ResCarta Standards

RC2007.bk.1.000.8.1

The ResCarta Foundation, Inc.

ResCarta Source Format Type:

Book

Monographs/Serials

DRAFT Version 1.0

ResCarta Foundation, Inc.
La Crosse, Wisconsin
www.ResCarta.org

November 1, 2008

THE RESCARTA STANDARDS

ResCarta Foundation, Inc. was founded to encourage the development of a single set of open community standards and open source implementations of those standards. The objective is to produce a subset of existing, far-reaching standards and implementations that guide the creation, storage, indexing and retrieval of digital collections.

In creating these standards, the format strives to maintain simplicity that can and should stop creators and organizers of digital material from overloading collections with too many unnecessary features. The standards are designed to be the minimum set of rules to which all digital collections should adhere so that true interoperability may be achieved.

The ResCarta Standards currently support the following source types:

- Published works known as books. (e.g. monographs, serials, serial monographs)
- Published or unpublished documents (e.g. loose papers, letters, correspondence)
- Conference Proceedings
- Microforms
- Photographs
- Negatives
- Binary output computer files (e.g. MsWord Documents, Lotus 123 files, Symphony)
- Audio recordings
- Video recordings with and without audio
- Websites
- E-books
- Emails

In proceeding toward the stated objective, ResCarta Foundation has begun the task of creating standards for the various ResCarta-supported source types. ResCarta Foundation makes available draft standards for review and comment. This ensures that the standards meet the best practices of our industry. We value your thoughts on our standards and rely on your comments to assist us in completing this important task. Please provide comments on any of the ResCarta Standards by contacting us at info@rescarta.org.

TABLE OF CONTENTS

1.0 Definitions 5

2.0 General WorkFlow 10

2.1 Basic Workflow explanations 10

2.2 Workflow Source Format issues: Books 10

3.0 Specifications 11

3.1 Object Creation 11

3.2 Directory Structures 11

3.3 File Naming 12

3.4 ResCarta Object Metadata for books 13

3.5 ResCarta Container Structural Metadata for Books 14

3.6 Example of Monograph Described to ResCarta Standards 15

1.0 DEFINITIONS

Base Collection Index

Binary output files

Proprietary file formats used to store output from specific programs (e.g. Microsoft Word, Lotus 123, Adobe Acrobat, Word Perfect)

Book

A bound ResCarta container that has page images.

Checksum

A method of providing information for error detection, usually calculated by summing a set of values.

Convert

To change from one system or format to another.

Container

That which defines a group of objects (e.g. a book defines a group of pages)

Container Structural Metadata File

Metadata that is located at the container level.

Descriptive Metadata

Information on the intellectual content of an item, analogous to the main part of a traditional library catalogue record.

Digital Identifier System®

A system that provides a mechanism to interoperably identify and exchange intellectual property in the digital environment. It is an identifier. DOI conforms to a URI specification. It provides an extensible framework for managing intellectual content based on proven standards of digital object architecture and intellectual property management.

Documents

Papers, letters, correspondence or other paper-based source material, typically either not bound or loosely bound. Documents may be structured (organized) or unstructured (disorganized).

Dublin Core

A basic set of 15 metadata elements designed to represent core fields for the description of any electronic resource.

Dumb number

An identifier that does not include any codes or other signifying elements. Its purpose is to identify, not to provide any type of descriptive information about the thing being named.

Handle System

A comprehensive system for assigning, managing, and resolving persistent identifiers, known as "handles," for digital objects and other resources on the Internet. Handles can be used as Uniform Resource Names (URNs).

Institutional identifier (ResCarta)

An alpha-numeric eight character word used to define a discrete institution creating ResCarta[tm] data.

Limiter

A ResCarta protection device to insure compatibility and reliability in storing digital objects.

Limiter/aggregator directory

A part of the directory structure whose sole purpose is to limit the number of other subdirectories in the parent directory.

Metadata

Data about other data, commonly divided into descriptive metadata such as bibliographic information, structural metadata about formats and structures, and administrative metadata, which is used to manage information.

METS

Metadata Encoding and Transmission Standard "The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation" (from www.loc.gov/standards/mets).

MODS

Metadata Object Description Schema is an XML schema "for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications (from www.loc.gov/standards/mets).

Monograph

A book that is complete in one physical piece, as opposed to a serial (or journal), which is produced in cumulative parts indefinitely

Object Metadata

Data about the particular ResCarta object.

Property

A property is a specific aspect, characteristic, attribute, or relation used to describe a resource.

