

Entity Extraction Workflow

This guide outlines the step-by-step process for using Entity Extraction.

Step 1: Upload Resources to Collections

To extract entities from digital resources, start by [adding digital representations to a collection](#).

Note

Note: Supported formats include JPG, PNG, TIFF, and textual PDF.

Once the titles are uploaded, you can run the Entity Extraction process.

1. Go to **Specto > Data Management > Manage Collections**.
2. In the list of collections, locate the relevant collection that contains digital content.
3. Select the **More Actions** (ellipses) button and select the **Digital Metadata Assistant** option. This is the Alma Specto homepage.

Step 2: Run Entity Extraction

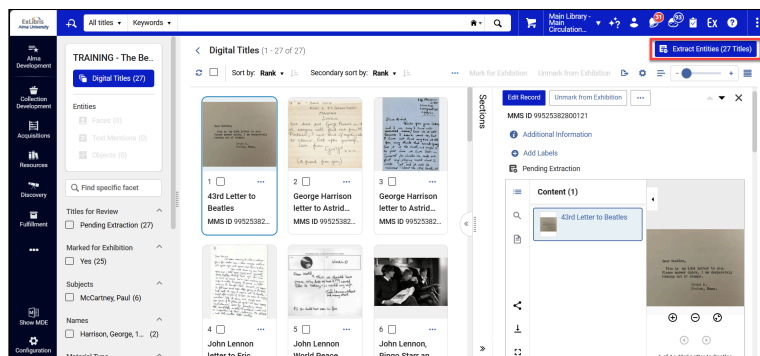
Once on the Alma Specto homepage for your collection, follow the steps to run entity extraction:

Note

For the best view of the titles and resources in your collection, switch to Gallery view in Specto by selecting the view options in the right-hand corner.

Images

1. Select the **Entity Extraction** button. This process reviews all relevant titles in the collection to identify and differentiate between images (pictures) and documents (containing text).



- The AI extracts **faces** and **objects** from the images. The process groups entities by similarity (e.g., all pictures of the same person are in one group).
- At each title level, face entities are visible in the image via a **bounding box** that highlights the extracted entity in the new Viewer.

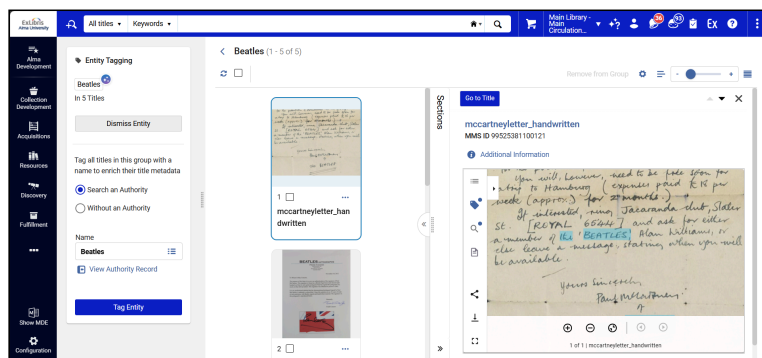
Entities are listed on the Viewer's left pane under **Entities**, allowing users to select and highlight the bounding boxes.

Documents

Note

The purpose of the ALTO file is to enable hit highlighting for document search, where the search term is highlighted directly in the document. This happens in the Specto environment in Alma.

- If none are available, the AI generates a full-text document and a corresponding ALTO for each file.
- The AI extracts names of **people**, **locations** and **organizations** from images.



- The process groups these entities by similarity (e.g., all mentions of the same person are in one group, all mentions of the same city are in another group).
- At the title level, each entity appears in the document under the Viewer **Entities** pane.
- Selecting an entity highlights the entity's mentions.
- With the generated full-text, search functionality is also available in the Viewer's **Search** pane and highlights keyword search results in the document.

