
批量创建数字内容

要管理数字资源，您需要具备下列角色：

- 数字馆藏操作员
- 数字馆藏操作员扩展（删除操作所需）
- 资源库馆藏操作员（添加新的数字表现时需要）

批量创建数字内容

Alma批量上传的工作流程包括以下步骤：

- 配置数字导入配置文件决定如何批量上传至Alma时如何处理数字资源。有关详细信息，见[管理导入配置文件](#)。
- 准备在导入文件时导入配置文件使用的书目记录的元数据文件。有关详细信息，见[准备XML元数据文件](#)和[准备CSV元数据文件](#)。
- 使用数字上传者将含有元数据的文档上传到Alma。有关更多信息，见[将文件上传到Alma](#)。

将文件上传到Alma

使用Alma的数字上传者将文件批量上传到Alma。题名（书目记录）放置在您上传时选择的资源库中。有关资源库的更多信息，见[管理资源库](#)。

在Alma中准备上传的每组文件称为上传批次。

准备上传批次分四个步骤进行：

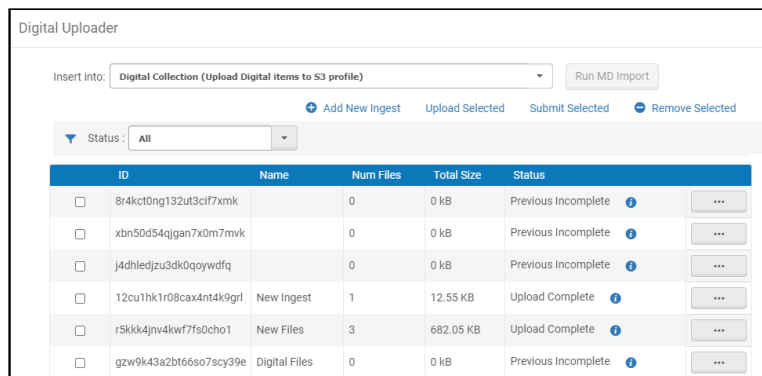
- 创建上传批次
- 将文件添加到上传批次（包括MD文件）
- 正在上传文件
- 提交上传批次

Note

上传批次文件夹信息存储在浏览器的本地存储中。因此，如果您从其他浏览器（或从相同的浏览器访问但是以其他用户身份登录），或者如果清除浏览器的本地存储，则不会看到以前创建的上传批次文件夹。这不影响已经提交的数据。

要上传和提交文件到Alma：

1. 打开数字上传者页面（[资源](#) > [高级工具](#) > [数字上传者](#)）。



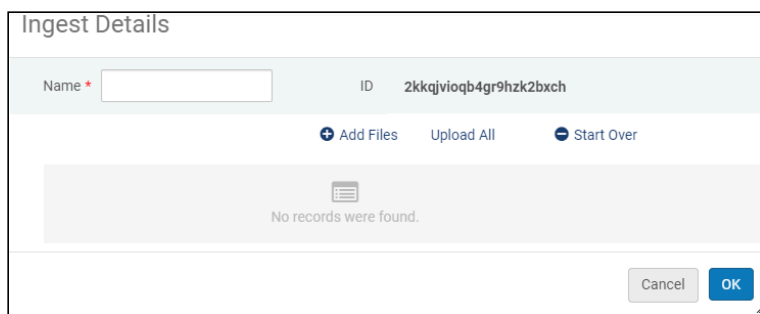
数字上传器

2. 从插入下拉列表中，选择要将文件和数字导入配置文件放在哪个资源库中，该配置文件定义文件导入到资源库中的方式。
3. 通过以下方式添加要提交到Alma的文件：
 - 拖放 - 通过将文件夹拖放到批次列表中添加上传批次，并通过将它们拖放到创建列表中的上传批次中以将文件添加到现有上传批次中。

