aspx"/>
            <page title="Purpose made building" url="http://eurojust.europa.eu/about/background/Pages/eurojust-core-business.aspx"/>
        </Subsection>
        <Subsection title="Legal framework">
            <page title="Eurojust legal framework" url="http://eurojust.europa.eu/about/legal-framework/Pages/eurojust-legal-framework.aspx"/>
            ....
        </Subsection>
    </Section>
</Sitemap>
```