Objects

- The AI extracts **types of objects**, **scenes**, and **actions** for the entire collection. Access them by selecting the **Objects** button on the left pane.
- Extracted objects for the collection is presented in a table with the following columns:

Object Name	Type	Category	Status	Number of Instances
1 Untagged Portrait	Portrait	Hobbies and Interests	Not Tagged	10
2 Untagged Photography	Photography	Hobbies and Interests	Not Tagged	10
3 Untagged Clothing	Clothing	Apparel and Accessories	Not Tagged	9
4 Untagged Accessories	Accessories	Apparel and Accessories	Not Tagged	7
5 Untagged Tie	Tie	Apparel and Accessories	Not Tagged	6
6 Untagged Formal wear	Formal wear	Apparel and Accessories	Not Tagged	6
7 Untagged Coat	Coat	Apparel and Accessories	Not Tagged	4
8 Untagged Sport	Sport	Sports	Not Tagged	4
9 Untagged Suit	Suit	Apparel and Accessories	Not Tagged	4
10 Untagged Happy	Happy	Expressions and Emotions	Not Tagged	3

- Entity Group** – this group encompasses objects of the same type and status. The column also shows the tag

- name given to objects once they are tagged.
2. **Type** – represents the type of object that was extracted.
 3. **Category** – a larger group to which the object belongs.
 4. **Status** – showing whether the objects are tagged or untagged.
 5. **Number of Instances** – the number of times a specific object appears in the image/collection. An image may contain more than one instance of the same object.
3. The objects are grouped by **Entity Group** and **Type**.
 4. You can change the objects tagged in a single group by selecting the **More Options** button (ellipsis) and selecting **Edit**.

Note

Processing time may vary depending on the size of the collection. The job runs in the background, and you are notified when it is complete. Do not take any further actions during job processing.

Note

The extraction and grouping runs against all titles that have not previously been processed. Titles added to or modified in the collection (e.g., added files) are available for extraction using the Extract Entities action, which indicates the exact population of titles processed.

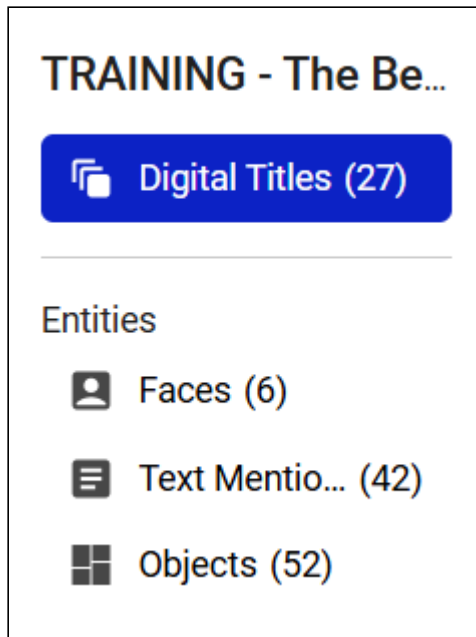
When choosing to extract, the pop-up:

- Indicates which titles are extracted:
 - X titles failed extraction
 - X titles pending extraction
- Each row is linkable and leads to the set of titles we recommend addressing before moving to the enrichment phase.

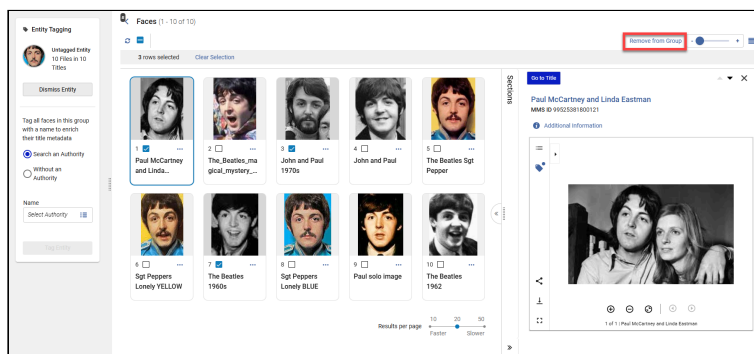


Step 3: Review Extracted Entities

1. Once the extraction and grouping process is complete, the entities appear in the left-side panel, under the collection name.
2. The **Digital Titles** section leads to the collection homepage and displays all the titles in the collection.
3. Select the type of entity you would like to review:
 1. **Faces** – to review the groups of faces detected in the images.
 2. **Text Mentions** – to review the people, organizations and locations mentioned in the documents.
 3. **Objects** – to review the table of objects detected in the images.