Note

目前仅Chrome支持文件夹拖放。使用支持的浏览器可以拖放文件。

- 使用添加新上传批次框：
 1. 从数字上传器页面，点击添加新上传批次。出现以下内容：



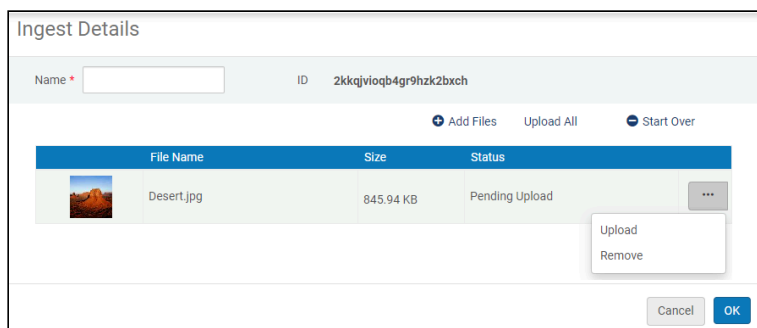
添加新上传批次

2. 输入上传批次名称。
3. 点击添加文件并选择要上传的文件。文件的名称、文件大小以及如果可能，以下文件格式的自动生成的缩略图将显示在上传批次对话框中。
 - jpg
 - png
 - mp4
 - wav
 - m4v
 - pdf

- doc
- ppt
- docx
- pptx
- jpeg2000

Note

- 每个文件最大可为1GB。
- 您可以在一个上传批次中最多包含1000个文件。
- 对于不属于以上格式的批次，可以直接上传到S3存储。有关更多信息，见[开发者网络](#)。



开始上传批次

Note

- 例如，您可以在上传批次中包含一个缩略图文件，该文件将显示在仓储检索结果和[数字查看器](#)中。有关详细信息，见[将缩略图文件添加到上传批次](#)。
- 您可以在ingest中包含与视频文件名称一致的.vtt格式字幕文件使其在播放视频时显示。字幕文件应命名为**abc.mp4.vtt.captions**，其中上传的视频文件为**abc.mp4**。有关更多信息，见[新数字查看器](#)。

-
4. 对上传批次（见[准备XML元数据文件](#)或[准备CSV元数据文件](#)）添加一个或多个元数据文件。

Note

如果不将元数据文件添加到上传批次中，则会在上传批次列表中显示警告图标



。

-
5. 要在上传批次中上传一个文件，点击上传。要一次性在上传批次中上传所有文件，选择上传全部。要移除文件并重新开始，点击重新来。
 6. 点击确定 返回数字上载器页面。见[数字上载器](#)。

以下信息在上传批次列表中可用：

- 上传批次的ID
- 上传批次的名称
- 上传批次文件的数量
- 上传批次的总大小
- 上传批次的状态以下是可能的状态：
 - 新 - 上传批次不包含任何文件。
 - 待处理上传 - 上传批次尚未上传。
 - 已完成上传 - 上传批次尚未提交。
 - 提交 - 已经提交了上传批次。

Note

- 要从上传批次中移除文件，选择操作按钮，然后选择移除。
- 要移除上传批次，选择上传批次并选择移除所选。

-
4. 要上传上传批次，选择上传批次，然后点击上传所选。
 5. 要提交上传批次进行处理，选择上传批次然后点击提交所选。

Note

- 提交步骤是必需的，以便您可以继续安全地将文件添加到您的上传批次，在您选择提交之前，该文件将不会被处理。
- 导入配置文件中的元数据文件名字段必须含有提交文件的正确路径和文件名。

-
6. 通常，元数据导入作业将按计划运行。要手动运行元数据导入作业，点击运行导入作业。

元数据导入作业运行并处理上传批次。

Note

超过30天的文件（锁定文件除外）由每周维护作业删除。

准备XML元数据文件

要使用数字上载器执行批量文件上传，您可以使用用于书目记录的元数据来准备一个MARC或DC格式XML文件。每个元数据文件可以包含多个书目记录信息。Alma使用此元数据文件中的信息创建文件的书目记录和表现。

文件中的记录元素应该嵌套在单个资源库元素中。

元数据文件中所需的信息类型取决于您如何配置导入配置文件。例如，如果您将导入配置文件配置为在MARC字段856子字段u中查找文件名，则元数据文件必须在此字段中包含此信息。

