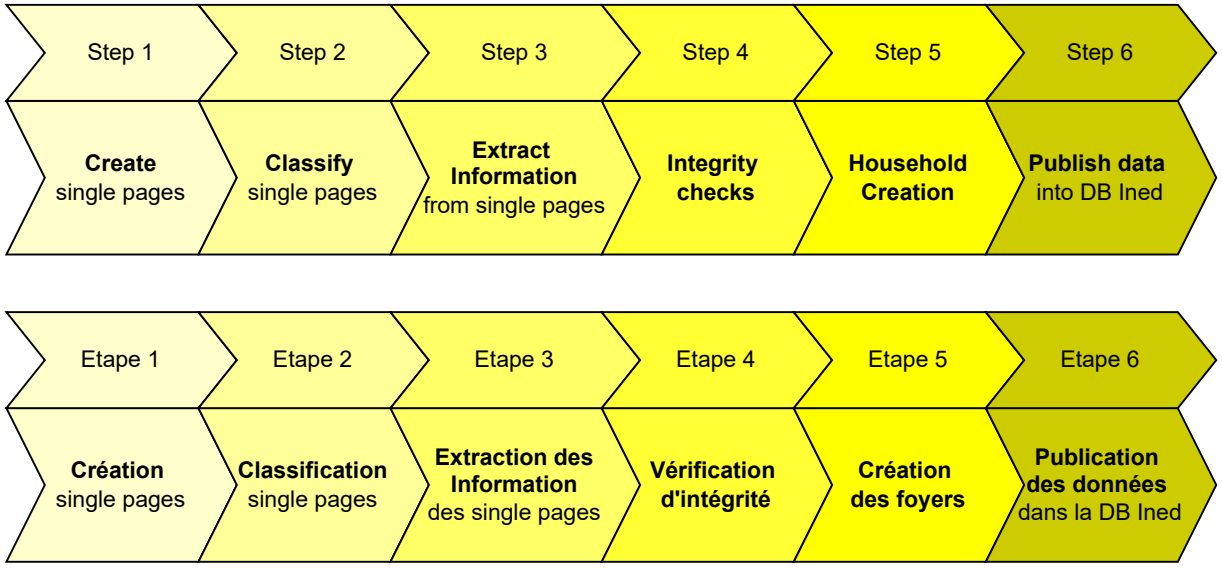
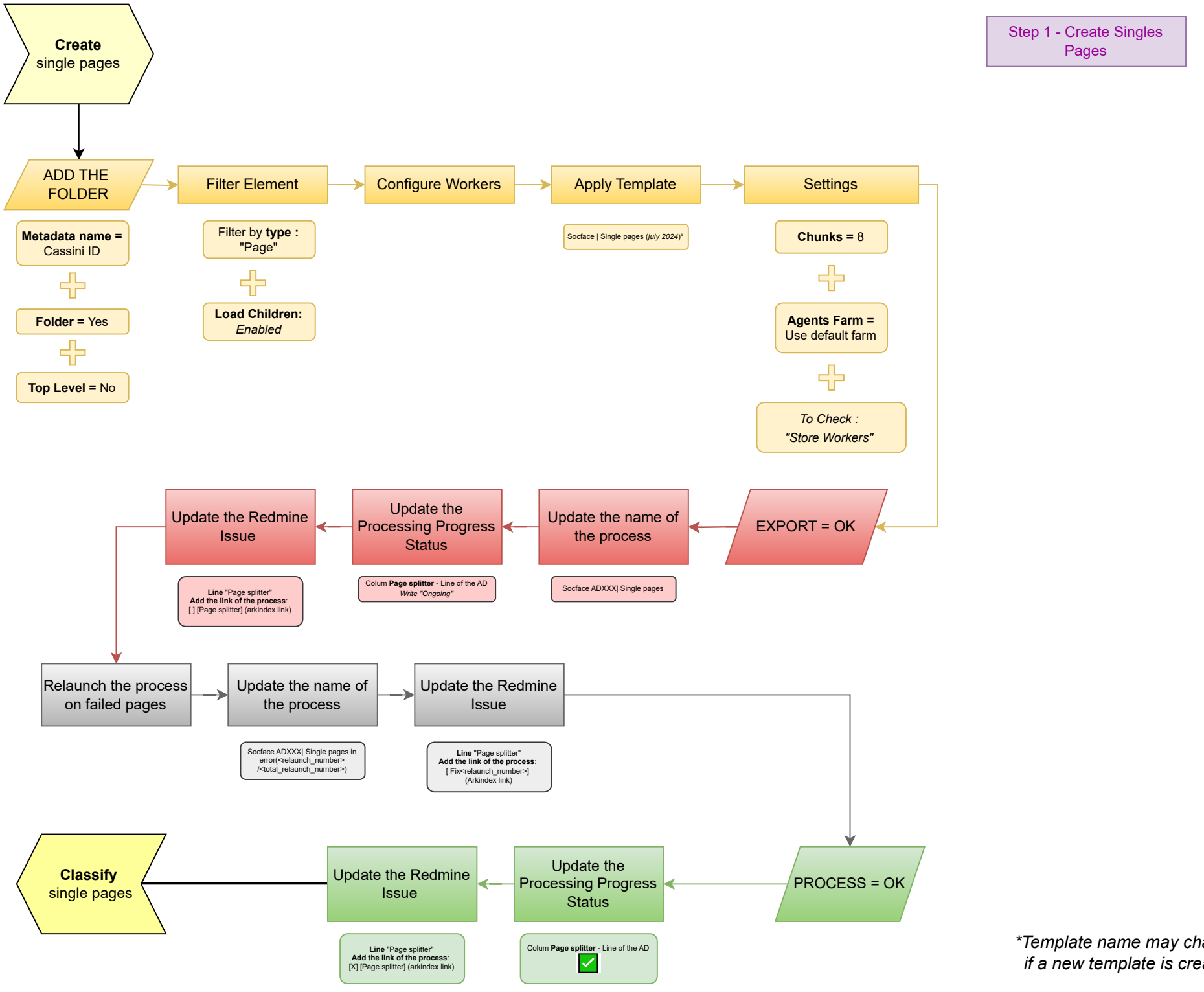
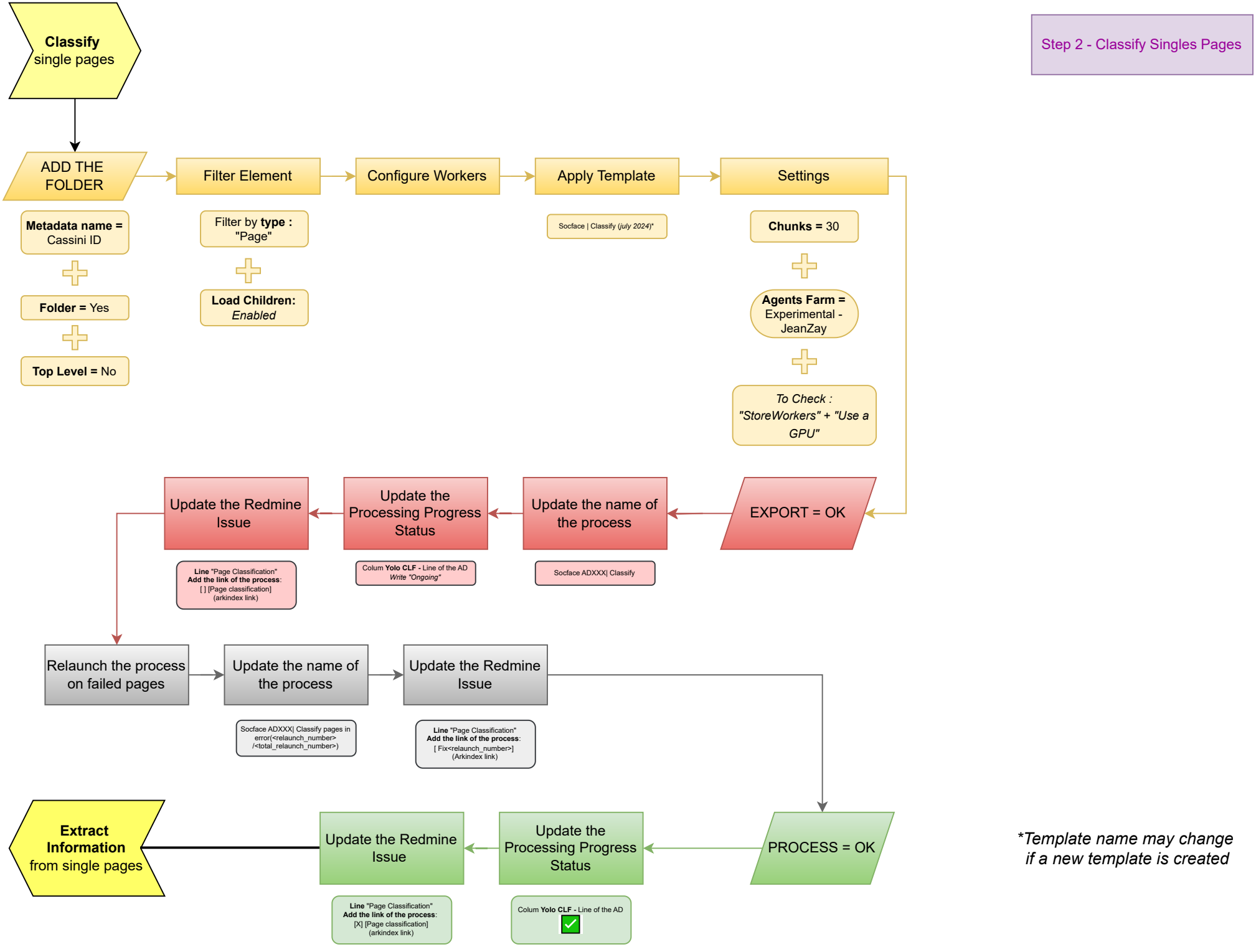
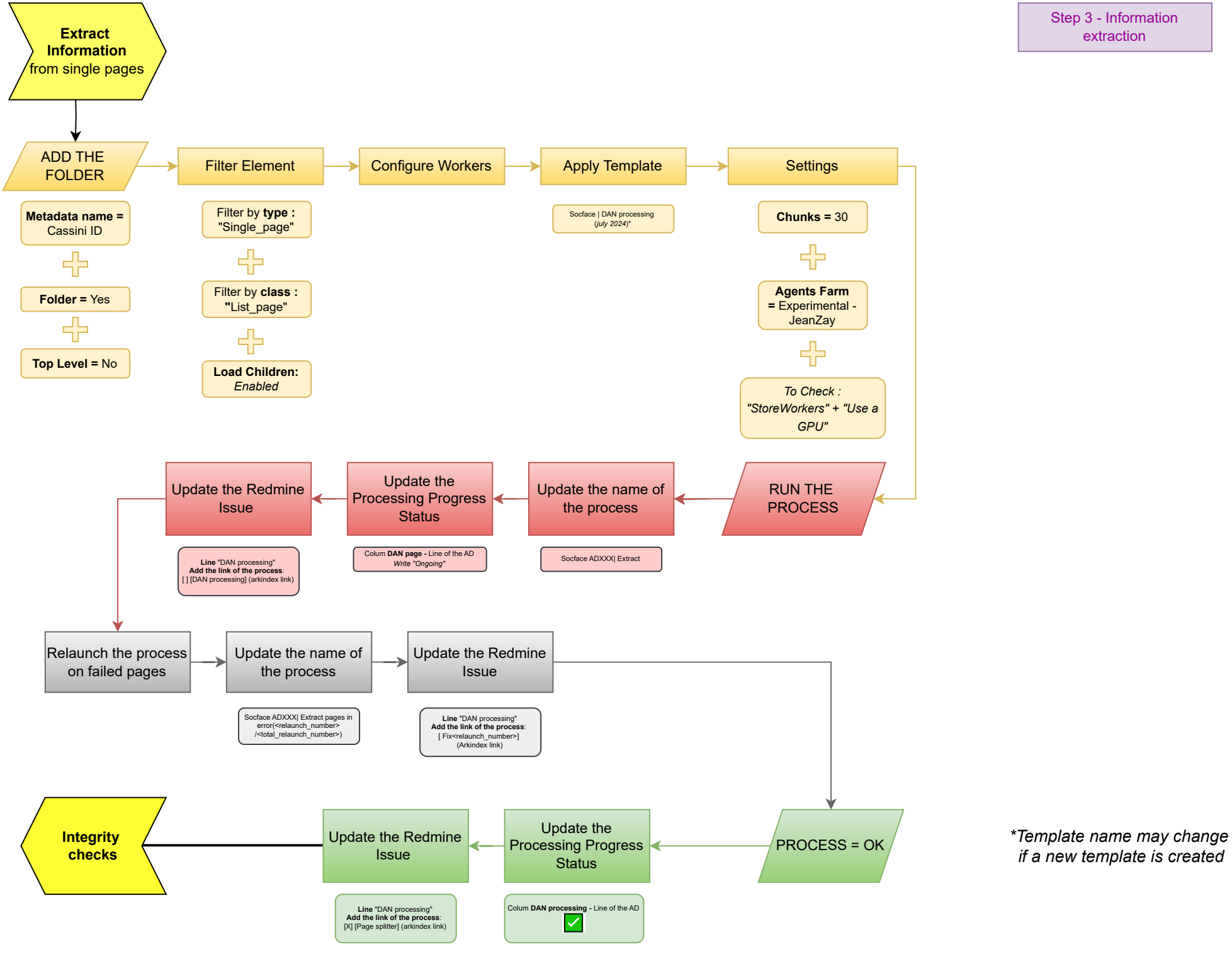
Vue Générale Machine Learning Workflow



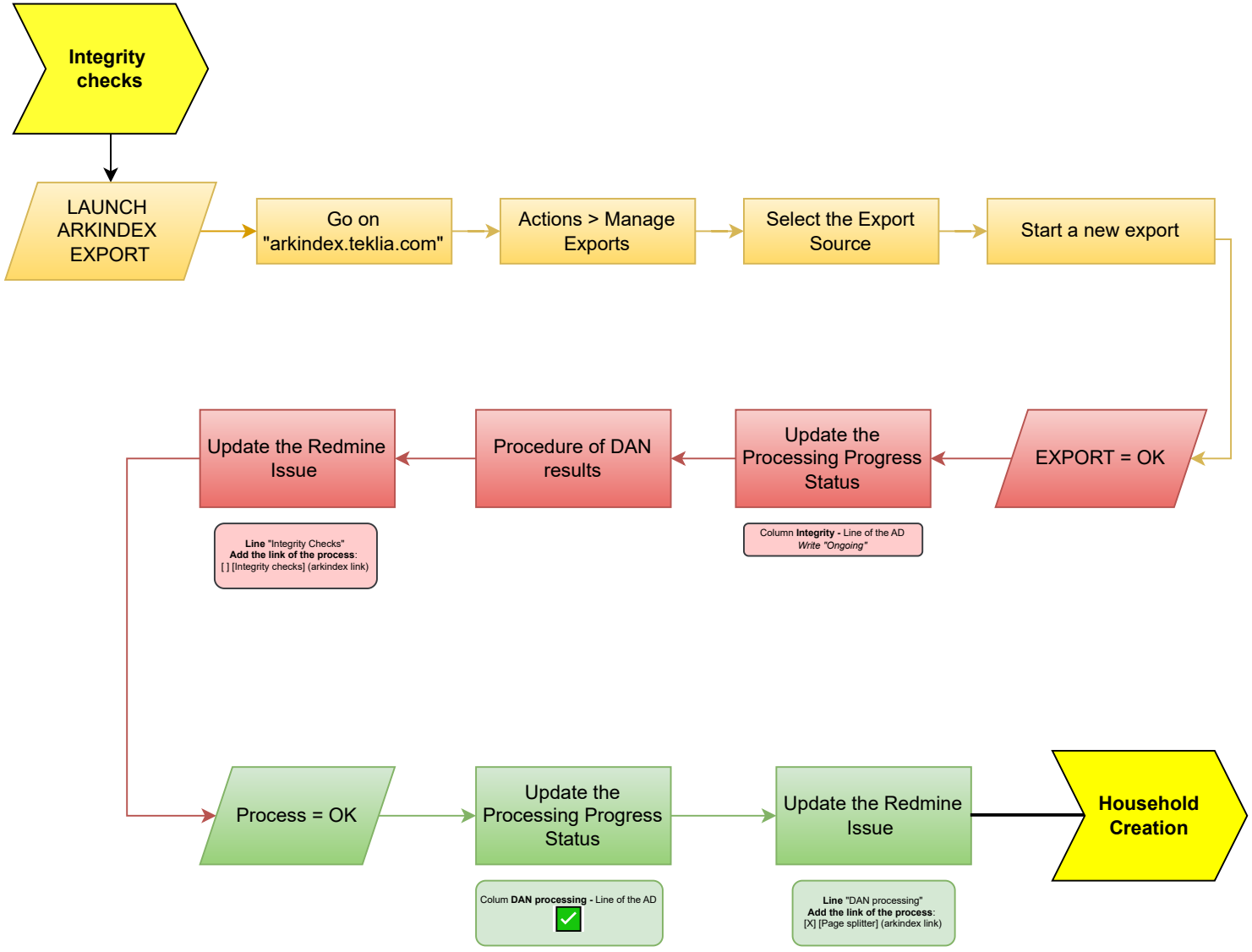

Step 1 - Create Singles Pages
Create single pages
ADD THE FOLDER
Metadata name = Cassini ID
Folder = Yes
Top Level = No
Filter Element
Filter by type : "Page"
Load Children: Enabled
Configure Workers
Apply Template