4. Within each section, you can select specific groups and review the groupings' accuracy to ensure that all group members are appropriately assigned before tagging.
 1. For grouped faces and text mentions, if one or more entities in a group are inaccurately associated, select the relevant items and use the **Remove from Group** bulk functionality to ensure the group is accurate for the next tagging step.



2. Object grouping is broader and is further elaborated in [tagging with or without linking to authority](#).
5. If you do not want to tag a detected entity or entities, you can dismiss the group in one of the following ways:
 1. On the entity section level, select the **More Actions** button (ellipses) under the image and select the option to **Dismiss Entity**.
 2. Within the entity itself, select the **Dismiss Entity** function from the left-side panel. Dismissed entities are not removed from the collection, but they are not available for tagging in this collection or marked on individual titles. The grouped entities are marked as **Dismissed** and can be restored later if necessary.
 3. For objects, you can dismiss a group by selecting the ellipses at the end of the row in the object table and selecting the **Dismiss Entity Group** option.

Note

Once you dismiss a group of objects, you can not find them again, unless you rerun the extraction process. Ensure you don't need the object group before dismissing it. Objects are not counted in the 'Pending tagging' counter and may be left untagged.

Step 4: Tag With or Without Linking to an Authority

Note

Use authority linking whenever possible to ensure consistency and interoperability across cataloging systems, and to support linked data in Primo.

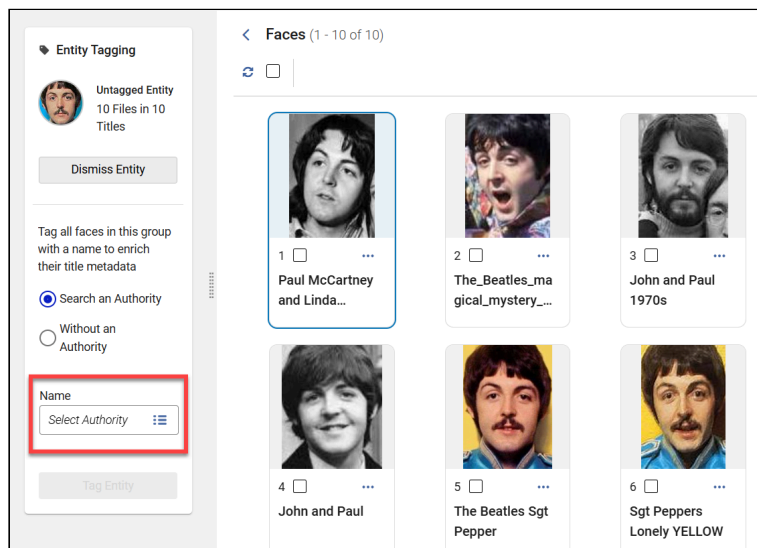
Faces

Once you are ready to tag the entities, select Edit Entity Group or select the item to begin. You can tag:

- With Authority Linking
- Without Authority Linking

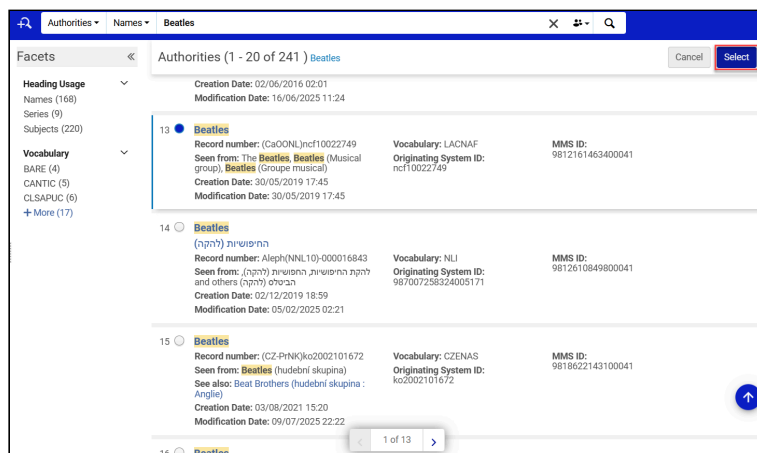
Tag with Authority Link

1. Select the **Search an Authority** option in the **Entity Tagging** box.



The screenshot shows the 'Entity Tagging' interface. On the left, there is a sidebar with the following options: 'Untagged Entity' (10 Files in 10 Titles), 'Dismiss Entity', and 'Tag all faces in this group with a name to enrich their title metadata'. Under 'Tag all faces in this group', the 'Search an Authority' radio button is selected, and the 'Without an Authority' radio button is unselected. Below these options is a text input field labeled 'Name' with the placeholder text 'Select Authority' and a dropdown arrow icon. At the bottom of the sidebar is a 'Tag Entity' button. The main area shows a grid of six face images with their corresponding labels: 1. Paul McCartney and Linda..., 2. The_Beatles_ma gical_mystery..., 3. John and Paul 1970s, 4. John and Paul, 5. The Beatles Sgt Pepper, 6. Sgt Peppers Lonely YELLOW.

2. Search the authority database and select the authority you wish to link to. Use filters and facets to narrow your search, and select individual records to view their metadata to ensure accuracy.
3. Select the relevant record.



The screenshot shows the Authority database search results for 'Beatles'. The search results are displayed in a table format with the following columns: Record number, Record number, Seen from, Creation Date, Modification Date, Vocabulary, Originating System ID, and MMS ID. The first record (13) is selected and highlighted. The record details are as follows:


Record number	Record number	Seen from	Creation Date	Modification Date	Vocabulary	Originating System ID	MMS ID
13	Beatles	Record number: (CaOONL)ncf10022749 Seen from: The Beatles, Beatles (Musical group), Beatles (Groupe musical)	30/05/2019 17:45	30/05/2019 17:45	LACNAF	ncf10022749	9812161463400041
14	Beatles	Record number: Aleph(NN10)-000016843 Seen from: התופשות (להקה), להקת התופשות החמישית (להקה) and others (להקה) הביטס (להקה)	02/12/2019 18:59	05/02/2025 02:21	NLI	987007258324005171	9812010849800041
15	Beatles	Record number: (CZ-PHNK)kc2002101672 Seen from: Beatles (hudební skupina) See also: Beat Brothers (hudební skupina : Anglie)	03/08/2021 15:20	09/07/2025 22:22	CZENAS	kc2002101672	9818622143100041

4. Select the **Tag Entity** button on the left panel. This tags ALL titles associated with the extracted entities in the group.

Tag without Authority Link

1. Select the **Without an Authority** option in the **Entity Tagging** box.

Entity Tagging



Untagged Entity
10 Files in 10
Titles

Dismiss Entity

Tag all faces in this group with a name to enrich their title metadata

Search an Authority

Without an Authority

Name*

Paul McCartney ×

Tag Entity

2. Enter the name manually in the **Name** field, meeting your institutional cataloging requirements.
3. Select the **Tag Entity** button. This tags ALL titles associated with the extracted entities in the group.

Note

Use authority linking whenever possible to ensure consistency and interoperability across cataloging systems, and to support linked data in Primo.

Text Mentions

The Text Mentions table includes the following columns:

- **Entity:** identified name of the grouped entity.
- **Tag Suggestion:**
 - **Stars:** AI-generated recommendation for an Authority record identified as appropriate for this entity.
 - **Lightbulb:** user-generated suggestion based on a previously tagged member of this group by a cataloger.
- **Entity Type:** distinguishes between people, locations and organizations.
- **Status:** untagged, tagged or dismissed.
- **Number of Titles:** number of titles in which this entity was mentioned. Note that there may be multiple mentions of the entity within each title.