如果书目记录包含资源库信息，记录将被分配给记录中指示的资源库。否则，记录将被分配给数字导入配置文件中选定为默认资源库分配的资源库。

Note

当上传格式不正确的XML文件或文件缺少元素时，元数据上传作业将会完成并带有错误。该作业将一直失败，直到用户删除或替换有故障的元数据文件。

MARC XML 资源库分配

对于MARC记录，如果激活含有**addBibToCollectionNormalizationTask**的规范化进程，根据以下规则优先级将记录分配给资源库：

- 如果MARC 787\$w包含有效的Alma资源库ID，则该记录将分配给此资源库。
 - 如果MARC 787\$o包含外部系统和外部ID字段的值，如在现有的Alma资源库中配置的，则记录将分配给此资源库。该字段必须包含以下格式的系统ID：({system}){ID}，例如，(Rosetta)123454321（类似于035字段结构）。
 - 如果MARC 787 \$t包含顶层Alma资源库的名称，则该记录将分配给此资源库。
 - 如果以上均不适用，则记录将分配给定义为导入配置文件默认值的Alma资源库。
-

Note

当上传含有单个记录的元数据文件时，不需要856字段。创建的记录包括在上传批次文件夹中的所有文件。

以下是XML元数据文件的示例：

```
<?xml version="1.0" encoding="UTF-8" ?>
<collection>
<record>
<leader>01139cam a2200277 a 4500</leader>
<controlfield tag="001">ocm27832725113 </controlfield>
<controlfield tag="003">OCoLC</controlfield>
<controlfield tag="005">20131108731707.4</controlfield>
<controlfield tag="008">970910r19451931pk b 000 0 eng </controlfield>
<datafield tag="035" ind1=" " ind2=" ">
<subfield code="a">(OCoLC)27832725113</subfield>
</datafield>
<datafield tag="100" ind1="1" ind2=" ">
<subfield code="a">Representation of rec 1</subfield>
<subfield code="b">jpg files</subfield>
</datafield>
<datafield tag="245" ind1="1" ind2="2">
<subfield code="a">Record 1 - Civil war</subfield>
</datafield>
<datafield tag="856" ind1=" " ind2=" ">
<subfield code="a">Soldier on a Horse</subfield>
<subfield code="u">civil war.jpg</subfield>
</datafield>
<datafield tag="856" ind1=" " ind2=" ">
<subfield code="a">North and South</subfield>
<subfield code="u">civil2.jpg</subfield>
</datafield>
```

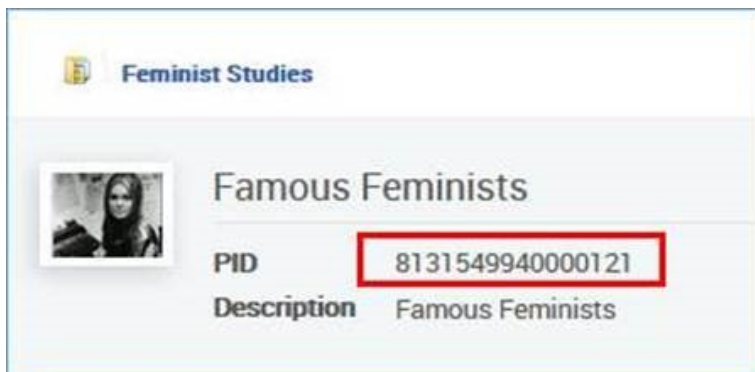
```
<datafield tag="650" ind1=" " ind2="0">
<subfield code="a">Digital items loaded to Amazon</subfield>
</datafield>
</record>
</collection>
```

以下为使用MARC 787 字段将记录分配给资源库的示例：

- 下列记录：

```
<datafield tag="245" ind1="1" ind2="4">
<subfield code="a">The politics of our lives :</subfield>
<subfield code="b">the Raising Her Voice in Pakistan experience /</subfield>
<subfield code="c">Jacky Repila.</subfield>
</datafield>
<datafield tag="787" ind1=" " ind2=" ">
<subfield code="w">8131549940000121</subfield>
</datafield>
```

将添加到此资源库：



示例1

- 下列记录：

```
<datafield tag="245" ind1="0" ind2="0">
<subfield code="a">New South Asian feminisms :</subfield>
<subfield code="b">paradoxes and possibilities /</subfield>
<subfield code="c">edited by Srila Roy.</subfield>
</datafield>
<datafield tag="787" ind1=" " ind2=" ">
<subfield code="w">8131549960000121</subfield>
</datafield>
```