Socface | Single pages (july 2024)*
Settings
Chunks = 8
Agents Farm = Use default farm
To Check : "Store Workers"
EXPORT = OK
Update the name of the process
Socface ADXXX| Single pages
Update the Processing Progress Status
Colum Page splitter - Line of the AD Write "Ongoing"
Update the Redmine Issue
Line "Page splitter" Add the link of the process: [ ] [Page splitter] (arkindex link)
Relaunch the process on failed pages
Update the name of the process
Socface ADXXX| Single pages in error(<relaunch_number> /<total_relaunch_number>)
Update the Redmine Issue
Line "Page splitter" Add the link of the process: [Fix<relaunch_number>] (Arkindex link)
PROCESS = OK
Update the Processing Progress Status
Colum Page splitter - Line of the AD
Update the Redmine Issue
Line "Page splitter" Add the link of the process: [X] [Page splitter] (arkindex link)
Classify single pages
*Template name may change if a new template is created


Step 2 - Classify Singles Pages
Classify single pages
ADD THE FOLDER
Metadata name = Cassini ID
Folder = Yes
Top Level = No
Filter Element
Filter by type : "Page"
Load Children: Enabled
Configure Workers
Apply Template
Socface | Classify (july 2024)*
Settings
Chunks = 30
Agents Farm = Experimental - JeanZay
To Check : "StoreWorkers" + "Use a GPU"
EXPORT = OK
Update the name of the process
Socface ADXXX| Classify
Update the Processing Progress Status
Colum Yolo CLF - Line of the AD
Write "Ongoing"