Entity	Tag Suggestion	Entity Type	Status	Number of Titles
1 Beatles	*, Beatles	Organization	Not Tagged	5
2 Lennon, John	*, Lennon, John, 1940-1980	Person	Not Tagged	2
3 Williams, Alan	*, Williams, Alan	Person	Not Tagged	2
4 Hamburg (Germany)	*, Hamburg (Germany)	Location	Not Tagged	2
5 McCartney, Paul		Person	Not Tagged	2
6 Harrison, George	*, Harrison, George, 1943-	Person	Not Tagged	2
7 Smith, Mimi	*, Smith, Mimi	Person	Not Tagged	1
8 Liverpool (England)	*, Liverpool (England)	Location	Not Tagged	1
9 United States	*, United States	Location	Not Tagged	1
10 Bangor World		Organization	Not Tagged	1

Once you are ready to tag the entities, select **Edit** or select the item to begin. You can tag:

- With Authority Linking
- Without Authority Linking

Tag With Authority Link

If an AI recommendation or user suggestion is already available, the **Name** field is pre-populated with the proposed Authority. You may review the record by selecting the **View Authority Record** to determine if this is the Authority to which you want to link the grouped entity.

1. If no recommendations or suggestions are available, select the **Search an Authority** option in the **Entity Tagging** box.
2. Search the authority database and select the authority you wish to link to. Use filters and facets to narrow your search, and select individual records to view their metadata to ensure accuracy.
3. Select the relevant record.

TRAINING - The Bea...
Digital Titles (27)

Entities
Faces (6)
Text Mentions (42)
Objects (52)

Find specific facet

Entity Type
 Location (19)
 Organization (8)
 Person (15)

Tagging Status
 Untagged (42)

< Text Mentions (1 - 42 of 42)
Sort by: Number of Titles • JF

Entity	Tag Suggestion	Entity Type	Status	Number of Titles
1 Beatles	↕ Beatles	Organization	Not Tagged	5
2 Lennon, John	↕ Lennon, John, 1940-1980	Person	Not Tagged	2
3 Williams, Alan	↕ Williams, Alan	Person	Not Tagged	2
4 Hamburg (Germany)	↕ Hamburg (Germany)	Location	Not Tagged	2
5 McCartney, Paul	↕ McCartney, Paul	Person	Not Tagged	2
6 Harrison, George	↕ Harrison, George, 1943...	Person	Not Tagged	2
7 Smith, Mimi	↕ Smith, Mimi	Person	Not Tagged	1
8 Liverpool (England)	↕ Liverpool (England)	Location	Not Tagged	1
9 United States	↕ United States	Location	Not Tagged	1
10 Barjo World		Organization	Not Tagged	1

- Select the **Tag Entity** button on the left panel. This tags ALL titles associated with the extracted entities in the group.

Entity Tagging

Beatles

In 5 Titles

[Dismiss Entity](#)

Tag all titles in this group with a name to enrich their title metadata

Search an Authority

Without an Authority

Name

[View Authority Record](#)

[Tag Entity](#)

Tag Without Authority Link

- Select the **Without an Authority** option in the **Entity Tagging** box.
- Enter the name manually in the **Name** field below, meeting your institutional cataloging requirements.
- Select the **Tag Entity** button below. This tags ALL titles associated with the extracted entities in the group.

Entity Tagging

Beatles

In 5 Titles

Dismiss Entity

Tag all titles in this group with a name to enrich their title metadata

Search an Authority

Without an Authority

Name*

Beatles

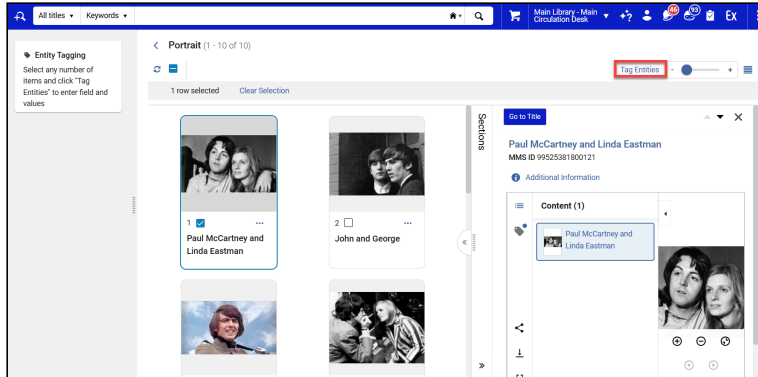
Tag Entity

Objects

- To tag extracted objects, select the object's name in the **Entity Group** column in the object table or select the ellipses at the end of the table row and select the **Edit Entity Group** option. A new page opens. On the left side, you see all the appearances of the selected object type in the collection, and on the right side, you have the viewer.

