

## Harvard E-Journal Archive

# Submission Information Package (SIP) Specification Version 1.0 DRAFT — December 19, 2001

Harvard University Library  
Office for Information Systems  
1280 Massachusetts Avenue, Suite 404  
Cambridge, MA 02138 USA

<p><i>Please direct any comments on this draft to:</i> Stephen Abrams &lt;stephen_abrams@harvard.edu&gt; Marilyn Geller &lt;marilyn.geller@mindspring.com&gt;</p>
---

## Abstract

The purpose of the Harvard University E-Journal Archive is to preserve the significant intellectual content of journals independent of the form in which that content was originally delivered in order to assure that this content will be available to the scholarly community for the indefinite future. Functionally, the archive is designed to render text and still images and other formats as practical with no significant loss in intellectual content. The archive reserves the right to freely manipulate the internal format of the manifestation over time as long as the plain meaning of the intellectual content is preserved.

The framework for discussing the architecture and operation of the archive is provided by the Open Archival Information System (OAIS) Reference Model. Under the OAIS model, material from a content provider is transmitted to the archive in a form called a Submission Information Package (SIP). The format of the SIP acceptable to the Harvard archive is described normatively by this specification.

## Table of Contents

1	Introduction .....	1
1.1	<i>Change History</i> .....	1
1.2	<i>Typographic Conventions</i> .....	1
1.3	<i>Terminology</i> .....	1
1.4	<i>Audience</i> .....	1
2	Desiderata and Policy .....	2
2.1	<i>Conceptual Framework</i> .....	2
2.2	<i>Specification Maintenance</i> .....	3
2.3	<i>Version Identification</i> .....	3
3	SIP Structure .....	4
3.1	<i>SIP Components</i> .....	4
3.2	<i>Normative Data Formats</i> .....	4
3.3	<i>Character Set Issues</i> .....	5
3.4	<i>Directory Structure</i> .....	6
3.4.1	Title-level directory .....	6
3.4.2	Issue-level directory .....	7
3.4.3	Item-level directories .....	7
3.5	<i>Issue-level File Components</i> .....	8
3.5.1	Issue-level metadata file .....	8
3.5.2	Issue-level content files .....	14
3.6	<i>Item-level File Components</i> .....	15
3.6.1	Item-level metadata file .....	15
3.6.2	Item-level content files .....	22
3.7	<i>SIP Aggregation</i> .....	22
4	SIP Submission Session .....	23
4.1	<i>SIP Ingest</i> .....	23
A	METS Framework .....	24
A.1	<i>SIP Issue-Level Metadata File (issue-md.xml)</i> .....	26
A.2	<i>SIP Item-Level Metadata File (item-md.xml)</i> .....	28
B	Normative Data Formats .....	31
B.1	<i>Metadata, Issue and Item-Level Text, and Item-Level Links</i> .....	31
B.1.1	Structural metadata schema .....	31
B.1.2	Descriptive and administrative metadata schema .....	31
B.1.3	Issue-level text schema .....	32
B.1.3	Item-level text schema .....	32
B.1.4	Item-level mathematics DTD .....	32
B.1.4	Item reference links schema .....	32
B.2	<i>Raster Still Image</i> .....	33
B.3	<i>Vector Still Image</i> .....	33
B.4	<i>Page Description</i> .....	33
B.5	<i>Audio</i> .....	33
B.6	<i>Video</i> .....	34
C	Non-normative Data Formats .....	35
C.1	<i>Raster Still Image</i> .....	35
C.1.1	GIF .....	35
C.1.2	JPEG / JFIF .....	35
C.2	<i>Page Description</i> .....	36
C.2.1	Encapsulated PostScript .....	36
C.2.2	PostScript .....	36
	References .....	37

# 1 Introduction

The Harvard University Library is developing an archive for the preservation of scholarly electronic journals. The framework for discussing the architecture and operation of such an archive is provided by the Open Archival Information System (OAIS) Reference Model [OAIS]. Under the OAIS model, material from a content provider is transmitted to the archive in a form called a Submission Information Package (SIP). The archive Ingest function accepts the SIP and potentially transforms its contents into an internal form called an Archival Information Package (AIP) for long-term preservation.

## 1.1 Change History

<b>Version</b>	<b>Date</b>	<b>Status</b>
1.0	December 19, 2001	First public draft.

Changes from the previously published version are indicated by a vertical bar in the right-hand margin.

## 1.2 Typographic Conventions

Directories, file names, and the literal contents of files are displayed in a `non-proportional` typeface. Abstract references to file contents that should be replaced with literal values are displayed in an *italic non-proportional typeface*.

## 1.3 Terminology

The use of the term “must” in this document indicates an absolute requirement necessary for compliance with this specification. The use of the term “should” indicates a strong preference that can be ignored only within the conditions specified in subsequent text.

## 1.4 Audience

This specification is prepared for E-journal content providers who wish to deposit material in the archive. Compliance to this specification is required for acceptance of submitted material.

## 2 Desiderata and Policy

The purpose of the Harvard E-Journal Archive is to preserve and deliver the intellectual content of scholarly electronic journals. Intellectual content is the human interpretable meaning of the physical instantiation of the journal as originally electronically published. Note that the archive is responsible for preserving intellectual content, not any particular representation of that content. The community to whom content is delivered is assumed to be conversant with the scholarly domain of that content.

The Submission Information Package (SIP) is the physical container in which content providers submit E-journal content to the archive.

### 2.1 Conceptual Framework

The design of the Harvard E-journal archive, and its SIP, is based upon the following conceptual underpinnings:

**Distinction between work and manifestation.** The archive's purpose is to preserve the intellectual content of journals independent of the tangible form in which that content was published; in other words, archiving takes place at a semantic, not a syntactic level. The archive reserves the right to freely manipulate the internal format of the manifestation over time as long as the plain meaning of the intellectual content is preserved.

Note that is possible for typographical manifestation to impart significant semantic value, as in the case of poetry and other forms of creative expression. In such cases where the manifestation forms an intrinsic part of the intellectual content it will be captured and preserved by the archive.

**Separation of content and metadata; data modeling at issue and item level.** Within the SIP and the archive these distinctions are enforced both at a conceptual level and at the level of physical objects. Issue is defined loosely as a publisher-specified aggregation of individual items. Item is defined as an indivisible piece of citable content such as an article, editorial, review, letter, erratum, etc. All individual archival objects can be categorized according to the following two-dimensional matrix:

	<b>Content</b>	<b>Metadata</b>
<b>Issue</b>	<i>Issue content</i>	<i>Issue metadata</i>
<b>Item</b>	<i>Item content</i>	<i>Item metadata</i>

**Distinction between primary and supplementary content.** Primary content includes the text (including non-verbal notation such as mathematical formulæ) and images directly supporting an intellectual argument, and without which that argument would not be sustainable or understandable. Supplementary content is additional material not necessary to understanding the intellectual argument, although perhaps useful for the purposes of verifying that argument, including research source material or data files.

**Normative vs. non-normative data formats.** The archive defines a limited number of data formats as being normative. The archive will undertake to guaranty the continued usefulness of data stored internally in normative formats over archival time-spans. Data submitted in non-normative formats will be transformed upon ingest into analogous normative formats whenever possible without substantive loss of semantic value.

**Common item-level schema.** Within the SIP and the archive all item-level textual content is normalized to a common schema, designed to preserve the semantic integrity of item intellectual content. Purely syntactic or representational aspects of items' original electronic publication may not be captured or preserved under this schema.

**Self-documenting.** As far as practicable, SIP structuring and naming conventions are designed to allow predictable and unambiguous identification of file type and function independent of the contents of those files.

## **2.2 Specification Maintenance**

This document represents the authoritative specification of the format of the SIP used by publishers for submission of E-journal content to the archive. In order to minimize any disparate impact that changes to this specification might have upon publishers' internal systems and work-flow, the following update process will be followed:

- Suggestions for enhancements to this specification may originate either internally from the archive development team or externally from the publishing community.
- Updates to this specification will occur, if necessary, on no more frequent a basis than once a year.
- All changes proposed for a given update cycle will be aggregated together and formally announced, beginning a two month public review and comment period. During this time the proposed changes may be refined in light of ongoing public discussions.
- At the end of this review period the archive development team will perform a final evaluation of the proposed changes in light of received comments. Changes will proceed towards promulgation only in the cases of adequate public consensus or significant technical necessity, as determined by the archive development team.
- All accepted changes will be formally promulgated and a new edition of this document will be published with all necessary emendations six months prior to the changes becoming operationally effective.
- All changes to this specification will be indicated by the appropriate increment of the version identifier.

## **2.3 Version Identification**

All versions of this specification are assigned a unique two-part identifier of the form *major.minor*. Incompatible changes between subsequent versions will always be reflected by an increment in the major number. Minor number increments indicate an upwardly compatible change from previous versions. (Aspects of the specification may be deprecated during minor number increments, but these features will always remain valid until a major number update.)

For purpose of comparison the major and minor version numbers should be separately interpreted numerically. For example, version 2.3 precedes version 2.13, which in turn precedes version 12.3.

All SIP instances contain an internal statement of the version of this specification to which they conform.

### **3 SIP Structure**

The unit of submission to the archive is the E-journal issue. An issue can be loosely defined as a publisher-specified aggregation of citable content items, such as articles, editorials, reviews, correspondence, errata, etc. Note that this definition does not imply a fixed periodic publication schedule for the journal.

The component items of an issue are manifested in physical files. There is no requirement that the content of an item must always be completely encapsulated in a single file. In general, an item's content will be manifested in multiple files insofar as different portions of that content require independent preservation paths over archival time-spans.

The file components of the SIP are created and populated within a structured file system directory hierarchy and are then aggregated into a single file for transmission to the archive.

#### **3.1 SIP Components**

The individual component files of the SIP fall into two categories: content and metadata. Content files are the primary carriers of the intellectual meaning of the issue. Metadata files provide pertinent information about the content and the content files themselves. Content and metadata occur at two separate levels: issue and item-level.

Issue-level content includes bibliographic information and other editorial material such as the masthead, cover image, table of contents, board of editors, and editorial policy statement. Item-level content includes citable scholarly material such as peer-reviewed articles, editorials, reviews, correspondence, and errata, as well as supplementary, though pertinent materials such as data files associated with individual items.