Update the Redmine Issue
Line "Page Classification"
Add the link of the process:
[ ] [Page classification]
(arkindex link)
Relaunch the process on failed pages
Update the name of the process
Socface ADXXX| Classify pages in error(<relaunch_number> /<total_relaunch_number>)
Update the Redmine Issue
Line "Page Classification"
Add the link of the process:
[Fix<relaunch_number>]
(Arkindex link)
PROCESS = OK
Update the Processing Progress Status
Colum Yolo CLF - Line of the AD
Update the Redmine Issue
Line "Page Classification"
Add the link of the process:
[X] [Page classification]
(arkindex link)
Extract Information from single pages
*Template name may change if a new template is created

Step 3 - Information extraction

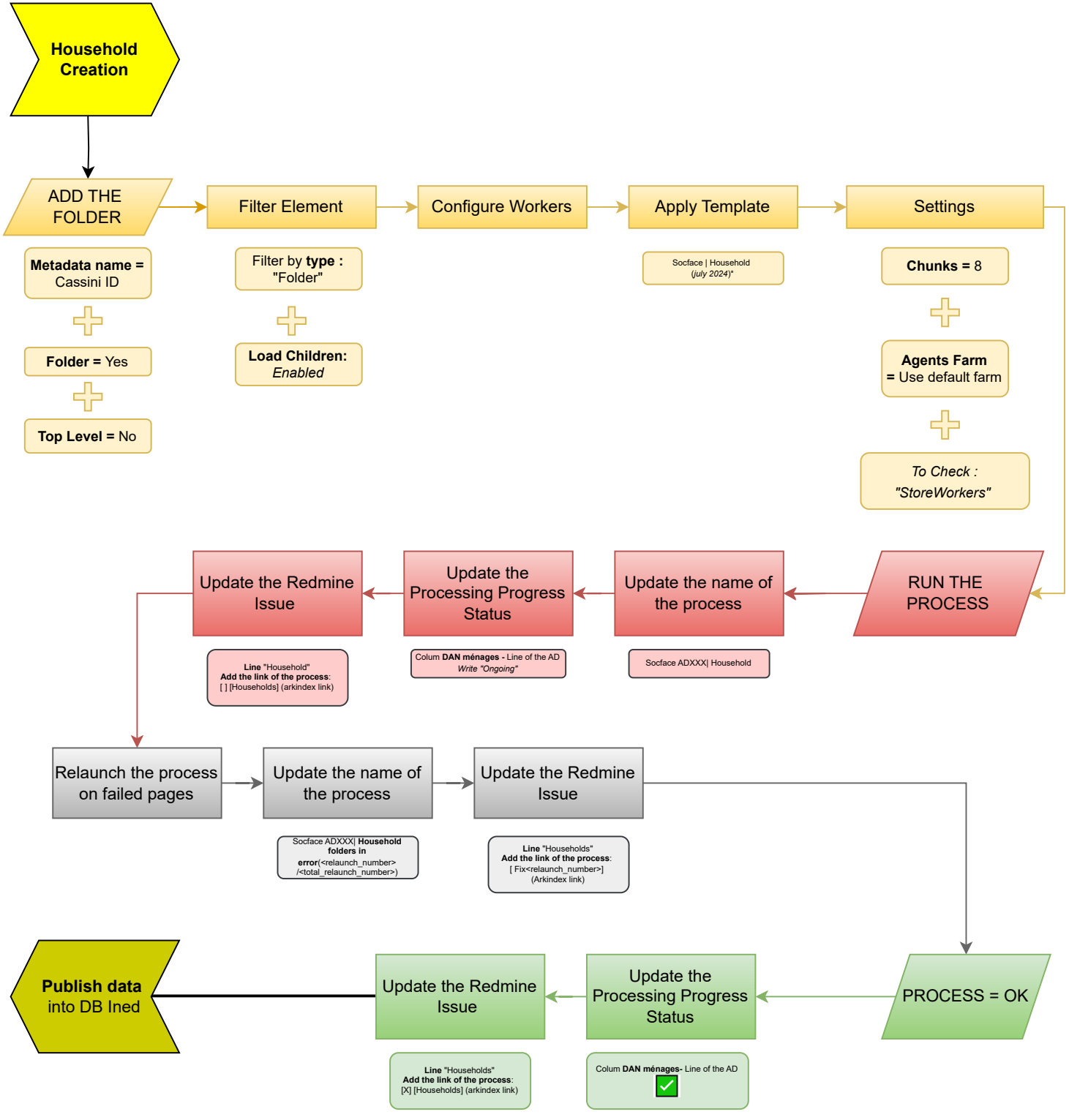


*Template name may change if a new template is created



Step 4 - Integrity Checks

Step 5 - Household creation



Template name may change if a new template is created