将添加到此资源库：



示例2

DC XML 资源库分配

对于都柏林核心元数据记录，如果激活含有**addBibToCollectionNormalizationTask**的规范化进程，Alma检查dc:relation和dcterms:isPartOf字段。Alma按照以下顺序检查每个字段并分配导入记录到资源库。

1. 通过内部资源库ID：

```
<dc:relation>alma:{INST_CODE}/bibs/collections/{COLLECTION_ID}</dc:relation>
```

2. 通过外部资源库ID：

```
<dc:relation>any text</dc:relation>
```

3. 通过资源库名称（仅限顶层）：

```
<dc:relation>any text</dc:relation>
```

4. 通过默认资源库（仅当1-3与现有资源库不匹配时）。

MODS XML 资源库分配

创建MODS的添加书目到资源库规范化规则。这与对DC的处理类似，允许根据资源库ID、资源库外部ID和资源库名称分配：

- `<relatedItem@type="host"><identifier>alma:{INST_CODE}/bibs/collections/{COLLECTION_ID}</identifier></relatedItem>`
- `<relatedItem@type="host"><identifier>{external_id}</identifier></relatedItem>`
- `<relatedItem@type="host"><identifier>{collection_name - top level only}</identifier></relatedItem>`

准备CSV元数据文件

要使用数字上载器来执行文件批量上传，您可以使用用于书目记录的元数据来准备一个CSV文件。CSV文件是基于馆藏的，这意味着每一行都代表一个表现。每一行都可以有完整的书目信息（生成一个新书目记录），或者可以引用现有的书目记录，与匹配规则匹配，并将数字馆藏添加到引用的书目记录中。文件级别内容包含在同一行中，包含完整的文件路径（相对于上传批次文件夹）和（可选）标签。每个表现支持多个文件。每个书目记录的多个表现由包含完整书目记录的两行表现（具有防止创建多个记录的规则）。在此情况下，向相同的表现分配group_id，从而Alma知道这些表现是同一书目记录的一部分。

Note

CSV文件必须为UTF-8编码（非UTF-8-BOM）。

导入元数据时，您可以提供带有外部系统和ID的CSV记录，以将其分配给资源库外部系统列中的资源库（与MARC记录格式相同）。该字段是可重复的。

CSV可以包含以下类型的字段：

- 资源库 - 用于资源库分配。包含分配书目记录的资源库名称或ID（资源库名称仅为顶层资源库）。该字段可选。（如果资源库字段不存在，则书目记录将分配给元数据导入配置文件中定义的默认资源库）。可重复。

如果导入配置文件的格式设置为MARC，CSV主题名称如下映射到MARC字段：

- collection_name - 787 t
- collection_id - 787 w
- collection_external - 787 o

如果导入配置文件中的格式设置为都柏林核心，CSV头名称映射到dc:relation字段。

Note

您必须在导入配置文件中设置规范化规则添加书目到资源库使导入作业根据这些字段分配资源库。否则使用默认资源库。

-
- 书目记录级别字段 - 有两种子类型：
 - mms_id和originating_system_id - 用于将记录与现有记录进行匹配。如果要在mms_id上匹配并添加originating_system_id（由规则处理），则两个字段可以一起存在。mms_id不可重复。
 - 其他书目记录字段 - 根据开发人员网络的表格映射到目标元数据格式（<https://developers.exlibrisgroup.com/alma/integrations/digital/almadigital/ingest>）。有关重复性细节，见表格。（不可重复的字段标记为NR。）
 - 表现级别字段 - 映射到表现属性。保留字段是不可重复的（rep_note除外）。
 - 文件等级字段。可重复。

应根据Excel解析器创建和处理CSV内容。例如，带逗号的字段必须用引号括起来：

War and Peace,"Tolstoy, Leo",1862,...

csv支持MARC XML支持的所有字符。

有关CSV文件的示例，见[CSV示例](#)。

CSV到都柏林核心映射

当您选择CSV作为纸本源格式和都柏林核心作为目标格式时，Alma将CSV记录中的信息转换为都柏林核心格式。大多数匹配是直观的，例如：责任者映射到dc:contributor，但是请注意以下事项：

- MMS_ID只能用于匹配现有的Alma MMS ID。语法是：alma:{INST_CODE}/bibs/{MMS_ID}。此字段不作为导入记录的一部分存储。
- ISBN和ISSN映射到具有前缀urn的<dc:identifier xsi:type="dcterms:URI">。
- Originating_system_id映射到dc:identifier。
- 对于资源库，collection_id按以下格式映射到dc:relation:<Inst-code>/bibs/collections/<collection id>