Descriptive metadata provides information useful for resource discovery at the issue and item level. Administrative metadata encompasses rights metadata, which provides information concerning the intellectual property rights associated with the journal issue and its individual item-level components; provenance metadata, which provides information concerning the creation and fixity of those components; and technical metadata, which provides information useful for the archival preservation and delivery of issue and item-level components. Structural metadata provides the necessary information to successfully re-aggregate the individual file components of the SIP into unified issue items and, ultimately, an issue.

#### **3.2 Normative Data Formats**

All SIP metadata is encoded in XML files using namespace-qualified elements defined in the archive's EJAR (E-Journal ARchive) metadata schema. In general, textual content, including non-verbal notation such as mathematical formulæ, is also encoded in XML files using the EJAR issue and item-level schemas. The exception to this occurs in the case of item-level content being included in the SIP in a page description format such as PDF, which can encapsulate arbitrary aggregations of text and images.

Any given SIP component (e.g., metadata, content text or image) fulfills some specific function in regard to encapsulating an issue's intellectual content. The archive has established normative data formats for these various component functions in the SIP, based on criteria of open standards, mature and robust technology, generality of representation and ability to capture the highest resolution for given media types, creation as far upstream in the production process as possible, and suitability for preservation over archival time-spans. These formats are those used by the archive for its internal data storage and are the preferred formats for submission of SIP components.

The normative data formats are listed in the following table. A more complete descriptive registry of these formats is provided in Appendix B.

<b>Component Function</b>	<b>Normative Format</b>
Structural Metadata	XML 1.0 (METS 1.0 schema)
Descriptive Metadata	XML 1.0 (EJAR 1.0 descriptive and administrative metadata schema)
Administrative Metadata	
Issue-level Text	XML 1.0 (EJAR 1.0 issue schema)
Item-level Text	XML 1.0 (EJAR 1.0 item schema)
Item Reference Links	XML 1.0 (EJAR 1.0 reference links schema)
Raster Still Image	TIFF 6.0
Vector Still Image	XML 1.0 (SVG 1.0 schema)
Page Description	PDF 1.4
Audio	* AIFF 1.3
Video	* MPEG-2

Table 3.1 Normative data formats.

(\* At the time of publication these formats are being evaluated for inclusion in the normative category.)

All SIP components should be provided in the normative format established for that component's function. However, non-normative formats may be accepted in the SIP. Upon ingest, certain non-normative file components will be transformed into their normative analogues. The transformable non-normative formats are listed in Appendix C. SIP components supplied in other non-transformable, non-normative formats are accepted into the archive strictly under the proviso that they are not susceptible to archival preservation beyond the simple maintenance of their bit streams as supplied in the SIP, and that their usefulness upon delivery may become problematic over archival time-spans. In general, the non-transformable formats are those for which there does not exist a mechanism for transformation to a normative format that is essentially lossless with regard to the intellectual content of the underlying formatted data.

### 3.3 Character Set Issues

Within SIP components text characters can be grouped into three categories: ASCII; the non-ASCII characters in the Unicode repertoire (excluding the private use area); and characters outside of the Unicode repertoire [UNICODE]. Within XML-encoded files characters in these groups must be specified as follows:

<b>Category</b>	<b>Code Points</b>	<b>XML Encoding</b>
ASCII	U+0000 – U+007F	The ASCII character
Non-ASCII Unicode	U+0080 – U+DFFF, U+F900 – U+FFFD	Numeric character reference: <code>&amp;#xhhhh;</code> , where <i>hhhh</i> is the four digit hexadecimal representation of the character code point
Non-Unicode		Named character entity: <code>&amp;name;</code> , as defined by the archive's common entity set.

Table 3.3 XML character encoding

The archive's common entity set will contain all non-Unicode repertoire characters in use by registered archive content providers. Registration of additional characters in the entity set requires the specification of sufficient metadata to define the character's semantic and syntactic characteristics, including generation of an appropriately scaled glyph during delivery of the textual content.

### 3.4 Directory Structure

The SIP component files exist in a three-level file system directory hierarchy: a single empty directory corresponding to the journal title, which is the parent of a single sub-directory for all issue-level components, which is the common parent of sub-directories for every issue item.

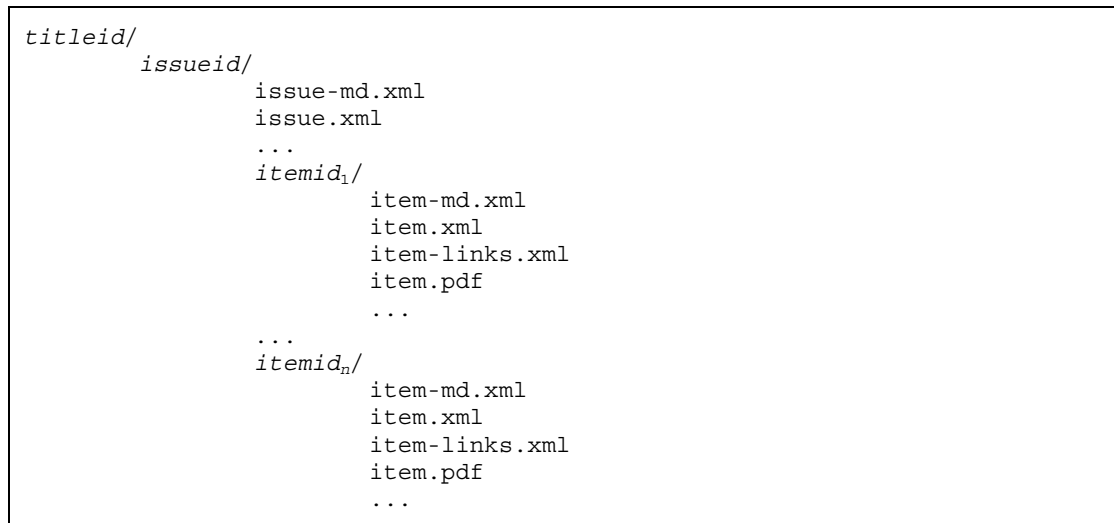


Fig 3.1 SIP directory structure.

The names given to these directories, *titleid*, *issueid*, and *itemid*, are used to identify the directories not only within the SIP but also to qualify file components within the archive. The format of these identifiers was designed to insure uniqueness within the archive, to be easily generated locally by individual SIP submitters, and to avoid conflict with file system and command shell semantics under the Unix, Windows, and Macintosh operating systems.

#### 3.4.1 Title-level directory

The title-level directory must be named with the journal's *titleid*, the SIP identifier for the journal title. The *titleid* must be the journal's ISSN numeric code in hyphenated form [ISSN]:

*nnnn-nnnn*

The ISSN should be that registered for the electronic edition of the journal title, if such an ISSN exists. Otherwise the ISSN of the print edition can be used. Note that no SIPs will be accepted by the archive for journal titles that do not have valid registered ISSNs. As part of the archive's administrative procedures, publishers must register all *titleids* with the archive prior to their use within a SIP.

The title-level directory is empty and serves only as the parent for the issue-level subdirectory.



### 3.4.2 Issue-level directory

The issue-level directory must be named with the issue's *issueid*. The *issueid* is based on the format defined for the chronology and enumeration fields of a SICI Item Segment [SICI]. However, within these fields every instance of a forward slash character, “/”, must be replaced with a hyphen character, “-”, and every instance of a colon character, “:”, must be replaced with a period character, “.”, in order to avoid potential collision with operating system-specific file system and command shell semantics. The *issueid* is then formed by concatenating the two fields together separated by an underscore character, “\_”:

*chron\_enum*

For example:

SICI:	0361-526X(199021/22)17:3/4<>1.0.CO;2-P
<i>issueid</i> :	199021-22_17.3-4

Ex. 3.1 SICI to *issueid* mapping.

Unlike a SICI, both the chronology and enumeration components of an *issueid* are required to be non-null. The minimal valid chronology is the year of publication. If no enumeration is associated with a published journal issue, the SIP submitter must create an enumeration designation such that the qualified directory path *titleid/issueid* is unique in the archive; in other words, it must be unique with regard to all previously deposited SIPs.

The issue-level directory contains, at a minimum, the issue-level metadata and textual content files, and optionally, additional issue-level content files, such as a cover image. This directory serves as the parent of the item-level subdirectories.

### 3.4.3 Item-level directories

Each item-level directory must be named with that item's *itemid*. The *itemid* should be that item's DOI, if one is registered for the item [DOI]. However, every instance of a forward slash character, “/”, within the DOI must be replaced with a hyphen character, “-”, in order to avoid potential collision with operating system-specific file system and command shell semantics:

10.reg-dss

where *reg* and *dss* are the numeric Registrant Code and DOI Suffix String components of the DOI, respectively. For example:

DOI:	10.1000/186
<i>itemid</i> :	10.1000-186

Ex. 3.2 DOI to *itemid* mapping.

All non-printable US-ASCII characters in the DOI must be encoded using the URI %-escape mechanism under which each non-ASCII character is replaced with the string “%xx” where *xx* is the hexadecimal representation of the non-ASCII character's ISO/IEC 10646 code point [URI].

If no DOI is registered for the item, a publisher-specified identifier can be used. This identifier must be unique with respect to case insensitivity among all items in a given SIP. The identifier must be a US-ASCII alphanumeric character string and must not mimic the syntactic structure of a SIP DOI-based item

identifier, i.e., it must not begin with the prefix “10.” followed by hyphen-separated numeric and character strings.

Each item-level directory contains, at a minimum, the item-level metadata and textual content files, and optionally, additional item-level content files, such as figures and supplementary content files.

### 3.5 Issue-Level File Components

The issue-level directory will always contain at least two files: a metadata file (`issue-md.xml`) and an issue-level textual content file (`issue.xml`).

The textual content file, `issue.xml`, aggregates all issue-level textual material. This includes the issue masthead information, copyright statement, table of contents, board of editors, and statement of editorial policy. Additional non-textual issue-level content data, such as a cover image, may be present in individual content files.

