

Rockefeller Archive Center BagIt Specification

This page describes the Rockefeller Archive Center's BagIt specification, developed with the goal of facilitating consistently structured bags – or packages – of digital records which can be programatically validated and verified.

Donor organizations are responsible for creating bags which comply to this specification and transferring them via agreed-upon protocols and schedules. To assist donors in meeting these requirements, a bag profile for organizational records (`organizational-bag-profile.json`) as well as example Python scripts (`example-scripts/`) to create and validate bags are available in this repository. Other libraries for this specification are available in languages including Java (<https://github.com/LibraryOfCongress/bagit-java>), Python (<https://github.com/LibraryOfCongress/bagit-python>), PHP (<https://github.com/scholarslab/BatItPHP>) and Ruby (<https://github.com/topr/bagit>).

Specification

1. Rockefeller Archive Center bags conform to the BagIt packaging specification (<https://tools.ietf.org/html/draft-kunze-bagit-14>)
2. Rockefeller Archive Center bags may be
 1. serialized (single ZIP file)
 2. un-serialized
3. All bags must valid according to the Rockefeller Archive Center BagIt profile (https://github.com/RockefellerArchiveCenter/project_electron/blob/master/transfer/bagit_spec.json)

RAC BagIt Structure

This section includes a simple example of a Rockefeller Archive Center BagIt Specification-compliant bag. Although the Rockefeller Archive Center accepts both serialized and un-serialized bags, this example specifies an unserialized bag.

```
<RAC-BAG- ID> /
| bagit.txt
| manifest-md5.txt
| bag-info.txt
\--- data/
    | [payload files]
    | metadata.json
```

Description

RAC-BAG-ID

The name of the root directory of the bag.

bagit.txt

As required in the BagIt spec. Includes BagIt version and tag file character encoding.

```
BagIt-Version: 0.97
Tag-File-Character-Encoding: UTF-8
```

manifest-md5.txt

This is a required element from the BagIt spec that contains a checksum for every item included in the bag's

payload. For this example we have chosen md5, but sha256 is also acceptable.

bag-info.txt

The Rockefeller Archive Center requires some of the below fields in our specification, and some are reserved BagIt fields as indicated in the **BagIt Specification**. For more details on the content, requirements, and usage of each field, please see the **Bag-Info Field Specification** section at the bottom of this document. The RAC will index metadata from bag-info.txt as structured data.

```
Source-Organization
External-Identifier
Internal-Sender-Description
Title
Date-Start
Date-End
Record-Creators
Record-Type
Language
Restrictions
Bagging-Date
Bag-Count
Bag-Group-Identifier
Payload-Oxum
BagIt-Profile-Identifier
```

data

Required directory for payload items. This directory may be serialized as a single file (for example a ZIP file), or unserialized as a number of files and subdirectories.

metadata.json

Valid JSON or JSON-LD file that includes metadata elements included in bag-info.txt as well as any additional elements donors wish to provide to the RAC.

Bag-Info Field Specifications

Source-Organization

- **Definition:** The organization responsible for sending the content.
- **Purpose:** Provides information to the Rockefeller Archive Center about the organization sending the records.
- **Data type:** Locally controlled. See the Rockefeller Archive Center BagIt profile (https://github.com/RockefellerArchiveCenter/project_electron/blob/master/transfer/bagit_spec.json) for acceptable values.
- **Obligation:** Required
- **Repeatability:** No
- **Examples:**
 - "Ford Foundation"

External-Identifier

- **Definition:** A unique identifier applied to each group of records composed of characters, numbers or letters, or a combination thereof, that uniquely identify the record within a given domain.
- **Purpose:** Ensures that records can be uniquely identified within the Rockefeller Archive Center's systems during archival processes.
- **Data type:** String

- **Obligation:** Required
- **Repeatability:** No
- **Examples:**
 - "OyGpXmSFVkcPds7i4gRv"
 - "Grant2561"

Internal-Sender-Description

- **Definition:** A prose description of the nature and contents of the group of records.
- **Purpose:** Aids user discovery by providing description about the content and context of records.
- **Data type:** String
- **Obligation:** Required
- **Repeatability:** No
- **Examples:**
 - "Annual reports discussing the accomplishments and major strategic initiatives of the Ford Foundation."