Record

A record is some structured metadata about a resource, comprising one or more properties and their associated values.

ResCarta Container

Analog or digital source of ResCarta objects.

ResCarta Collection

A group of ResCarta containers of same or different ResCarta source format types that are contained by upper level metadata.

ResCarta Conversion Format

A named ResCarta digital format. (e. g. Book, Newspaper, Photograph, etc.)

ResCarta Object

The finest granularity of a particular ResCarta source format type.

ResCarta Source Format Type

One of the enumerated analog or digital categories used to separate materials into the proper ResCarta conversion format.

ResCarta System

A collection of ResCarta[tm] data residing in an operating system dependent file structure.

Resource

A resource is anything that has an identity. Familiar examples include an electronic document, an image, and a collection of other resources. Not all resources are network "retrievable." For example, human beings, corporations, and bound books in a library are considered resources.

Root identifier

An alpha-numeric eight character word used to name the digital object's containing directory. It must be unique within the aggregator which contains it.

Serial

A publication in any medium issued in successive parts bearing numerical or chronological designations. Intended to be continued indefinitely. Included are periodicals, newspapers, annuals, proceedings, transactions, monographic series, etc.

Serial monograph

A series of somewhat related monographs published in a limited time frame.

Structural metadata

Information about how the item is put together or arranged such as the table of contents page, individual page numbers, illustration and plates pages, et cetera. Information that describes the structure of an item, such as a book, so that all of the pages of that item can be displayed in the correct order. Structural metadata may also include information that supports navigation among the components of a complex object.

TAGS

A reference to the identifying number of a TIFF field.

XML

eXtensible Markup Language

ZLIB compression.

A compression library initially developed as part of the PNG graphics standard that is not protected by patents.

2.0 GENERAL WORKFLOW

1) Determine Source Format Type
2) Convert/Import
3) Extract/Create Metadata
4) Map to the Directory Structure and name files
5) Assign to a Base Collection Index
6) Assign a DOI

2.1 BASIC WORKFLOW EXPLANATIONS

2.1.1.Assigning to a Base Collection Index: Typically the item is assigned to the primary supplier of the ResCarta container: the institution that is converting the material or having the material converted. The registered institution identifier will become the top-level directory separator. The intermediate aggregator/limiter is also an original collection identifier. The resulting ResCarta container may be moved from its original collection but it never changes its original identifier. This reference may be considered a dumb number by other institutions but used originally as a collector. The collection level metadata is the top level authority collector.

2.1.2 Assigning a DOI: A unique identifier is assigned to each container by concatenating the institution key, the intermediate aggregator/limiter and the root identifier. This identifier can be registered with DOI.org registrars to ensure permanence of the key. The use of DOI as opposed to creating a new system based on handle reduces the complexity of handle use and standardizes the approach. Institutions may register directly with a DOI registry or use the registry of the ResCarta Foundation.

2.2 WORKFLOW SOURCE FORMAT ISSUES: BOOKS

When gathering analog and digital books into the ResCarta system, several source format specific issues arise

The book must be cataloged as a monograph or a serial before proceeding too far into the conversion since the kinds of object metadata used by monographs and serials vary. The structural metadata of serials and monographs also varies and requires additional fields for serial publications or serialized monographs than for monographs.

The directory structures for the ResCarta system specify a single directory structure to contain monographs, serials and serial monographs. The introduction of the additional metadata driven directory levels for serials causes a disjuncture of the ResCarta concepts of container and object. Containers hold discrete objects. ResCarta defines a book as a container of page images. The page image is the finest granularity that is recognized for this type of source format. Therefore the directory structure for the ResCarta book type also has two main parts, the container root identifier and the images location.

3.0 SPECIFICATIONS

3.1 OBJECT CREATION

Each page of a book is to be converted to a ResCarta object. For book containers there are three file types available for use as ResCarta objects.

1. Bitonal, 600DPI, G4 compressed TIFF Files
2. Gray, 300DPI, 8bit, uncompressed TIFF Files
3. Color, 300DPI, 24bit, uncompressed TIFF Files

Pages comprised of mainly of text, half-toned photographs and etchings are to be scanned bitonally at 600dpi. The resulting image files are stored as G4 compressed TIFFs.

Pages comprised of mainly continuous tone black and white images are to be scanned as 8bit grayscale, at 300dpi. The resulting image files are to be stored as uncompressed TIFFs.

Pages having spot color or continuous tone color images are to be scanned as 8bits per channel RGB Color, at 300dpi. The resulting image files are to be stored as uncompressed TIFFs.