#### 3.5.1 Issue-level metadata file

The issue-level metadata file (`issue-md.xml`) is XML-encoded according to the METS schema [METS], as described in Appendix A. This file contains descriptive, administrative, and structural metadata related to the issue and all issue-level SIP components. An example issue-level metadata file is provided in Appendix A.1.

The descriptive, administrative, and structural metadata, and file descriptors for individual issue-level SIP components are defined independently in the METS framework. The appropriate relationships between these independent sections are specified using the standard XML ID/IDREF mechanism [XML]. These sections are each given a unique two-part identifier: an initial section type, and a unique string to identify the particular component; alphabetic for standard components, numeric for optional components.

*section:component*

Type prefixes are defined for descriptive and administrative metadata, and for file descriptor sections:

<i>section</i>	<b>Section</b>
<code>descr</code>	Descriptive metadata (<dmdSec> element)
<code>admin</code>	Administrative metadata (<admSec> element)
<code>file</code>	File descriptor (<file> element)

Table 3.3 Issue-level section type identifiers.

<i>component</i>	<b>Component</b>
<code>issue</code>	Entire issue
<code>issue-content</code>	Issue textual content file ( <code>issue.xml</code> )
<code>n</code>	Numeric identifier defined sequentially for optional file components: 1, 2, ...

Table 3.4 Issue-level section component identifiers.

Thus the identifier for the issue descriptive metadata is `descr:issue`; the identifiers for the administrative metadata and file descriptor for the issue-level textual content file are `admin:issue-content` and `file:issue-content`; the identifiers for the administrative metadata and file descriptor for the first optional file component are `admin:1` and `file:1`, etc.

The SIP-specific element tags and attributes of the issue-level metadata file are defined in the following tables. Elements without any defined content are empty elements: `<tag attr="value" ... />`.

### 3.5.1.1 METS root

The <mets> element is the root element of the issue-level metadata file. It is the structural parent of all other elements.

<mets>		METS root element
<i>Attributes</i>	TYPE="EJARISSUE-version" OBJID="issueid" LABEL="citation" PROFILE="EJAR"	SIP version identifier in the form <i>major.minor</i> Issue-level identifier Issue-level bibliographic citation
<i>Content</i>	<metsHdr> <dmdSec> <admSec> ... <fileSec> <structMap>	

### 3.5.1.2 METS header

The <metsHdr> element captures metadata relative to the issue-level metadata file itself, rather than to the abstract SIP issue object.

<metsHdr>		METS header
<i>Attributes</i>	CREATEDATE="yyyy-mm-dd"	issue-md.xml file creation date
<i>Content</i>	<agent>	
<i>Parent</i>	<mets>	

<agent>		Depositing agent
<i>Attributes</i>	ROLE="CREATOR" TYPE="ORGANIZATION"	
<i>Content</i>	<name>	
<i>Parent</i>	<metsHdr>	

<name>		Depositing agent name
<i>Content</i>	<i>publisher</i>	Registered name of content provider
<i>Parent</i>	<agent>	

### 3.5.1.3 Issue-level descriptive metadata

The <dmdSec> captures descriptive metadata for the issue as a whole.

<dmdSec>		Issue descriptive metadata
<i>Attributes</i>	ID="descr:issue"	
<i>Content</i>	<mdWrap> <ejar:...>	Wrapped issue-level descriptive metadata as defined by the EJAR descriptive metadata schema
<i>Parent</i>	<mets>	

The <mdWrap> element is defined in Appendix A. The <ejar:...> element represents as-yet-undefined descriptive metadata.

### 3.5.1.4 Issue-level administrative metadata

Administrative metadata is captured at varying levels of granularity corresponding to the issue as a whole and to the specific file components of the issue.

#### 3.5.1.4.1 Issue administrative metadata

This <admSec> element captures administrative metadata about the issue as a whole.

<admSec>		Issue administrative metadata
<i>Attributes</i>	ID="admin:issue"	
<i>Content</i>	<rightsMD>	
<i>Parent</i>	<mets>	

<rightsMD>		Issue rights metadata
<i>Content</i>	<mdWrap> <ejar:...>	Wrapped issue digital rights management metadata as defined by the EJAR administrative metadata schema
<i>Parent</i>	<admSec ID="admin:issue">	

The <mdWrap> element is defined in Appendix A. The <ejar:...> element represents as-yet-undefined administrative metadata.

#### 3.5.1.4.2 Issue textual content administrative metadata

This <admSec> element captures administrative metadata about the issue-level textual content file (issue.xml).

<admSec>		Issue textual content administrative metadata
<i>Attributes</i>	ID="admin:issue-content"	
<i>Content</i>	<techMD> <digiprovMD>	
<i>Parent</i>	<mets>	

<techMD>		Issue textual content technical metadata
<i>Content</i>	<mdWrap> <ejar:...>	Wrapped technical metadata as defined by the EJAR administrative metadata schema
<i>Parent</i>	<admSec ID="admin:issue-content">	

<digiprovMD>		Issue textual content provenance metadata
<i>Content</i>	<mdWrap> <ejar:...>	Wrapped checksum as defined by the EJAR administrative metadata schema
<i>Parent</i>	<admSec ID="admin:issue-content">	

The <mdWrap> element is defined in Appendix A. The <ejar:...> element represents as-yet-undefined administrative metadata.

### 3.5.1.4.3 Optional issue-level component administrative metadata

Additional <admSec> elements are required to capture administrative metadata about any other optional issue-level content files.

<admSec>		Optional issue-level component administrative metadata
<i>Attributes</i>	ID="admin:n"	Sequentially assigned numeric identifier: 1, 2, ...
<i>Content</i>	<techMD> <rightsMD> <digiprovdMD>	
<i>Parent</i>	<mets>	

<techMD>		Optional issue-level component technical metadata
<i>Content</i>	<mdWrap> <...>	Wrapped component technical metadata as defined by the EJAR administrative schema or by other appropriate external schemas incorporated normatively into this specification
<i>Parent</i>	<admSec ID="admin:n" >	

<digiprovdMD>		Optional issue-level component provenance metadata
<i>Content</i>	<mdWrap> <ejar:...>	Wrapped component checksum as defined by the EJAR administrative metadata schema
<i>Parent</i>	<admSec ID="admin:n" >	

<rightsMD>		Optional issue-level component rights metadata
<i>Content</i>	<mdWrap> <ejar:...>	Wrapped component digital rights management metadata as defined by the EJAR administrative metadata schema
<i>Parent</i>	<admSec ID="admin:n" >	

This <rightsMD> element is only required if the optional issue-level component has different digital rights management metadata than that defined for the issue as a whole.

The <mdWrap> element is defined in Appendix A. The <ejar:...> element represents as-yet-undefined administrative metadata.

### 3.5.1.5 Issue-level file inventory

<fileSec>		Issue-level file inventory section
<i>Content</i>	<fileGrp>	
<i>Parent</i>	<mets>	

<fileGrp>		Issue-level file inventory
<i>Attributes</i>	ADMID="admin:issue"	
<i>Content</i>	<file> ...	Issue-level file descriptors
<i>Parent</i>	<fileSec>	

### 3.5.1.5.1 Issue textual content file descriptor

<b>&lt;file&gt;</b>		Issue textual content file descriptor
<b>Attributes</b>	ID="file:issue-content" ADMID="admin:issue-content" CREATED="yyyy-mm-dd" MIMETYPE="text/xml" OWNERID="id" SIZE="n"	File creation date  Optional publisher-specific local identifier File size in bytes
<b>Content</b>	<FLocat>	
<b>Parent</b>	<fileGrp>	

<b>&lt;FLocat&gt;</b>		Issue textual content file locator
<b>Attributes</b>	xlink:type="simple" xlink:href="issue.xml"	
<b>Parent</b>	<file ID="file:issue-content" >	

### 3.5.1.5.2 Optional issue-level file component descriptor

Additional <file> elements are required to capture file descriptor information about any other optional issue-level content files.

<b>&lt;file&gt;</b>		Optional issue-level file component descriptor
<b>Attributes</b>	ID="file:n" ADMID="admin:n" CREATED="yyyy-mm-dd" MIMETYPE="mime" OWNERID="id" SIZE="n"	Sequentially assigned numeric identifier: 1, 2, ...  File creation date File MIME type Optional publisher-specific local identifier File size in bytes
<b>Content</b>	<FLocat>	
<b>Parent</b>	<fileGrp>	

<b>&lt;FLocat&gt;</b>		Optional issue-level file component locator
<b>Attributes</b>	xlink:type="simple" xlink:href="file.ext"	Component file name
<b>Parent</b>	<file ID="file:n" >	

### 3.5.1.6 Issue-level structural metadata

The <structMap> element captures the structural relationships that occur amongst the various individual components of the SIP issue, arranged in a series of nested divisions.

<b>&lt;structMap&gt;</b>		Issue-level structural metadata
<b>Attributes</b>	type="LOGICAL"	
<b>Content</b>	<div TYPE="EJARISSUE" >	
<b>Parent</b>	<mets>	

### 3.5.1.6.1 Issue division

This top-most <div> element represents the issue as a whole.

<div>		Issue-level division
<b>Attributes</b>	TYPE="EJARISSUE" ADMID="admin:issue" DMD="descr:issue" LABEL="citation"	Issue-level bibliographic citation
<b>Content</b>	<fptr FILEID= "issue-content"> <fptr FILEID="file:n" <div TYPE="EJARSECTION"> ...>	Sequentially assigned numeric identifier: 1, 2, ...
<b>Parent</b>	<structMap>	

<fptr>		Issue textual content file pointer
<b>Attributes</b>	FILEID="file:issue-content"	
<b>Parent</b>	<div TYPE="EJARISSUE">	

Additional <fptr> elements are required to reference file descriptor information about any other optional issue-level content files.

<fptr>		Optional issue-level file pointer
<b>Attributes</b>	FILEID="file:n"	
<b>Parent</b>	<div TYPE="EJARISSUE">	

### 3.5.1.6.2 Section divisions

Intermediate <div> elements are required to represent typed aggregations of individual items, for example, correspondence, articles, reviews, errata, etc, as would appear on the issue table of contents.