- 所有字段都是可重复的，没有一个是必须的。
- 建议使用DCMI类型词汇表（不包括不应使用的资源库）。
- 通过将表现分配给相同的group_id来支持为单个书目记录映射多个表现。
- 支持编码标准和语言。ISO 639-1格式要求使用两个字母的语言代码，ISO 639-2/3要求使用三个字母的语言代码。

句法如下：

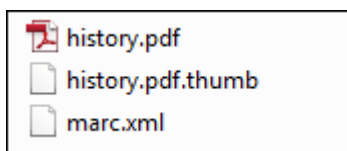
- 仅编码 – property.schema (例如：dc:subject.dcterms:LCSH)
- 仅语言 – property lang={2 or 3 letter code} (例如， dc:subject lang=en)
- 编码和语言 – property.schema lang={2 or 3 letter code} (例如：dc:subject.dcterms:LCSH lang=en)

准备Excel元数据文件

有关准备Excel元数据文件的信息，见[使用CSV或Excel文件导入记录](#)。

将缩略图文件添加到上传批次

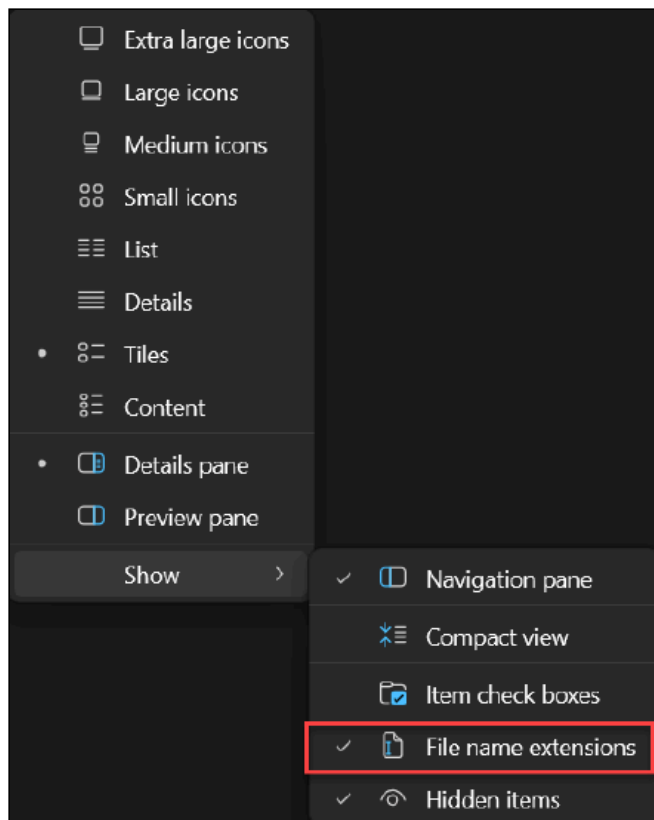
您可以将缩略图文件添加到与缩略图扩展名的数字文件相同的文字中（仅限jpg，png，gif格式）。例如：



缩略图文件

这允许Alma将.thumb图像存储为相应文件的缩略图。缩略图将显示在仓储检索结果、[数字查看器](#)和Primo中。

要成功使用.thumb扩展名重命名文件，您需要在Windows文件资源管理器中启用扩展名。否则，文件将保留其格式并在名称中添加.thumb。要在文件资源管理器中启用文件扩展名，请选择视图下拉菜单，然后选择显示，接下来选择选项文件扩展名。



将全文文件添加到上传批次

您可以添加plain或alto格式的全文文件到上传批次，使用与数字文件相同的名称并带有.text.plain或.text.alto扩展名。这允许您对Book Reader查看器中的图像执行文本检索。

在删除过程中，可以将OCR文件与其他数字文件关联。

如果您有名为**abc.tif**的.tif文件以及与.tif文件相关的纯文本类型OCR文件，则准备类似于如下的上传文件夹：

<文件夹名称>

abc.tif

abc.tif.text.plain

使用CSV文件进行上传时，请按照上述说明包含带该文件的文件夹。

您无需在CSV文件中列出OCR文件名。