3.2 DIRECTORY STRUCTURES

The First-level directory name is always eight characters in length and named from the ResCarta institutional identifier of the original converter of the data.

The Second level directory is always eight characters in length and named from a sequential dumb number generated at the time of production. This is a ResCarta limiter designed to keep the number of container level directories from exceeding operating system limits. It can contain up to a maximum of 512 container level directories and if ISO-9660 should not exceed 620 megabytes in total storage capacity. All levels restrict names to upper case letters, digits and underscores ("_").

The third level directory, the ResCarta container level, is always eight characters in length and is given an institutionally unique identifying name also known as the root identifier.

There is no need for a limiter/aggregator directory when dealing with published works since the format of the modern book is normally limited to a handful of binding widths. This will limit the container to fewer than two thousand ResCarta objects in real world applications.

3.3 FILE NAMING

The first four characters contain a leading zero padded, sequentially incremented, image sequence number, starting with 0001. The final four characters will contain a representation of the page number printed on the page, formulated according to the following rules.

A logical page number or appropriate tag to accompany it is designated for every image. This page number or tag will be used in the container metadata and amended, if necessary, to accommodate ISO 9660 file name restrictions in the image file name. Page numbers containing characters that are not permitted in ISO 9660 file names are recorded with an underscore character in place of the illegal character. For example, page 22.6 is recorded as 22_6. In situations such as this, the unmodified page number should be recorded in the container metadata as described in the section on the container level metadata.

The word "pagination" refers to the logical sequential pagination of a series of pages. For example, a page without a printed number on it, which is located between pages imprinted 2 and 4 can be assumed to be page 3. Similarly, four pages without page numbers printed on them followed by pages 5, 6, 7, etc., can be assumed to be pages 1, 2, 3 and 4.

All pages included within the logical pagination are designated with their actual page numbers.

If there are pages before the logical pagination begins, they are designated 000A, 000B, 000C, 000D, 000E, etc. All levels restrict names to upper case letters, digits and underscores ("_").

Pagination that appears as Roman numerals in the source material are translated into Arabic numerals and appended with a leading "R" for file names (e.g.: page vii becomes page R007). In the absence of printed page numbers, it is to be assumed that Roman numerals continue until logical Arabic numeral pagination begins. In the situation where sequential pagination continues through a change from Roman numerals to Arabic numerals, the Arabic numerals are assumed to start at the change in type of document content. Typically this occurs in transitions from the front matter to the content section.

When there are pages in the material that are not included in the sequential pagination (commonly occurring with plates), the pages are designated by the number of the preceding paginated page appended with a trailing letter This designation will increase sequentially for each page. (e.g. 0031, 0032, 032A, 032B, 0033, 0034)

Page numbers that actually contain letter prefixes will be recorded according to the same rules as standard Arabic numbered pages, except that punctuation between the prefix and the number should be dropped. Thus a page from Appendix A labeled as A-9, is recorded as 00A9.

Adornments around page numbers, including dashes, asterisks, parentheses, brackets, etc., are ignored and not entered.

Page numbers that result in more than four characters will be truncated to the last four characters. For example, pages 99.1, 99.2, 100.1, 100.2 are recorded as 99_1, 99_2, 00_1, and 00_2. The first quartet containing the series will keep the names unique and in order.

3.4 RESCARTA OBJECT METADATA FOR BOOKS

Container metadata is stored in the header of each tiff file. This ensures that at the finest granularity (i.e, the ResCarta object), the ResCarta system will carry sufficient information to provide provenance and a base to rebuild the collection metadata.

Container level metadata carries basic (i.e., Dublin Core) information for the source material as well as basic technical metadata on the conversion. This metadata is expressed in Metadata Object Description Schema (MODS) and placed within the tiff header of the object in the Document Description TAG = 270(10E).

3.4.1 Minimum data fields required under the ResCarta standards for books:

Title
Label: titleinfo
Definition: The name given to the resource, usually by the creator or publisher.

Creator
Label: name
Definition The person, persons or organization primarily responsible for creating the intellectual content of the resource. For example, authors in the case of written documents.

Publisher
Label: origininfo/publisher
Definition The entity responsible for making the resource available in its original analog form, such as a publishing house, a university department, or a corporate entity.

Label: origininfo/publisher
Definition The entity responsible for making the resource available in its present digital form, such as a publishing house, a university department, or a corporate entity.

Date
Label: origininfo/dateIssued
Definition The date of publication of the original analog source

Label: origininfo/dateCaptured
Definition The date of capturing the original analog source to digital format.