<div>		Item-level section division
<b>Attributes</b>	TYPE="EJARSECTION" LABEL="section"  ORDER="n"	Item section label, e.g., Correspondence, Articles, Reviews, Errata, etc. Item section ordinal position: 1, 2, ...
<b>Content</b>	<div TYPE="EJARITEM"> ...>	
<b>Parent</b>	<div TYPE="EJARISSUE">	

### 3.5.1.6.3 Item-level divisions

Terminal `<div>` elements are required to wrap the pointers to the individual item-level metadata files (*itemid/item-md.xml*).

<code>&lt;div&gt;</code>		Item-level division
<i>Attributes</i>	TYPE="EJARITEM" LABEL="citation" ORDERLABEL="n"	Item-level bibliographic citation Item position within section, e.g., page number, ordinal position, etc.
<i>Content</i>	<code>&lt;mptr&gt;</code>	
<i>Parent</i>	<code>&lt;div TYPE="EJARSECTION"&gt;</code>	

<code>&lt;mptr&gt;</code>		Item-level METS file pointer
<i>Attributes</i>	xlink:type="simple" xlink:href= "itemid/item-md.xml"	Path name of item-level metadata file relative to issue-level directory
<i>Parent</i>	<code>&lt;div TYPE="EJARSECTION"&gt;</code>	

### 3.5.2 Issue-level content files

The issue-level textual content file (*issue.xml*) is XML-encoded according to the archive EJAR issue schema.

Other content files may exist to provide additional issue-level content, such as a cover image. For the purposes of these additional files, the file names *issue-md.xml* and *issue.xml* are reserved. Although file names can be composed using both upper and lower-case characters, all file names must be unique with regard to case-insensitivity. The file extensions established for the normative and non-normative formats specified in Appendices B and C must be used.

Mathematical notation occurring in the issue-level textual content must be encoded into XML in the format specified by the MathML 2.0 specification [MATHML].



### 3.6 Item-Level File Components

Each item-level directory will always contain at least two files: a metadata file (`item-md.xml`) and either an item-level textual content file (`item.xml`) or page description content file (`item.pdf`). Even if an XML-encoded textual content file is supplied, a single page description file encapsulating the entire item should also be provided, if possible. (This file will help to facilitate the quality assurance phase of SIP ingest.) If no XML-encoded textual content file is supplied, then the entire item must be encapsulated into a single page description file. If XML-encoded textual content is supplied, all item-level reference links are aggregated into a single XML-encoded linkage file (`item-links.xml`). Additional non-textual item-level content data such as figures, or supplemental material such as data files, may be present in individual content files.

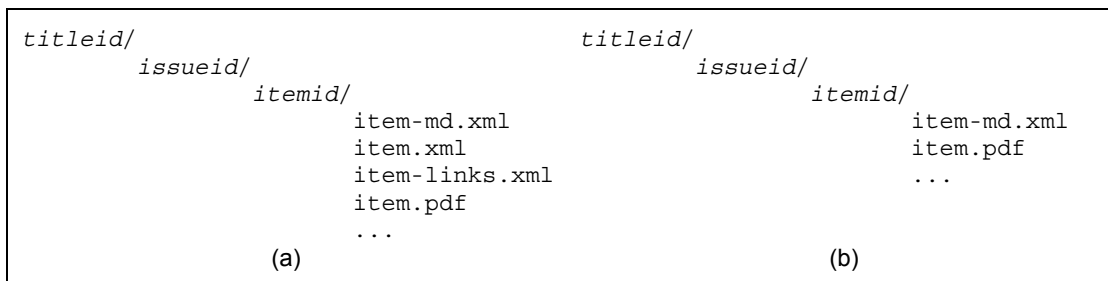


Fig. 3.3 Item-level file components: (a) a single XML-encoded textual content file and optional supplemental files; or (b) a single PDF file encapsulating all primary item content; supplemental material is still present in additional files.

#### 3.6.1 Item-level metadata file

The item-level metadata file is XML-encoded according to the METS schema, as described in Appendix A. This file contains descriptive, administrative, and technical metadata related to the item and all item-level SIP components. An example item-level metadata file is provided in Section A.2.

The descriptive, administrative, and structural metadata, and file descriptors for individual item-level SIP components are defined independently in the METS framework. The appropriate relationships between these independent sections are specified using the standard XML ID/IDREF mechanism [XML]. These sections are each given a unique two-part identifier: an initial section type, and a unique string to identify the particular component; alphabetic for standard components, numeric for optional components.

*section:component*

Type prefixes are defined for descriptive and administrative metadata, and for file descriptor sections:

<i>section</i>	<b>Section</b>
descr	Descriptive metadata (<dmSec> element)
admin	Administrative metadata (<admSec> element)
file	File descriptor (<file> element)

Table 3.5 Item-level section type identifiers.

<i>component</i>	<b>Component</b>
item	Entire issue
item-content	Item XML-encoded textual content file ( <code>item.xml</code> )
item-page	Item page description content file ( <code>item.pdf</code> )
item-links	Item reference links file ( <code>item-links.xml</code> )
<i>n</i>	Numeric identifier defined sequentially for optional file components: 1, 2, ...

Table 3.6 Item-level section component identifiers.

Thus the identifier for the item descriptive metadata is `descr:item`; the identifiers for the administrative metadata and file descriptor for the item-level XML-encoded textual content file are `admin:item-content` and `file:item-content`; the identifiers for the administrative metadata and file descriptor for the first optional file component are `admin:1` and `file:1`, etc.

The SIP-specific element tags and attributes of the item-level metadata file are defined in the following tables. Elements without any defined content are empty elements: `<tag attr="value" ... />`.

### 3.6.1.1 METS root

The `<mets>` element is the root element of the item-level metadata file. It is the structural parent of all other elements.

<code>&lt;mets&gt;</code>		METS root element
<b>Attributes</b>	TYPE="EJARITEM-version" OBJID="itemid" LABEL="citation" PROFILE="EJAR"	SIP version identifier in the form <i>major.minor</i> Item-level identifier Item-level bibliographic citation
<b>Content</b>	<code>&lt;metsHdr&gt;</code> <code>&lt;dmdSec&gt;</code> <code>&lt;admSec&gt;</code> ... <code>&lt;fileSec&gt;</code> <code>&lt;structMap&gt;</code>	

### 3.6.1.2 METS header

The `<metsHdr>` element captures metadata relative to the item-level metadata file itself, rather than to the abstract SIP item object.

<code>&lt;metsHdr&gt;</code>		METS header
<b>Attributes</b>	CREATEDATE="yyyy-mm-dd"	item-md.xml file creation date
<b>Content</b>	<code>&lt;agent&gt;</code>	
<b>Parent</b>	<code>&lt;mets&gt;</code>	

<code>&lt;agent&gt;</code>		Depositing agent
<b>Attributes</b>	ROLE="CREATOR" TYPE="ORGANIZATION"	
<b>Content</b>	<code>&lt;name&gt;</code>	
<b>Parent</b>	<code>&lt;metsHdr&gt;</code>	

<code>&lt;name&gt;</code>		Depositing agent name
<b>Content</b>	<i>publisher</i>	Registered name of content provider
<b>Parent</b>	<code>&lt;agent&gt;</code>	

### 3.6.1.3 Item-level descriptive metadata

The <dmdSec> captures descriptive metadata for the item as a whole.

<dmdSec>		Item descriptive metadata
<i>Attributes</i>	ID="descr:item"	
<i>Content</i>	<mdWrap> <ejar:...>	Wrapped item-level descriptive metadata as defined by the EJAR descriptive metadata schema
<i>Parent</i>	<mets>	

The <mdWrap> element is defined in Appendix A. The <ejar:...> element represents as-yet-undefined descriptive metadata.

### 3.6.1.4 Item-level administrative metadata

Administrative metadata is captured at very levels of granularity corresponding to the item as a whole and the specific file components of the item.

#### 3.6.1.4.1 Item administrative metadata

This <admSec> element captures administrative metadata about the item as a whole.

<admSec>		Item administrative metadata
<i>Attributes</i>	ID="admin:item"	
<i>Content</i>	<rightsMD>	
<i>Parent</i>	<mets>	

<rightsMD>		Item rights metadata
<i>Content</i>	<mdWrap> <ejar:...>	
<i>Parent</i>	<admSec ID="admin:item">	

The <mdWrap> element is defined in Appendix A. The <ejar:...> element represents as-yet-undefined administrative metadata.

#### 3.6.1.4.2 Item textual content administrative metadata

This <admSec> element captures administrative metadata about the item-level XML-encoded textual content file (item.xml).

<admSec>		Item textual content administrative metadata
<i>Attributes</i>	ID="admin:item-content"	
<i>Content</i>	<techMD> <digiprovMD>	
<i>Parent</i>	<mets>	

<techMD>		Item textual content technical metadata
<i>Content</i>	<mdWrap> <ejar:...>	Wrapped technical metadata as defined by the EJAR administrative metadata schema
<i>Parent</i>	<admSec ID="admin:item-content">	

<digiprovMD>		Item textual content provenance metadata
<b>Content</b>	<mdWrap> <ejar:...>	Wrapped checksum as defined by the EJAR administrative metadata schema
<b>Parent</b>	<admSec ID="admin:item-content">	

The <mdWrap> element is defined in Appendix A. The <ejar:...> element represents as-yet-undefined administrative metadata.

#### 3.6.1.4.3 Item page description administrative metadata

This <admSec> element captures administrative metadata about the item-level page description file (item.pdf).

<admSec>		Item page description administrative metadata
<b>Attributes</b>	ID="admin:item-page"	
<b>Content</b>	<techMD> <digiprovMD>	
<b>Parent</b>	<mets>	

<techMD>		Item page description technical metadata
<b>Content</b>	<mdWrap> <ejar:...>	Wrapped technical metadata as defined by the EJAR administrative metadata schema
<b>Parent</b>	<admSec ID="admin:item-page">	

<digiprovMD>		Item page description provenance metadata
<b>Content</b>	<mdWrap> <ejar:...>	Wrapped checksum as defined by the EJAR administrative metadata schema
<b>Parent</b>	<admSec ID="admin:item-page">	

The <mdWrap> element is defined in Appendix A. The <ejar:...> element represents as-yet-undefined administrative metadata.