Title

- **Definition:** The title of a group of records. Do not include dates in the title element. This should not be a description of the type of materials, but instead a declarative title for all of the records in the bag.
- **Purpose:** Enables information search and discovery, facilitates user choice and provides additional context for archival records.
- **Data type:** String
- **Obligation:** Required
- **Repeatability:** No
- **Examples:**
 - "Annual Reports"

Date-Start

- **Definition:** The earliest date on which records in the bag were created.
- **Purpose:** Enables information search and discovery, facilitates user choice and provides additional context for archival records.
- **Data type:** Values must meet *ISO 8601: Standard for Representation of Dates and Times*.
- **Obligation:** Required
- **Repeatability:** No
- **Examples:**
 - "1995-01-01"
 - "2002"

Date-End

- **Definition:** The latest date on which records in the transfer were created. Use only if this value differs from Date-Start.
- **Purpose:** Enables information search and discovery, facilitates user choice and provides additional context for archival records.
- **Data type:** Values must meet *ISO 8601: Standard for Representation of Dates and Times*.
- **Obligation:** Required for date ranges
- **Repeatability:** No

- **Examples:**
 - "1997-12-31"

Record-Creators

- **Definition:** Identifies the individuals, organizations or departments that created the group of records.
- **Purpose:** Provides evidence and context for records by identifying the creator of the record and supports user search, allowing information to be filtered.
- **Data type:** String
- **Obligation:** Required
- **Repeatability:** Yes
- **Examples:**
 - "Ford Foundation Communications Office"
 - "Walker, Darren"

Record-Type

- **Definition:** The broad category into which the records fall.
- **Purpose:** Allows records to be staged appropriately for archival processes and facilitates the appropriate application of embargo periods, which vary by record type.
- **Data type:** Locally controlled. See the Rockefeller Archive Center BagIt profile (https://github.com/RockefellerArchiveCenter/project_electron/blob/master/transfer/bagit_spec.json) for acceptable values.
- **Obligation:** Required
- **Repeatability:** Yes
- **Examples:**
 - "annual reports"
 - "grant records"

Languages

- **Definition:** The natural language(s) in which the materials are written.
- **Purpose:** Assists user evaluation of relevance and facilitates machine translation.
- **Data type:** Values must meet *ISO 639-2: Codes for the Representation of Names of Languages*. Use URIs linking directly to the language code. If materials have no language, please use "nil"
- **Obligation:** Required
- **Repeatability:** Yes
- **Examples:**
 - "http://id.loc.gov/vocabulary/iso639-2/eng (<http://id.loc.gov/vocabulary/iso639-2/eng>)"
 - "http://id.loc.gov/vocabulary/iso639-2/spa (<http://id.loc.gov/vocabulary/iso639-2/spa>)"

Restrictions

- **Definition:** Any restrictions related specifically to the group of records in the bag. Do not record general restrictions related to all records (or record types) from a given person or organization.
- **Purpose:** Provides information to users about the availability of records, and assists archivists in responsibly providing access to records.
- **Data type:** String
- **Obligation:** Optional
- **Repeatability:** No
- **Examples:**

- "The annual report for 1996 might have IP restrictions due to photographs for which the Ford Foundation does not own copyright."

Bagging-Date

- **Definition:** Date that the content was prepared for delivery.
- **Purpose:** Provides the date the bag was created to ensure authenticity.
- **Data type:** Values must meet *ISO 8601: Standard for Representation of Dates and Times*.
- **Obligation:** Required
- **Repeatability:** No
- **Examples:**
 - "2016-04-24"

Bag-Count

- **Definition:** Two numbers separated by 'of', in particular, 'N of T', where T is the total number of bags in a group of bags and N is the ordinal number within the group; if T is not known, specify it as '?' (question mark).
- **Purpose:** Allows donors to split large bags into multiple parts that can be reassembled later.
- **Data type:** String
- **Obligation:** Optional
- **Repeatability:** No
- **Examples:**
 - "1 of 2"
 - "4 of 4"
 - "3 of ?"

Bag-Group-Identifier

- **Definition:** A unique identifier for the entire set of bags to which this bag belongs.
- **Purpose:** Provide an identifier to logically group multi-bag transfers.
- **Data type:** String
- **Obligation:** Optional
- **Repeatability:** No
- **Examples:**
 - "xOmy5"
 - "AnnualReports"
 - "Group1"

Payload-Oxum

- **Definition:** The "octetstream sum" of the payload, namely, a two-part number of the form `OctetCount.StreamCount`, where `OctetCount` is the total number of octets (8-bit bytes) across all payload file content and `StreamCount` is the total number of payload files. Payload-Oxum should be included in `bag-info.txt` if at all possible.
- **Purpose:** Payload-Oxum is intended for machine consumption.
- **Data type:** `OctetCount.StreamCount`
- **Obligation:** Required
- **Repeatability:** No
- **Examples:**

- "279164409832.1198"

BagIt-Profile-Identifier

- **Definition:** An HTTP URI that
- **Data type:** Locally controlled, temporary URL:
"https://raw.githubusercontent.com/RockefellerArchiveCenter/project_electron/master/transfer/organizational-bag-profile.json (https://raw.githubusercontent.com/RockefellerArchiveCenter/project_electron/master/transfer/organizational-bag-profile.json) "
- **Obligation:** Required
- **Repeatability:** No
- **Examples:**
 - "https://standards.rockarch.org/bagit/organizational-bag-profile.json
(https://standards.rockarch.org/bagit/organizational-bag-profile.json) "
 - "http://rockarch.org/bagitprofiles/bag-profile.json (http://rockarch.org/bagitprofiles/bag-profile.json) "