Resource Type
Label: typeOfResource
Definition The category of the resource, such as home page, novel, poem, working paper, technical report, essay, dictionary. For the sake of interoperability, Type should be selected from the ResCarta enumerated list.

Format
Label: physicalDescription
Definition The data format and, optionally, dimensions of the resource. The format is used to identify the software and hardware needed to display or operate the resource. For the sake of interoperability, the format should be selected from the ResCarta enumerated list.

Resource Identifier
Label: identifier
Definition A string used to uniquely identify the resource. The concatenation of the institutional identifier, the disk limiter, and the root identifier

Language
Label: language
Definition The language(s) of the intellectual content of the resource selected from the ResCarta enumerated list.

Rights Management
Label: note/ownership
Definition A rights management statement, an identifier that links to a rights management statement, or an identifier that links to a service providing information about rights management for the resource

Additional technical metadata concerning the size, resolution, color-depth, compression, ResCarta software versioning and Date/Time of creation are stored in the TIFF tags as specified by the TIFF 6.0 specification. Specifically the following eighteen tag numbers will be populated, 254-259,262, 266, 273, 274, 277-279, 282-284, 296, 305, and 306. Images will be represented in orientation 1 and this will be reflected in tag 274.

ResCarta optional object metadata includes full, unedited output from OCR engines. The output is reduced to word found and found word location. These items are placed into the private TAGS registered for this purpose.

Raw OCR Output
Label: Content
Direct textual output obtained from passing the image through an optical character recognition program and stored in tiff Tag = 5688(1638) resulting in one line per word expressed in the character set appropriate for the language found in the collection level metadata. If no language is specified, it defaults to US English.

Word Location Data
Label: WordLoc
The x,y location of each word found on the image by the optical character recognition program, along with its perceived font and font size stored in tiff tag = 5689(1639) resulting in one line per word. There will be the same number of lines in this data as in the Raw OCR Output data. ASCII encoded and ZLIB compressed.

3.5 RESCARTA CONTAINER STRUCTURAL METADATA FOR BOOKS

ResCarta container structural metadata has been adapted to the METS schema. The ResCarta standard for books need only complete a portion of the metadata described by the METS schema to reduce the pressure on the creating or converting institution. By using the METS schema, ResCarta allows for future growth and the current deep needs of more advanced conversions.

Beyond the base metadata used primarily as core data stock for discovery, there exist structural differences between containers in the book format. At the top level, structural differences between

a monograph and a serial publication must be recognized. These structural differences are described in other systems by divergent directory structures or by an all inclusive directory structure with the ability to not fill unnecessary items.

In an effort to reduce the "case statements" necessary in programming for search, display, and printing, ResCarta standards keep the directory structure mirrored to the container model. All arbitrary non-physical structures placed upon the object by the creator are recreated in the appropriate metadata level. The modeling of the physical publication of issues as separate items and collections of volumes as separate items becomes the basis to remove the 'unused' directory structures.

If the container is "Journal of Political Statements Vol 1," then the container is the volume and the pages of that volume are the objects. The issue level is then a non-physical structure, which is defined by metadata. Individual articles in the issue are defined by additional metadata. All logical document identification is carried in the container structural metadata file.

Thus, if another institution gets a single analog copy of the "Journal of Political Statements Vol 1 Issue 13" to digitized, that container's pages become the objects and the relation to the previous ResCarta container is found in the metadata. These two containers will have different top-level directory names, different limiters and different root identifiers. They will also be issued two distinct DOIs. This structure will not allow the data to collide if loaded to a server, nor will such a load supercede any of the metadata.

The METADATA.XML file located under the root identifier carries all the container structural metadata. Following METS, the objects are defined as participating in a chapter of a monograph, an article of a particular issue of a particular volume of a serial publication, or any number of artificially generated subsets produced by the original author or publisher.

This file also carries redundant metadata from the minimum data fields required by ResCarta already stored in the header of each object. This ensures the ability to recreate the metadata in a single object if the metadata is lost. Additional container metadata can be added at this point by ResCarta production methods.

3.6 EXAMPLE OF A SERIAL DESCRIBED TO RESCARTA STANDARDS

Hannibal Free Public Library (HFS) owns a collection of city directories dating from the 1800s and wishes to digitize these books to ResCarta format. Included in the collection are *Hannibal City Directories* published as a serial, and a monograph *The History of Hannibal MO*.