#### 3.6.1.4.4 Item reference links administrative metadata

This <admSec> element captures administrative metadata about the item reference links file (item-links.xml).

<admSec>		Item reference links administrative metadata
<b>Attributes</b>	ID="admin:item-links"	
<b>Content</b>	<techMD> <digiprovMD>	
<b>Parent</b>	<mets>	

<techMD>		Item reference links technical metadata
<b>Content</b>	<mdWrap> <ejar:...>	Wrapped technical metadata as defined by the EJAR administrative metadata schema
<b>Parent</b>	<admSec ID="admin:item-links">	

<digiprovMD>		Item reference links provenance metadata
<b>Content</b>	<mdWrap> <ejar:...>	Wrapped checksum as defined by the EJAR administrative metadata schema
<b>Parent</b>	<admSec ID="admin:item-links">	

The <mdWrap> element is defined in Appendix A. The <ejar:...> element represents as-yet-undefined administrative metadata.

### 3.6.1.4.5 Optional item-level component administrative metadata

Additional <admSec> elements are required to capture administrative metadata about any other optional item-level content files.

<admSec>		Optional item-level component administrative metadata
<i>Attributes</i>	ID="admin:n"	Sequentially assigned numeric identifier: 1, 2, ...
<i>Content</i>	<techMD> <rightsMD> <digiprovMD>	
<i>Parent</i>	<mets>	

<techMD>		Optional item-level component technical metadata
<i>Content</i>	<mdWrap> <...>	Wrapped component technical metadata as defined by the EJAR administrative schema or by other appropriate external schemas incorporated normatively into this specification
<i>Parent</i>	<admSec ID="admin:n" >	

<digiprovMD>		Optional item-level component provenance metadata
<i>Content</i>	<mdWrap> <ejar:...>	Wrapped component checksum as defined by the EJAR administrative metadata schema
<i>Parent</i>	<admSec ID="admin:n" >	

<rightsMD>		Optional item-level component rights metadata
<i>Content</i>	<mdWrap> <ejar:...>	Wrapped component digital rights management metadata as defined by the EJAR administrative metadata schema
<i>Parent</i>	<admSec ID="admin:n" >	

The <mdWrap> element is defined in Appendix A. The <ejar:...> element represents as-yet-undefined administrative metadata.

### 3.6.1.5 Item-level file inventory

<fileSec>		Item-level file inventory section
<i>Content</i>	<fileGrp>	
<i>Parent</i>	<mets>	

<fileGrp>		Item-level file inventory
<i>Attributes</i>	ADMID="admin:item"	
<i>Content</i>	<file> ...	Item-level file descriptors
<i>Parent</i>	<fileSec>	

The <fileGrp> element will contain either: (1) one to three <file> elements, one each for the item XML-encoded textual content file (*item.xml*), and optionally, the item page description and reference

links files (`item-links.xml`); or (2) one `<file>` element for a single item page description file (`item.pdf`).

### 3.6.1.5.1 Item XML-encoded textual content, page description, and reference links file descriptors

Either one to three `<file>` elements must appear together, one each for the XML-encoded textual content file, and optionally, the page description and reference links files; or one `<file>` element must appear by itself for a single page description.

<code>&lt;file&gt;</code>		Item XML-encoded textual content file descriptor
<b>Attributes</b>	ID="file:item-content" ADMID="admin:item-content" CREATED="yyyy-mm-dd" MIMETYPE="text/xml" OWNERID="id" SIZE="n"	File creation date  Optional publisher-specific local identifier File size in bytes
<b>Content</b>	<code>&lt;FLocat&gt;</code>	
<b>Parent</b>	<code>&lt;fileGrp&gt;</code>	

<code>&lt;FLocat&gt;</code>		Item XML-encoded textual content file locator
<b>Attributes</b>	xlink:type="simple" xlink:href="item.xml"	
<b>Parent</b>	<code>&lt;file ID="file:item-content"&gt;</code>	

<code>&lt;file&gt;</code>		Item page description file descriptor
<b>Attributes</b>	ID="file:item-page" ADMID="admin:item-page" CREATED="yyyy-mm-dd" MIMETYPE="application/pdf" OWNERID="id" SIZE="n"	File creation date  Optional publisher-specific local identifier File size in bytes
<b>Content</b>	<code>&lt;FLocat&gt;</code>	
<b>Parent</b>	<code>&lt;fileGrp&gt;</code>	

<code>&lt;FLocat&gt;</code>		Item page description file locator
<b>Attributes</b>	xlink:type="simple" xlink:href="item.pdf"	
<b>Parent</b>	<code>&lt;file ID="file:item-page"&gt;</code>	

<code>&lt;file&gt;</code>		Item reference links file descriptor
<b>Attributes</b>	ID="file:item-links" ADMID="admin:item-links" CREATED="yyyy-mm-dd" MIMETYPE="text/xml" OWNERID="id" SIZE="n"	File creation date  Optional publisher-specific local identifier File size in bytes
<b>Content</b>	<code>&lt;FLocat&gt;</code>	
<b>Parent</b>	<code>&lt;fileGrp&gt;</code>	

<code>&lt;FLocat&gt;</code>		Item reference links file locator
<b>Attributes</b>	xlink:type="simple" xlink:href="item-links.xml"	
<b>Parent</b>	<code>&lt;file ID="file:item-links"&gt;</code>	

### 3.6.1.5.2 Optional item-level content file descriptor

Additional <file> elements are required to capture file descriptor information about any other optional item-level content files.

<file>		Optional item-level file component descriptor
<b>Attributes</b>	ID="file:n" ADMID="admin:n" CREATED="yyyy-mm-dd" MIMETYPE="mime" OWNERID="id" SIZE="n"	Sequentially assigned numeric identifier: 1, 2, ...  File creation date File MIME type Optional publisher-specific local identifier File size in bytes
<b>Content</b>	<FLocat>	
<b>Parent</b>	<fileGrp>	

<FLocat>		Optional item-level file component locator
<b>Attributes</b>	xlink:type="simple" xlink:href="file.ext"	Component file name
<b>Parent</b>	<file ID="file:n">	

### 3.6.1.6 Item-level structural metadata

At the item-level the only structural relationship between individual components is aggregation, thus the <structMap> element contains a single item-level division as the parent of the file referencing <fptr> elements.

<structMap>		Item-level structural metadata
<b>Attributes</b>	type="LOGICAL"	
<b>Content</b>	<div>	
<b>Parent</b>	<mets>	

<div>		Item level division
<b>Attributes</b>	TYPE="EJARITEM" ADMID="admin:item" DMD="descr:item" LABEL="citation"	Item-level bibliographic citation
<b>Content</b>	<fptr FILEID="item-content"> <fptr FILEID="item-page"> <fptr FILEID="item-links"> <fptr FILEID="file:n"> ...	
<b>Parent</b>	<structMap>	

<fptr>		Item XML-encoded textual content file pointer
<b>Attributes</b>	FILEID="file:item-content"	
<b>Parent</b>	<div>	

<fptr>		Item page description file pointer
<b>Attributes</b>	FILEID="file:item-page"	
<b>Parent</b>	<div>	

<fptr>		Item textual content file pointer
<i>Attributes</i>	FILEID="file:item-links"	
<i>Parent</i>	<div>	

<fptr>		Optional item-level file pointer
<i>Attributes</i>	FILEID="file:n"	
<i>Parent</i>	<div>	

### 3.6.2 Item-level content files

Item-level content must be either unencapsulated in a set of files; or entirely encapsulated in a single page description file (`item.pdf`).

Under the preferred, unencapsulated scheme, one of item-level files must be a textual content file (`item.xml`), XML-encoded according to the archive EJAR item schema. The existence of any other item-level files is optional. Even if an XML-encoded textual content file is provided, a single page description file (`item.pdf`) encapsulating the entire item should also be supplied, if possible. (This file will help to facilitate the quality assurance phase of SIP ingest.) For items that contain reference links, those links are contained in a linkage file (`item-links.xml`), XML-encoded according to the archive EJAR linkage schema. Additional non-textual item-level files might include figures or data files.

Other content files may exist to provide additional item-level content, such as figures. For the purposes of these additional files, the file names `item-md.xml`, `item.xml`, `item-links.xml`, and `item.pdf` are reserved. Although file names can be composed using both upper and lower-case characters, all file names must be unique with regard to case-insensitivity. The file extensions established for the normative and non-normative formats specified in Appendices B and C must be used.

Mathematical notation occurring in the item-level textual content must be encoded into XML in the format specified by the MathML 2.0 specification [MATHML].

### 3.7 SIP Aggregation

After the entire SIP directory structure has been created and fully populated, the SIP is aggregated into a single file and compressed.

```
cd titleid/..
tar cvf issueid.tar titleid/issueid
gzip issueid.tar
```

Fig. 3.3 SIP aggregation and compression (based on Unix tar and gzip).



## 4 Submission Session

A Submission Session (SS) is the operational process of submission of a SIP to the archive by a content provider.

The compressed aggregated SIP is transmitted to the archive via FTP [FTP]. All content providers registered with the archive are provided with a unique *publisherid* and password that must be used for FTP session authentication.

```
cd titleid/..
ftp ejar.harvard.edu
Name: publisherid
Password: *****
ftp> binary
ftp> cd titleid
ftp> send issueid.tar.gz
ftp> quit
```

Fig. 4.1 SIP Submission Session (based on a typical Unix command-line oriented ftp client).

Distribution of *publisherids* and passwords occurs as part of the administrative registration procedures between content providers and the archive.

### 4.1 SIP Ingest

After the archive has received a SIP transmitted by a publisher to the archive in an SS, the SIP is processed by the archive Ingest function. This processing includes decompression, disaggregation, verification of all appropriate file components, validation by format, and transformation to normative formats, if appropriate. Once all ingest processing is complete, the archive sends a submission confirmation to the publisher via e-mail indicating the success or failure of the ingest. Invalid SIPs that fail during ingest processing are not added to the archive; the publisher must resubmit the SIP with all necessary corrections as indicated in the negative confirmation message.

