

BTE transition: 4Science implementation proposal and estimation

Implementation of the start submission via an uploaded file (estimation 3 days - required previous)

This task is related to rewriting the `org.dspace.app.rest.repository.WorkspaceItemRestRepository.upload(Context, HttpServletRequest, MultipartFile)` to internally use Live Import instead than BTE to eventually extract metadata for one or more submission from the provided file.

Indeed, most of files (BibTeX, EndNote, etc.) allow to describe list of publications a special case could be represented by a PDF that will bring the data only about the related publication.

No rest contract changes are expected so no impact on the angular side as well.

The task will **include the creation of a first example of data extraction from a file** via the Live Import Framework, such as BibTeX or EndNote, and corresponding Integration Test that will verify that the data are correctly added to the created items

Implementation of the Extract Panel using the live import framework (estimation 2 day - require previous)

This task is related to the step `org.dspace.submit.step.MetadataStep`, `org.dspace.app.rest.submit.step.ExtractMetadataStep` currently in the master.

According to the javadoc

This submission steps allow to extract metadata from an uploaded file to enrich or initialize a submission. The processing is delegated to a list of extractor specialized by format (i.e. a Grobid extractor to get data from a PDF file, an extractor to get data from bibliographic file such as BibTeX, etc)

the plan is to provide equivalent steps that will use extractors build with the Live Import framework instead than BTE.

The current class should be flagged as deprecated and renamed in

- `MetadataStepDeprecatedBTEStep`
- `ExtractMetadataDeprecatedBTEStep`

This step, that can be enabled or disabled via the `item-submission.xml` has the effect to automatically enrich an inprogress submission when new information that can be used by the extractor (doi, pmid, pdf file, etc.) are added to the workspace (or workflow) item.

The whole feature is already implemented and has been demonstrated at OR and the youtube video of the DSpace 7 preview, there is no interaction with the user about the additional information found in the external provider but the extra changes automatically applied are notified with the usual sticky messages.

Implementation of a search and import from an external source (Estimation REST: 1day, Angular: 4 days) - this is the goal of the original DS-4281

A separate UI linked from the MyDSpace and/or the left menu should allow the user to search an external datasource and import one or more results.

The UI can be organized with panel or tab for each external source as retrieved via the endpoint

<https://github.com/DSpace/Rest7Contract/blob/master/external-authority-sources.md>

and allow the user to perform a search, visualize the result, select one or more via a checkbox and import them in a specific collection (using the same dropdown that will be introduced in the `myspace/admin` menu) starting a new submission.

On the rest side **this task will include the porting to the live import of the CrossRef BTE connector** that also provide the ability to search by query other than by identifier.

Implementation of a metadata quality tool for inprogress submission (TBD after 7.0)

This is a new feature, that was initially discussed and named metadata suggestions and confused with the search functionality above.

For this feature we ask to have a wider discussion in the full team and temporary withdrawn/put on hold the contract (<https://github.com/DSpace/Rest7Contract/blob/master/metadata-suggestions.md>) that was approved in the Entities Working Group. This discussion can be started once that the version 7.0 is out or at least we are sure that discussion this new feature will not delay more urgent work.

Some initial thoughts:

- can we design that to fit in the panel/step submission framework instead than use a new completely independent endpoint?
- can we design that to be a bit more general so to have a framework to provide feedback about an existing submission? (antiplagiarism, sherpa/romeo, unpaywall, etc.)?