Hannibal Free is registered with the ResCarta Foundation as an institution and has received the identifier of "hannibal." Additionally, HFS uses a library cataloging system that gives a unique identifier to each item. HFS has chosen not to use this identifier as its root identifier due to irregularities in the identifiers.

The source materials are comprised of text and etchings so they are scanned bitonally at 600dpi. The resulting image files are stored as G4 compressed TIFFs.

Since this is the first conversion of the materials, the following directory structure is produced for the serial *Hannibal City Directory 1859 - 1860*.

Table 1. Directory Structure for the Serial
Hannibal City Directory 1859 - 1860

\\hannibal\hf000001\00000001

hannibal is the institutional identifier,
hf000001 is the disk limiter, and
00000001 is the unique root identifier.

Files are named with a dual quartet, the first quartet being the sequence of the image in the container and the second quarter being the last four characters of the as-printed page name. File naming follows the procedure discussion section 3.2 of this document and looks like this:

Table 2. File Naming for the Serial
Hannibal City Directory 1859 - 1860

0001000A.tif	0007R005.tif	0013R011.tif
0002000B.tif	0008R006.tif	0014R012.tif
0003R001.tif	0009R007.tif	00150001.tif
0004R002.tif	0010R008.tif	00160002.tif
0005R003.tif	0011R009.tif	…
0006R004.tif	0012R010.tif	02280214.tif

ResCarta Object metadata is written as a MODS file into the headers of each TIFF file in the ResCarta Metadata Tag = 5687(1637) as follows:

Table 3. ResCarta Object Metadata for the Serial
Hannibal City Directory 1859 - 1860

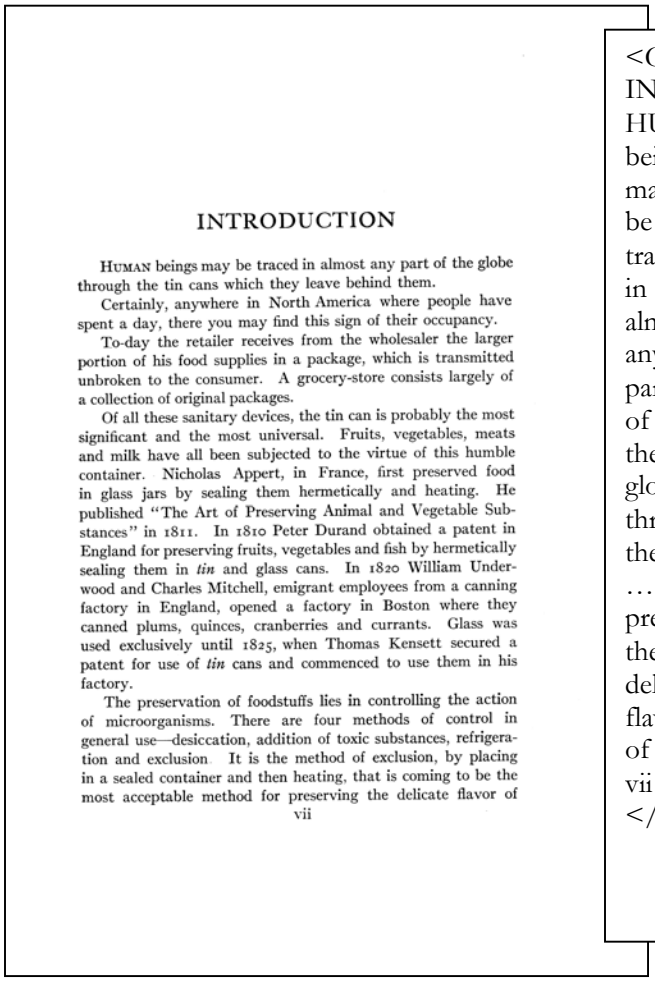
```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<mods version="3.0" xsi:schemaLocation="http://www.loc.gov/mods/v3
http://loc.gov/standards/mods/v3/mods-3-0.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://www.loc.gov/mods/v3">
    <relatedItem>
        <titleInfo>
            <title>Hannibal City Directory 1859 - 1860</title>
        </titleInfo>
        <name authority="local" type="corporate">
            <namePart>Hannibal Public Library</namePart>
            <role>
                <roleTerm authority="marcrelator"
type="text">Owner</roleTerm>
                <roleTerm authority="marcrelator"
type="code">own</roleTerm>
            </role>
```

```
        </name>
        <typeOfResource>text</typeOfResource>
        <originInfo>
            <dateIssued encoding="iso8601">1859</dateIssued>
            <dateCaptured encoding="iso8601">2005-06-21</dateCaptured>
            <issuance>continuing</issuance>
        </originInfo>
        <language>
            <languageTerm authority="iso639-2b"
type="code">eng</languageTerm>
        </language>
        <physicalDescription>
            <form authority="marcform">electronic</form>
            <extent>5.938 x 9.313 in.</extent>
            <digitalOrigin>reformatted digital</digitalOrigin>
        </physicalDescription>
        <subject>
            <hierarchicalGeographic>
                <country>US</country>
                <state>MO</state>
                <county>Marion, Ralls</county>
            </hierarchicalGeographic>
        </subject>
        <subject authority="lcsh">
            <topic>Hannibal (Mo.)</topic>
        </subject>
        <subject authority="lcsh">
            <topic>Directories.</topic>
        </subject>
        <subject authority="lcsh">
            <topic>History.</topic>
        </subject>
        <location>
            <url>hannibal/hf000004/00000001</url>
        </location>
        <identifier type="local">hannibal/hf000004/00000001
</identifier>
        <part>
            <detail level="1" type="volume">
                <number>1859</number>
            </detail>
            <detail level="2" type="issue">
                <number>000</number>
            </detail>
            <extent unit="page">
                <start>0002</start>
                <end>0002</end>
            </extent>
        </part>
    </relatedItem>
    <location>
        <url>hannibal/hf000004/00000001/00010002.TIF</url>
    </location>
    <identifier
type="local">hannibal/hf000004/00000001/00010002</identifier>
</mods></name>
```

```
<typeOfResource>text</typeOfResource>
<genre authority="marc">bibliography</genre>
<originInfo>
<place>
<code authority="marc">manhi</code>
<text>Andover, MA 01810, USA</text>
</place>
<publisher>John Wiley & sons, Inc.; [etc.]</publisher>
<dateIssued>1916</dateIssued>
<dateCaptured>2004</dateCaptured>
<issuance>monographic</issuance>
</originInfo>
<language authority="iso639-2b">eng</language>
</mods>
```