Following the OAIS functional model, valid SIPs successfully passing the Ingest function are transformed internally into AIPs (Archival Information Packages), which are then processed by the Archival Storage function. After the processing of an AIP by the Archive Storage function, the archive sends an archival confirmation to the publisher via e-mail. Only after receiving an affirmative archival confirmation is the publisher assured that the archive has assumed active archival responsibility for the E-journal content submitted in the SIP.

## A METS Framework

METS (Metadata Encoding and Transmission Standard) is a XML-formatted metadata framework for encoding descriptive, administrative, and structural metadata of digital library objects [METS]. It was developed as an initiative of the Digital Library Federation, and is built upon work previously performed for the DLF-funded Making of America II project coordinated at the University of California, Berkeley.

The METS mechanisms for defining structural metadata and synchronization were derived in part from TEI [TEI] and SMIL [SMIL]. METS is a metadata framework capturing structural relationships and providing containers for descriptive and administrative metadata encoded according to standards external to METS itself.

A METS file is composed of five main sections:

<b>Element</b>	<b>Section</b>
<code>&lt;metsHdr&gt;</code>	METS header, recording information about the METS file itself, not the object the file abstractly represents.
<code>&lt;dmdSec&gt;</code>	Descriptive metadata section.
<code>&lt;admSec&gt;</code>	Administrative metadata section.
<code>&lt;fileSec&gt;</code>	File inventory section.
<code>&lt;structMap&gt;</code>	Structural metadata section.

Table A.1 METS main sections

For the purposes of the SIP, the issue and item-level metadata files (`issue-md.xml` and `item-md.xml`) have the following general structure, with global scoping of the METS, EJAR metadata, XML Schema, and XLink namespaces, is:

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<mets xmlns="http://www.loc.gov/METS/"
      xmlns:ejar="http://hul.harvard.edu/EJAR/METADATA"
      xmlns:xlink="http://www.w3.org/1999/xlink"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.loc.gov/METS
                        http://www.loc.gov/standards/mets/mets.xsd"
      xsi:schemaLocation="http://hul.harvard.edu/EJAR/METADATA
                        http://hul.harvard.edu/EJAR/metadata.xsd"
...>
<metsHdr ...> <!-- METS file metadata --> </metsHdr>
<dmdSec ...> <!-- descriptive metadata --> </dmdSec>
<admSec ...> <!-- administrative metadata --> </admSec>
...
<fileSec> <!-- file inventory --> </fileSec>
<structMap ...> <!-- structural metadata --> </structMap>
</mets>
```

Fig. A.1 METS structure.

METS does not define the form or encoding of descriptive or administrative metadata elements. Instead, it provides containers into which externally defined metadata can be placed. XML-encoded metadata can be inserted directly into a METS file by placing it within an `<mdWrap>` metadata wrapper element:

<mdWrap>		Metadata wrapper
<i>Attributes</i>	MIMETYPE="text/xml" MDTYPE="type" OTHERMDTYPE="othertype"	METS registered metadata type, or OTHER If MDTYPE="OTHER", non-registered metadata type
<i>Content</i>	<...>	Externally defined metadata
<i>Parent</i>	<dmdSec> <techMD> <rightsMD> <digiprovMD>	

The following metadata types are used in SIP issue and item-level metadata files:

<i>MDTYPE</i>	<i>OTHERMDTYPE</i>	<i>Metadata Type</i>
NISOIMG		NISO raster still image metadata schema [NISOIMG]
OTHER	EJAR	EJAR descriptive or administrative metadata schema

Table A.1 SIP metadata types.

## A.1 SIP Issue-Level Metadata File (*issue-md.xml*)

The following informative example provides a general example of the form of an issue-level metadata file. Note that this example presupposes the existence of EJAR and NISO namespace-qualified metadata elements whose form has not been standardized as of the time of publication.

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<mets xmlns="http://www.loc.gov/METS/"
      xmlns:ejar="http://hul.harvard.edu/EJAR/METADATA/"
      xmlns:xlink="http://www.w3.org/1999/xlink"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.loc.gov/METS
                        http://www.loc.gov/standards/mets/mets.xsd"
      xsi:schemaLocation="http://hul.harvard.edu/EJAR/METADATA
                        http://hul.harvard.edu/EJAR/metadata.xsd"
      TYPE="EJARISSUE-major.minor" OBJID="issueid"
      LABEL="issue bibliographic citation" PROFILE="EJAR">

  <metsHdr CREATEDATE="yyyy-mm-dd">
    <agent ROLE="CREATOR" TYPE="ORGANIZATION">
      <name>content provider</name>
    </agent>
  </metsHdr>

  <dmdSec ID="descr:issue">
    <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
      <ejar:descr type="issue">issue descriptive metadata</ejar:descr/>
    </mdWrap>
  </dmdSec>

  <admSec ID="admin:issue">
    <rightsMD>
      <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
        <ejar:copyright>issue copyright metadata</ejar:copyright>
      </mdWrap>
    </rightsMD>
  </admSec>

  <admSec ID="admin:issue-content">
    <techMD>
      <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
        <ejar:tech type="TEXT">issue content technical metadata</ejar:tech>
      </mdWrap>
    </techMD>

    <digiprovMD>
      <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
        <ejar:checksum type="MD5">content file checksum</ejar:checksum>
      </mdWrap>
    </digiprovMD>
  </admSec>

  ...
```

Fig. A.2(a) SIP issue-level metadata file (*issue-md.xml*)  
(continued on next page).

```

...
<admSec ID="admin:1">
  <techMD>
    <mdWrap MIMETYPE="text/xml" MDTYPE="NISOIMG">
      <niso:...>cover image technical metadata</niso:...>
    </mdWrap>
  </techMD>

  <rightsMD>
    <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
      <ejar:copyright>cover image copyright metadata</ejar:copyright>
    </mdWrap>
  </rightsMD>

  <digiprovMD>
    <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
      <ejar:checksum type="MD5">cover image checksum</ejar:checksum>
    </mdWrap>
  </digiprovMD>
</admSec>

<fileSec>
  <fileGrp ADMID="admin:issue">
    <file ID="file:issue-content" ADMID="admin:issue-content"
      CREATED="yyyy-mm-dd" MIMETYPE="text/xml" OWNERID="id" SIZE="n">
      <Flocat xlink:type="simple" xlink:href="issue.xml"/>
    </file>

    <file ID="file:1" ADMID="admin:1" CREATED="yyyy-mm-dd"
      MIMETYPE="image/tiff" OWNERID="id" SIZE="n">
      <Flocat xlink:type="simple" xlink:href="cover.tif"/>
    </file>

    ...
  </fileGrp>
</fileSec>

<structMap TYPE="LOGICAL">
  <div TYPE="EJARISSUE" ADMID="admin:issue" DMD="descr:issue"
    LABEL="issue bibliographic citation">
    <fptr FILEID="file:issue-content"/>
    <fptr FILEID="file:1"/>

    <div TYPE="EJARSECTION" LABEL="section label" ORDER="n">
      <div TYPE="EJARITEM" LABEL="item bibliographic citation"
        ORDERLABEL="n">
        <mptr xlink:type="simple" xlink:href="itemid1/item-md.xml"/>
      </div>

      ...
    </div>

    ...
  </div>
</structMap>
</mets>

```

Fig. A.2(b) SIP issue-level metadata file (*issue-md.xml*)  
(continued from previous page).

## A.2 SIP Item-Level Metadata File (*item-md.xml*)

The following informative example provides a general example of the form of an item-level metadata file. Note that this example presupposes the existence of EJAR and NISO namespace-qualified metadata elements whose form has not been standardized as of the time of publication.

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<mets xmlns="http://www.loc.gov/METS/"
      xmlns:ejar="http://hul.harvard.edu/EJAR/METADATA"
      xmlns:xlink="http://www.w3.org/1999/xlink"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.loc.gov/METS
                        http://www.loc.gov/standards/mets/mets.xsd"
      xsi:schemaLocation="http://hul.harvard.edu/EJAR/METADATA
                        http://hul.harvard.edu/EJAR/metadata.xsd"
      TYPE="EJARITEM-major.minor" OBJID="itemid"
      LABEL="item bibliographic citation" PROFILE="EJAR">

  <metsHdr CREATEDATE="yyyy-mm-dd">
    <agent ROLE="CREATOR" TYPE="ORGANIZATION">
      <name>content provider</name>
    </agent>
  </metsHdr>

  <dmdSec ID="descr:item">
    <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
      <ejar:descr="item">item descriptive metadata</ejar:descr>
    </mdWrap>
  </dmdSec>

  <admSec ID="admin:item">
    <rightsMD>
      <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
        <ejar:copyright>item copyright metadata</ejar:copyright>
      </mdWrap>
    </rightsMD>
  </admSec>

  <admSec ID="admin:item-content">
    <techMD>
      <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
        <ejar:tech type="TEXT">item content technical metadata</ejar:tech>
      </mdWrap>
    </techMD>

    <digiprovMD>
      <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
        <ejar:checksum type="MD5">item content file checksum</ejar:checksum>
      </mdWrap>
    </digiprovMD>
  </admSec>

  ...
```

Fig. A.3(a) SIP item-level metadata file (*item-md.xml*)  
(continued onto next page).

```

...

<admSec ID="admin:item-page">
  <techMD>
    <mdWrap MIMETYPE="application/pdf" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
      <ejar:tech type="TEXT">item page technical metadata</ejar:tech>
    </mdWrap>
  </techMD>

  <digiprovMD>
    <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
      <ejar:checksum type="MD5">item page file checksum</ejar:checksum>
    </mdWrap>
  </digiprovMD>
</admSec>

<admSec ID="admin:item-links">
  <techMD>
    <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
      <ejar:tech type="LINKS">item links technical metadata</ejar:tech>
    </mdWrap>
  </techMD>

  <digiprovMD>
    <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
      <ejar:checksum type="MD5">item links checksum</ejar:checksum>
    </mdWrap>
  </digiprovMD>
</admSec>

<admSec ID="admin:1">
  <techMD>
    <mdWrap MIMETYPE="text/xml" MDTYPE="NISOIMG">
      <niso:...>image technical metadata...</niso:...>
    </mdWrap>
  </techMD>

  <rightsMD>
    <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
      <ejar:copyright>image copyright metadata</ejar:copyright>
    </mdWrap>
  </rightsMD>

  <digiprovMD>
    <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
      <ejar:checksum type="MD5">image checksum</ejar:checksum>
    </mdWrap>
  </digiprovMD>
</admSec>

<admSec ID="admin:2">
  <techMD>
    <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
      <ejar:tech type="DATA">data technical metadata</ejar:tech>
    </mdWrap>
  </techMD>

...

```

Fig. A.3(b) SIP item-level metadata file (item-md.xml)  
(continued from previous and onto next page).

