Archive Extractor How to configure

On this page:

Step 1. Launch Aspire and open the Content Source Management Page

Launch Aspire (if it's not already running). See:

- Launch Control
- Browse to: http://localhost:50505. For details on using the Aspire Content Source Management page, please refer to Admin UI

Step 2. Add a new Content Source

- For this step please follow the step from the Configuration Tutorial of the connector of you choice, please refer to Connector list

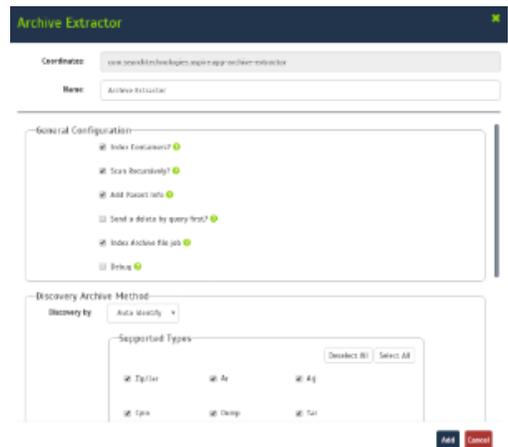## Step 3. Add a new Archive Extractor to the Workflow

To add an Archive Extractor drag from the **Archive Extractor** rule from the *Applications Workflow Library* and drop to the *On Add Update Workflow Tree*. This will automatically open the Archive Extractor window for the configuration.

Step 3a. Specify Archive Information

In the Archive Extractor window, specify the desired options for .

1. General Configuration
    a. Index Containers
    b. Scan Recursively
    c. Add Parent Info
    d. Send Delete By Query first
    e. Index Archive file job
    f. Debug

2. Discovery Archive Method
    a. Auto Identify (Select Supported Types)
    b. Regex

3. Extract Text
    a. Timeout
    b. Size
    c. Disable extraction

4. Routing
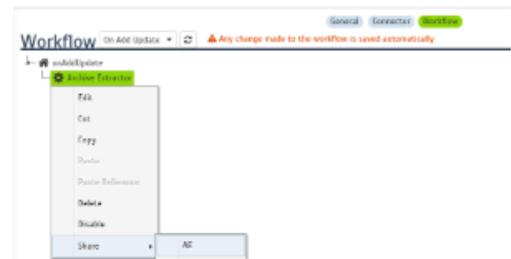   a. Workflow for Add/Update jobs
    b. Workflow for delete jobs

c. Workflow for error jobs

Step 3b. Share the rule into a new Library

Once you save the component, share it in a library (this is required).

Step 3c. Copy the shared rule

Add it into the Delete pipeline (from the shared library, this is required)

> ⓘ In order to extract the content of the files inside the Archive File you need to disable the extract text of the connector and Configure it in the Archive File Component. So you need to add a rule for the extract text of the others jobs from the crawl (you can share the extract text in the same library used before).
>
> You can use some rule like:

Once you've clicked on the *Add* button, it will take a moment for Aspire to download all of the necessary components (the Jar files) from the Maven repository and load them into Aspire. Once that's done, the publisher will appear in the Workflow Tree.

> ⓘ For details on using the Workflow section, please refer to Workflow introduction.