Object Name	Type	Category	Status	Number of Instances
1 Untagged Portrait	Portrait	Hobbies and Interests	Not Tagged	10
2 Untagged Photography	Photography	Hobbies and Interests	Not Tagged	10
3 Untagged Clothing	Clothing	Apparel and Accessories	Not Tagged	9
4 Untagged Accessories	Accessories	Apparel and Accessories	Not Tagged	7
5 Untagged Tie	Tie	Apparel and Accessories	Not Tagged	6
6 Untagged Formal wear	Formal wear	Apparel and Accessories	Not Tagged	6
7 Untagged Coat	Coat	Apparel and Accessories	Not Tagged	4
8 Untagged Sport	Sport	Sports	Not Tagged	4
9 Untagged Suit	Suit	Apparel and Accessories	Not Tagged	4
10 Untagged Happy	Happy	Expressions and Emotions	Not Tagged	3

- Choose the object or objects you wish to tag by selecting the checkbox under the title and selecting the **Tag Entities** button in the upper right side of the screen. Once tagged, the objects are removed from the Objects table and appear in the viewer within the Specto Title window.



Tag with Authority Link

1. Select the **Search an Authority** option in the **Entity Group Tagging** pop-up.

Entity Tagging

Portrait

1 Entity

Tag object(s) to enrich the title metadata. Select relevant MARC21 field to update in record. For Dublin Core (DC) enter "dc:subject".

Search an Authority
 Without an Authority

Field Value

Select Authority
☰

Field

1st Indicator

2st Indicator

Cancel
Tag

2. Search the authority database and select the authority you wish to link to. Use filters and facets to narrow your search, and select individual records to view their metadata to ensure accuracy.
3. Select the relevant record.

Authorities (1 - 5 of 5) Paul McCartney Linda Eastman

Sort by: Rank

1	McCartney, Linda מקרטני, לינדה	Record number: Aleph(NNL10)-004634558 Seen from: Eastman, Linda Creation Date: 30/01/2020 12:39 Modification Date: 01/02/2024 14:11	Vocabulary: NLI Originating System ID: 987007354502905171	MMS ID: 98138334095000041
2	McCartney, Linda	Record number: DCLn 87800870 Seen from: Eastman, Linda See also: Corporate body: Wings (Musical group) Creation Date: 26/02/2012 20:30 Modification Date: 31/12/2014 07:58	Vocabulary: LCNAMES Originating System ID: n 87800870	MMS ID: 98158777600000041
3	McCartney, Linda 1941-1998	Record number: (DE-588)119089955 Entity Type (GND): p Cataloging Level (GND): grnd1 TBK (GND): m, f, s Seen from: Eastman, Linda 1941-1998 , Eastman, Linda Louise 1941-1998 Frosherer Name: McCartney, L. 1941-1998 and others See also: (DE-588)012512266 McCartney, James 1977: Bezeichnung familiär , (DE-	Vocabulary: GND Originating System ID: 119089955	MMS ID: 989323981500041

4. Select the **Tag Entity Group** button on the left panel. This tags ALL selected entities in the group.

Tag without Authority Link

Entity Tagging

Portrait

1 Entity

Tag object(s) to enrich the title metadata. Select relevant MARC21 field to update in record. For Dublin Core (DC) enter "dc:subject".

Search an Authority
 Without an Authority

Field Value*

Portrait of Paul McCartney and Linda Eastman

Field*

999

1st Indicator

2st Indicator

1. Select the **Without an Authority** option in the **Entity Group Tagging** box.

2. Enter the metadata field number (a three-digit number) in the **Field** box, and enter the 1st and 2nd indicators according to your institutional cataloging requirements.
Ensure you don't enter a number of metadata fields that are already predetermined for another purpose.
3. Enter the subfield and name in the **Field Value** box below, meeting your institutional cataloging requirements. (You can enter only a name, but the system applies **\$\$a subfield** to the name.)
4. Select the **Tag Entity Group** button below. This tags the selected titles in the group. The tag appears in the metadata under the field you entered with the indicators you chose.