```

...

<rightsMD>
  <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
    <ejar:copyright>data copyright metadata</ejar:copyright>
  </mdWrap>
</rightsMD>

<digiprovMD>
  <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="EJAR">
    <ejar:checksum type="MD5">data file checksum</ejar:checksum>
  </mdWrap>
</digiprovMD>
</admSec>

<fileSec>
  <fileGrp ADMID="admin:item">
    <file ID="file:item-content" ADMID="admin:item-content"
      CREATED="yyyy-mm-dd" MIMETYPE="text/xml" OWNERID="id" SIZE="n">
      <Flocat xlink:type="simple" xlink:href="item.xml"/>
    </file>

    <file ID="file:item-page" ADMID="admin:item-page"
      CREATED="yyyy-mm-dd" MIMETYPE="application/pdf" OWNERID="id"
      SIZE="n">
      <Flocat xlink:type="simple" xlink:href="item.pdf"/>
    </file>

    <file ID="file:item-links" ADMID="admin:item-links"
      CREATED="yyyy-mm-dd" MIMETYPE="text/xml" OWNERID="id" SIZE="n">
      <Flocat xlink:type="simple" xlink:href="item-links.xml"/>
    </file>

    <file ID="file:1" ADMID="admin:1" CREATED="yyyy-mm-dd"
      MIMETYPE="image/tiff" OWNERID="id" SIZE="n">
      <Flocat xlink:type="simple" xlink:href="fig1.tif"/>
    </file>

    <file ID="file:2" ADMID="admin:2" CREATED="yyyy-mm-dd"
      MIMETYPE="application/x-type" OWNERID="id" SIZE="n">
      <Flocat xlink:type="simple" xlink:href="dataset.dat"/>
    </file>