The image files are OCRed and the raw text is written to Tag = 5688(1638), one word per line. This tag is a private tag but registered and published by the ResCarta Foundation for this purpose.

.



INTRODUCTION

HUMAN beings may be traced in almost any part of the globe through the tin cans which they leave behind them.

Certainly, anywhere in North America where people have spent a day, there you may find this sign of their occupancy.

To-day the retailer receives from the wholesaler the larger portion of his food supplies in a package, which is transmitted unbroken to the consumer. A grocery-store consists largely of a collection of original packages.

Of all these sanitary devices, the tin can is probably the most significant and the most universal. Fruits, vegetables, meats and milk have all been subjected to the virtue of this humble container. Nicholas Appert, in France, first preserved food in glass jars by sealing them hermetically and heating. He published "The Art of Preserving Animal and Vegetable Substances" in 1811. In 1810 Peter Durand obtained a patent in England for preserving fruits, vegetables and fish by hermetically sealing them in *tin* and glass cans. In 1820 William Underwood and Charles Mitchell, emigrant employees from a canning factory in England, opened a factory in Boston where they canned plums, quinces, cranberries and currants. Glass was used exclusively until 1825, when Thomas Kensett secured a patent for use of *tin* cans and commenced to use them in his factory.

The preservation of foodstuffs lies in controlling the action of microorganisms. There are four methods of control in general use—desiccation, addition of toxic substances, refrigeration and exclusion. It is the method of exclusion, by placing in a sealed container and then heating, that is coming to be the most acceptable method for preserving the delicate flavor of

vii

<CONTENT>
INTRODUCTION
HUMAN
beings
may
be
traced
in
almost
any
part
of
the
globe
through
the
…
preserving
the
delicate
flavor
of
vii
</CONTENT>

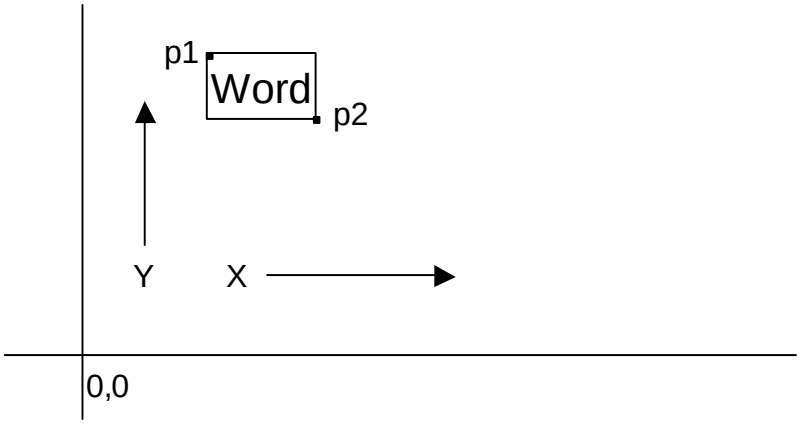


Figure 1: Word Location Coordinate System

OCR word location information is written to Tag = 5689(1639). This tag is a private tag but registered and published by the ResCarta Foundation for this purpose.