    ...
  </fileGrp>
</fileSec>

<structMap type="LOGICAL">
  <div TYPE="EJARITEM" ADMID="admin:item" DMD="descr:item"
    LABEL="item bibliographic citation">
    <fptr FILEID="file:item-content"/>
    <fptr FILEID="file:item-links"/>
    <fptr FILEID="file:1"/>
    <fptr FILEID="file:2"/>
    ...
  </div>
</structMap>
</mets>

```

Fig. A.3(b) SIP item-level metadata file (item-md.xml)  
(continued from previous page).



## B Normative Data Formats

The normative data formats are those preferred for representing content components within the SIP.

### B.1 Metadata, Issue and Item-Level Text, and Item-Level Linkage

Format	XML
Version	1.0
File Extension	.xml
MIME type	text/xml
Authoritative Specification	Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, and Eve Maler, eds., <i>Extensible Markup Language (XML) 1.0 (Second Edition)</i> , W3C Recommendation, October 6, 2000 < <a href="http://www.w3.org/TR/2000/REC-xml-20001006">http://www.w3.org/TR/2000/REC-xml-20001006</a> >.
Additional Requirements	All Unicode characters [UNICODE] outside of the US-ASCII repertoire [USASCII] must be represented in the marked-up text as character references:  <code>&amp;#xhhhh;</code>  where <i>hhhh</i> is the hexadecimal representation of the character's Unicode code point. <sup>1</sup>  All characters outside of the Unicode repertoire must be represented in the marked-up text as named entities drawn from the archive common entity set.

Table B.1 Normative metadata and issue and item-level text format (XML).

Beyond conformance to the XML 1.0 standard, certain SIP components must also conform to specific XML schemas.

#### B.1.1 Structural Metadata schema

Schema	METS
Version	1.0
Namespace	<a href="http://www.loc.gov/METS/">http://www.loc.gov/METS/</a>
Location	<a href="http://www.loc.gov/standards/mets/mets.xsd">http://www.loc.gov/standards/mets/mets.xsd</a>
Informative Specification	<i>Metadata Encoding &amp; Transmission Standard (METS)</i> , Washington, DC: Network Development and MARC Standards Office, Library of Congress < <a href="http://www.loc.gov/standards/mets/">http://www.loc.gov/standards/mets/</a> >.

Table B.2 Metadata schema.

#### B.1.2 Descriptive and Administrative Metadata schema

Schema	EJAR
Version	1.0
Namespace	<a href="http://hul.harvard.edu/EJAR/METADATA">http://hul.harvard.edu/EJAR/METADATA</a>
Location	<a href="http://hul.harvard.edu/EJAR/metadata.xsd">http://hul.harvard.edu/EJAR/metadata.xsd</a>
Informative Specification	<i>TBD</i>

Table B.3 Metadata schema.

<sup>1</sup> ISO/IEC 10646 defines a Universal Character Set (UCS) with a 31 bit encoding. The Unicode Standard [UNICODE] uses a 16 bit encoding. As of 1993 the Unicode repertoire is fully equivalent to the 10646-1 Basic Multilingual Plane (BMP) subset.

### B.1.3 Issue-level text schema

Schema	EJAR-ISSUE
Version	1.0
Namespace	<a href="http://hul.harvard.edu/EJAR/ISSUE/">http://hul.harvard.edu/EJAR/ISSUE/</a>
Location	<a href="http://hul.harvard.edu/EJAR/issue.xsd">http://hul.harvard.edu/EJAR/issue.xsd</a>
Informative Specification	<i>TBD</i>

Table B.4 Issue-level text schema.

### B.1.4 Item-level text schema

Schema	EJAR-ITEM
Version	1.0
Namespace	<a href="http://hul.harvard.edu/EJAR/ITEM/">http://hul.harvard.edu/EJAR/ITEM/</a>
Location	<a href="http://hul.harvard.edu/EJAR/item.xsd">http://hul.harvard.edu/EJAR/item.xsd</a>
Informative Specification	<i>TBD</i>

Table B.5 Item-level text schema.

### B.1.5 Item-level mathematics DTD

DTD	MathML
Version	2.0
DTD FPI	<code>-//W3C//DTD MathML 2.0//EN</code>
Location	<a href="http://www.w3.org/TR/MathML2/dtd/mathml2.dtd">http://www.w3.org/TR/MathML2/dtd/mathml2.dtd</a>
Informative Specification	David Carlisle, Patrick Ion, Robert Miner, and Nico Poppelier, eds., <i>Mathematical Markup Language (MathML) Version 2.0</i> , W3C Recommendation 21 February 2001 < <a href="http://www.w3.org/TR/MathML2/">http://www.w3.org/TR/MathML2/</a> >.

Table B.6 Item-level mathematics DTD.

### B.1.6 Item reference links schema

Schema	EJAR-LINKS
Version	1.0
Namespace	<a href="http://hul.harvard.edu/EJAR/LINKS/">http://hul.harvard.edu/EJAR/LINKS/</a>
Location	<a href="http://hul.harvard.edu/EJAR/links.xsd">http://hul.harvard.edu/EJAR/links.xsd</a>
Informative Specification	<i>TBD</i>

Table B.7 Item reference links schema.

## B.2 Raster Still Image

Format	TIFF
Version	6.0
File Extension	.tif
MIME type	image/tiff
Authoritative Specification	Adobe Systems, Inc., <i>TIFF Revision 6.0</i> , final — June 3, 1992 < <a href="http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf">http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf</a> >.
Additional Requirements	All TIFF files must contain a single image. All TIFF files must conform fully to all TIFF 6.0 Baseline requirements. RGB full color images must use the <i>Chunky</i> PlanarConfiguration format. Bilevel (1 bit) images must be compressed with ITU T.6 encoding (CCITT Facsimile Group IV) [ITU]. Grayscale, palette-color, and RGB full color images must be uncompressed.

Table B.7 Normative raster still image format (TIFF).

## B.3 Vector Still Image

Format	SVG
Version	1.0
File Extension	.svg
MIME type	image/svg+xml
Authoritative Specification	Jon Ferraiolo, ed., <i>Scalable Vector Graphics (SVG) 1.0 Specification</i> , W3C Proposed Recommendation, July 19, 2001 < <a href="http://www.w3.org/TR/2001/PR-SVG-20010719/">http://www.w3.org/TR/2001/PR-SVG-20010719/</a> >.

Table B.8 Normative vector still image format (SVG).

## B.4 Page Description

Format	PDF
Version	1.4
File Extension	.pdf
MIME type	application/pdf
Authoritative Specification	Adobe Systems, Inc., <i>Adobe Portable Document Format, Version 1.3</i> (2nd ed; Reading, MA: Addison-Wesley, 2000) < <a href="http://partners.adobe.com/asn/developer/acrosdk/docs/PDFRef.pdf">http://partners.adobe.com/asn/developer/acrosdk/docs/PDFRef.pdf</a> >.  _____, <i>Errata for PDF Reference</i> , 2nd ed., April 10, 2001 < <a href="http://partners.adobe.com/asn/developer/acrosdk/docs/PDFerrata.txt">http://partners.adobe.com/asn/developer/acrosdk/docs/PDFerrata.txt</a> >.  _____, <i>Portable Document Format: Changes from Version 1.3 to 1.4</i> , Technical Note #5409 (preliminary), June 11, 2001 < <a href="http://partners.adobe.com/asn/developer/acrosdk/docs/filefmtspecs/PDF14Deltas.pdf">http://partners.adobe.com/asn/developer/acrosdk/docs/filefmtspecs/PDF14Deltas.pdf</a> >.
Additional Requirements	All PDF files must be unencrypted, must have neither an owner or user password defined, and must be linearized.

Table B.9 Normative page description format (PDF).

## B.5 Audio

Format	AIFF
Version	1.3
File Extension	.aif
MIME type	audio/x-aiff
Authoritative Specification	Apple Computer, Inc., <i>Audio Interchange File Format: A Standard for Sampled Sound Files</i> , Version 1.3, January 4, 1989.

Table B.10 Audio format (AIFF).

## B.6 Video

Format	MPEG
Version	2
File Extension	.mpg
MIME type	application/pdf
Authoritative Specification	ISO/IEC 13818:2000, <i>Information technology -- Generic coding of moving pictures and associated audio information</i> . Geneva: ISO, 2000.

Table B.11 Video format (MPEG).

## C Non-Normative Formats

The following non-normative formats will be accepted by the archive in a SIP, but they will be transformed into their normative analogues during the ingest process.

<b>Component Function</b>	<b>Non-Normative Format</b>	<b>Normative Analogue</b>
Raster Still Image	GIF	TIFF 6.0
	JPEG (JFIF)	
Vector Still Image	Encapsulated PostScript *	XML 1.0 (SVG 1.0)
	PostScript *	
Page Description	Encapsulated PostScript	PDF 1.4
	PostScript	

Table C.1 Mappings between non-normative and normative formats.

\* Note that PostScript and Encapsulated PostScript will be transformed to SVG only if the instance files are composed entirely of vector-oriented images. The inclusion of any raster data will result in the files being transformed to PDF.

### C.1 Raster Still Image

All SIP components received in the following non-normative raster still image formats will be transformed upon ingest into TIFF 6.0.

#### C.1.1 GIF

Format	GIF
File Extension	.gif
MIME type	image/gif
Authoritative Specification	CompuServe, <i>GIF Graphics Interchange Format: A Standard Defining a Mechanism for the Storage and Transmission of Raster-based Graphics Information</i> , June 15, 1987. (Version 87a.)  _____, <i>Graphics Interchange Format, Version 89a</i> , July 31, 1990.
Normative Analogue	TIFF 6.0

Table C.2 Non-normative raster still image format (GIF).

#### C.1.1 JPEG / JFIF

Technically, the JPEG format defines an image compression standard, not a file exchange format. The JPEG File Interchange Format (JFIF) was developed subsequently to allow file interchange of JPEG-encoded data.

Format	JFIF
File Extension	.jpg
MIME type	image/jpeg
Authoritative Specification	Independent JPEG Group, <i>JPEG File Interchange Format</i> , Version 1.02, ed. Eric Hamilton, September 1, 1992 < <a href="http://www.w3.org/Graphics/JPEG/jfif3.pdf">http://www.w3.org/Graphics/JPEG/jfif3.pdf</a> >.  <i>Information Technology — Digital Compression and Coding of Continuous-tone Still Images — Requirements and Guidelines</i> , ITU-T Recommendation T.81, September 1992. Geneva: ITU/CCITT, 1993 < <a href="http://www.w3.org/Graphics/JPEG/itu-t81.pdf">http://www.w3.org/Graphics/JPEG/itu-t81.pdf</a> >.
Normative Analogue	TIFF 6.0

Table C.3 Non-normative raster still image format (JPEG/JFIF).

## C.2 Page Description

All SIP components received in the following non-normative page description formats will be transformed upon ingest into SVG 1.0, if entirely vector, otherwise PDF 1.4.

### C.2.1 Encapsulated PostScript

Encapsulated PostScript (EPS) is a format used to encapsulate a single PostScript page image. An EPS file is a valid PostScript file, subject to additional formatting rules and operator restrictions.

Format	EPS
File Extension	.eps
MIME type	application/postscript
Authoritative Specification	(See specifications for PostScript below.) Adobe System, Inc., <i>Encapsulated PostScript (EPS) File Format Specification Version 3.0</i> , Technical Note #5002, May 1, 1992 < <a href="http://partners.adobe.com/asn/developer/pdfs/tn/5002.EPSF_Spec.pdf">http://partners.adobe.com/asn/developer/pdfs/tn/5002.EPSF_Spec.pdf</a> >.
Normative Analogue	XML 1.0 (SVG 1.0) <i>if entirely vector</i> PDF 1.4 <i>if any portion is raster</i>

Table C.4 Non-normative page description format (EPS).

### C.2.2 PostScript

Format	PostScript
File Extension	.ps
MIME type	application/postscript
Authoritative Specification	Adobe Systems, Inc., <i>PostScript Language Reference Manual</i> . (3rd ed.; Reading, MA: Addison-Wesley, 1999) < <a href="http://partners.adobe.com/asn/developer/pdfs/tn/PLRM.pdf">http://partners.adobe.com/asn/developer/pdfs/tn/PLRM.pdf</a> >.  _____, <i>Errata for PostScript Language</i> , 3rd ed., December 8, 2000 < <a href="http://partners.adobe.com/asn/developer/pdfs/tn/PSerrata.txt">http://partners.adobe.com/asn/developer/pdfs/tn/PSerrata.txt</a> >.
Normative Analogue	XML 1.0 (SVG 1.0) <i>if entirely vector</i> PDF 1.4 <i>if any portion is raster</i>

Table C.5 Non-normative page description format (PostScript).

## References

- [DOI] *The DOI Handbook*, Version 1.0.0, International DOI Foundation, February 2001 <[http://www.doi.org/handbook\\_2000/010222DOI-Handbook-V100.pdf](http://www.doi.org/handbook_2000/010222DOI-Handbook-V100.pdf)>.
- [FONT] ISO/IEC 9541-2:1991, *Information Technology — Font Information Interchange — Part 2: Interchange Format*. Geneva: ISO, 1991.
- [FTP] J. Postel and J. Reynolds. *File Transfer Protocol (FTP)*, IETF STD 9, RFC 959, October 1985 <<http://www.ietf.org/rfc/rfc959.txt>>.
- [ISSN] ISO 3297:1998, *Information and Documentation — International Standard Serial Number (ISSN)*. Geneva: ISO, 1998.
- [ITU] *Facsimile Coding Schemes and Coding Control Functions for Group 4 Facsimile Apparatus*, ITU-T Recommendation T.6, November 1988 <<http://www.itu.int/rec/recommendation.asp?type=items&lang=e&parent=T-REC-T.6-198811-I>>.
- [MATHML] David Carlisle, Patrick Ion, Robert Miner, and Nico Poppelier, eds. *Mathematical Markup Language (MathML) Version 2.0*, W3C Recommendation 21 February 2001 <<http://www.w3.org/TR/MathML2/>>.
- [MD5] R. Rivest. *The MD5 Message-Digest Algorithm*, IETF RFC 1321, April 1992 <<http://www.ietf.org/rfc/rfc1321.txt>>.
- [METS] *Metadata Encoding & Transmission Standard (METS)*, Washington, DC: Network Development and MARC Standards Office, Library of Congress, 2001 <<http://www.loc.gov/standards/mets/>>.
- [NISOIMG] Robin Dale and Oya Y. Rieger, eds. *Data Dictionary: Technical Metadata for Digital Still Images*, Working Draft, 1.0, July 5, 2000. Bethesda, MD: NISO, 2000 <<http://www.niso.org/pdfs/DataDict.pdf>>.
- [OAIS] *Reference Model for an Open Archival Information System (OAIS)*, CCSDS 650.0-R-1.1 (Red Book). Oxford, UK: Consultative Committee for Space Data Systems, April 20, 2001 <[http://ssdoo.gsfc.nasa.gov/nost/isoas/us20/650x0\\_010510.pdf](http://ssdoo.gsfc.nasa.gov/nost/isoas/us20/650x0_010510.pdf)>.
- [SICI] NISO/ANSI Z39.56-1996, *Serial Item and Contribution Identifier (SICI)*, Version 2, August 14, 1996. Bethesda, MD: NISO Press, 1997 <<http://www.techstreet.com/cgi-bin/pdf/free/152629/z39-56.pdf>>.
- [SMIL] Jeff Ayars, Dick Bulterman, Aaron Cohen, Ken Day, Erik Hodge, Philipp Hoschka, Eric Hyché, Muriel Jourdan, Michelle Kim, Kenichi Kubota, Rob Lanphier, Nabil Layaïda, Thierry Michel, Debbie Newman, Jacco van Ossenbruggen, Lloyd Rutledge, Bridie Saccocio, Patrick Schmitz, and Warner ten Kate, eds. *Synchronized Multimedia Integration Language (SMIL 2.0)*, W3C Recommendation 07 August 2001 <<http://www.w3.org/TR/smil20/>>.
- [TEI] C. M. Sperberg-McQueen and Lou Burnard, eds. *Guidelines for Electronic Text Encoding and Interchange*, Rev. reprint, May 1999 <<http://www.tei-c.org/Guidelines/index.htm>>.
- [TYPE1] Adobe Systems, Inc. *Adobe Type 1 Font Format*. Reading, MA: Addison-Wesley, 1990 <[http://partners.adobe.com/asn/developer/pdfs/tn/T1\\_SPEC.PDF](http://partners.adobe.com/asn/developer/pdfs/tn/T1_SPEC.PDF)>. See also *Type 1 Font Format Supplement*, Technical Specification #5015, May 15, 1994 <[http://partners.adobe.com/asn/developer/pdfs/tn/5015.Type1\\_Supp.pdf](http://partners.adobe.com/asn/developer/pdfs/tn/5015.Type1_Supp.pdf)>.
- [UNICODE] *The Unicode Standard, Version 2.0*. Reading, MA: Addison-Wesley, 1996 <<http://www.unicode.org/unicode/uni2book/u2.html>>.

- [URI] T. Berners-Lee, R. Fielding, and L. Masinter. *Uniform Resource Identifiers (URI): Generic Syntax*, IETF RFC 2396, August 1998 <<http://www.ietf.org/rfc/rfc2396.txt>>.
- [UCS] ISO/IEC 10646-1:2000, *Information Technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane*. Geneva: ISO, 2000.
- [USASCII] ANSI X3.4-1986, *Information Systems — Coded Character Sets — 7-Bit American National Standard Code for Information Interchange (7-Bit ASCII)*. New York: ANSI, 1997.
- [XML] Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, and Eve Maler, eds. *Extensible Markup Language (XML) 1.0 (Second Edition)*, W3C Recommendation, October 6, 2000 <<http://www.w3.org/TR/2000/REC-xml-20001006>>.