OCR word location data is comprised of ZLIB compressed data in a set of pairs defining a space by x,y coordinate values from the lower left hand corner of the image with values defined by pixels in units at 72 ppi multiplied by 10 and represented by integers. Each word is bounded by a rectangle area formed by a point (p1) in the upper left hand corner of the area and a point (p2) in the lower right hand corner of the area. The font name and resolved font size is also encoded. The fields are comma separated and followed by a newline (10). The word sequence is determined by its location in the table listed in Tag = 5688(1638).

X of p1, Y of p1, X of p2, Y of p2,"Font Name",Point Size (Newline)

```
350,6132,3087,5882,"Times-Bold",32
3280,6143,4693,5913,"Times-Bold",32
2270,5308,2337,5258,"Times-Roman",10
2370,5349,2542,5259,"Helvetica",10
350,4661,1842,4381,"Times-Roman",38
2030,4629,2360,4399,"Times-Roman",31
2550,4715,3551,4405,"Times-Roman",40
3900,4699,4493,4419,"Times-Roman",38
360,3144,1054,2994,"Times-Roman",19
1140,3122,1920,3002,"Times-Bold",15
```

Technical metadata is recorded in the appropriate TIFF tags as specified in the TIFF 6.0 Final specification of June 3, 1993 shown here.

Table 4: Technical Metadata as Specific in the TIFF 6.0 Final Specification (June 3, 1993)		
NewSubfileType Tag = 254(FE) Type = LONG Length = 1 Value = 0	FillOrder Tag = 266(10A) Type = SHORT Length = 1 Value = 1	XResolution Tag = 282(11A) Type = RATIONAL Length = 1 Value = 600 / 1
ImageWidth Tag = 256(100) Type = SHORT Length = 1 Value = 3304	StripOffsets Tag = 273(111) Type = LONG Length = 1 Value = 486	YResolution Tag = 283(11B) Type = RATIONAL Length = 1 Value = 600 / 1
ImageLength Tag = 257(101) Type = SHORT Length = 1 Value = 5400	Orientation Tag = 274(112) Type = SHORT Length = 1 Value = 1	PlanarConfiguration Tag = 284(11C) Type = SHORT Length = 1 Value = 1
BitsPerSample Tag = 258(102) Type = SHORT Length = 1 Value = 1	SamplesPerPixel Tag = 277(115) Type = SHORT Length = 1 Value = 1	ResolutionUnit Tag = 296(128) Type = SHORT Length = 1 Value = 2
Compression Tag = 259(103) Type = SHORT Length = 1 Value = 4	RowsPerStrip Tag = 278(116) Type = LONG Length = 1 Value = 5400	Software Tag = 305(131) Type = ASCII Length = 15 Value = RC2004-SS1.001
PhotometricInterpretation Tag = 262(106) Type = SHORT Length = 1 Value = 0	StripByteCounts Tag = 279(117) Type = LONG Length = 1 Value = 69416	DateTime Tag = 306(132) Type = ASCII Length = 20 Value = 2004:01:08 13:02:07

These technical tags are the minimum required, the content of the software tag for ResCarta objects is defined in the ResCarta Standards Overview. Most of the above will vary by source materials ad time of conversion except for orientation tag, which is required to be 1. Other specified and private tags may be added at time of creation.

A checksum is created and placed into the <filesec> of the object level metadata file. Using the SHA1 checksum in this example, the contents would look like…

Table 5: SHA1 Cheksum File for the Serial
Hannibal City Directory 1859 - 1860

```
Hash      File Name                          Hash
Type
SHA1   (0001000A.TIF)   =   71591e6c8616bb7b4eaedeb17283db909578daa0
SHA1   (0002000B.TIF)   =   4944c9481c4c0a63220cb703dea3dbfa63eb24ad
SHA1   (0003R001.TIF)   =   75D2BC91554827E4906078CD9F072E62813D2532
SHA1   (0004R002.TIF)   =   D6AB389718B40957FB46C90D99E41BD45735FA00
SHA1   (0005R003.TIF)   =   1FF593931418AB30A34ED96AB4581978C42F1D1C
SHA1   (0006R004.TIF)   =   576B7F009A062F9215B8007A91C0B81644D2C0BB
SHA1   (0007R005.TIF)   =   EF829570ABCBA7EEEAB4E5B2DDE750F73B4BA6B1
SHA1   (0008R006.TIF)   =   C7EBDF4AEDF540164DF2681954CE364C8FDAC4EC
SHA1   (0009R007.TIF)   =   3ACB98DB140B382F3CB32BC9B4EED37AA862BB1A
SHA1   (0010R008.TIF)   =   66F79E8A8C46AE2202E0FBFD482FC985498F4D29
SHA1   (0011R009.TIF)   =   B5181055918DECB540BA3A5D6404F2ECFBB3A438
SHA1   (0012R010.TIF)   =   39CF3D136C669E63183192272F07EE19AED7A732
SHA1   (0013R011.TIF)   =   41E1EB1784C5318A9394E607A921B0750C06222E
SHA1   (0014R012.TIF)   =   F40CEF1ADF9655F429B1C04239AD1FD200C3B867
SHA1   (00150001.TIF)   =   FDE975932A1DB069EAA92D6D9EA30217E7901F68
SHA1   (00160002.TIF)   =   D227EB68A58A1FBD2A5692F957081FA255494F8B
…
SHA1   (02250211.TIF)   =   516AB6060187C2C08C0CF0C1C14302840108E5C4
SHA1   (02260212.TIF)   =   64910375C00C8B03443BF61C5007ABF2677B5376
SHA1   (02270213.TIF)   =   7321BE937420BFFB3C38D3565F318453537DE25C
SHA1   (02280214.TIF)   =   3F8F69E86CBC4A85D97E78C56D8DC6D8B4BC1ED0
```

The checksum is stored in the metadata.xml file like

```
<fileSec>
  <fileGrp ID="FG0001">
   <file ID="FID0001" MIMETYPE="image/tiff" SEQ="1" SIZE="457500"
CHECKSUM="71591e6c8616bb7b4eaedeb17283db909578daa0" CHECKSUMTYPE="SHA-1"
GROUPID="FG0001">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="file://./0001000A.TIF"/>
   </file>
   <file ID="FID0002" MIMETYPE="image/tiff" SEQ="2" SIZE="51683"
CHECKSUM="4944c9481c4c0a63220cb703dea3dbfa63eb24ad" CHECKSUMTYPE="SHA-1"
GROUPID="FG0001">
    <FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="file://./0002000B.TIF"/>
   </file>…
```

Structural metadata is stored in a METS file directly under the root identifier and named metadata.xml.

Table 6: Sructural Metadata Stored in a METS file for the Serial *Hannibal City Directory 1859 – 1860*


```
<?xml version="1.0" encoding="utf-8" standalone="no"?>
<mets OBJID="rescarta/HF000001/00000001" LABEL="Hannibal City Directory
1873-74." TYPE="book" xmlns="http://www.loc.gov/METS/"
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">
 <metsHdr CREATEDATE="2006-09-26T14:04:31" RECORDSTATUS="Complete">
  <agent ROLE="CREATOR" TYPE="ORGANIZATION">
   <name>ResCarta Tools</name>
   <note>From NM2000 formatted data</note>
  </agent>
 </metsHdr>
 <dmdSec ID="DMD0001">
  <mdWrap MIMETYPE="text/xml" MDTYPE="MODS">
   <xmlData xmlns:mods="http://www.loc.gov/mods/v3"
xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
    <mods:mods>
     <mods:titleInfo>
      <mods:title>Hannibal City Directory 1873-74.</mods:title>
     </mods:titleInfo>
     <mods:name authority="local" type="corporate">
      <mods:namePart>Hannibal Public Library</mods:namePart>
      <mods:role>
       <mods:roleTerm authority="marcrelator"
type="text">Owner</mods:roleTerm>
       <mods:roleTerm authority="marcrelator"
type="code">own</mods:roleTerm>
      </mods:role>
     </mods:name>
     <mods:typeOfResource>text</mods:typeOfResource>
     <mods:originInfo>
      <mods:dateIssued encoding="iso8601">1873</mods:dateIssued>
      <mods:dateCaptured encoding="iso8601">2005-06-
21</mods:dateCaptured>
      <mods:issuance>continuing</mods:issuance>
     </mods:originInfo>
     <mods:language>
      <mods:languageTerm authority="iso639-2b"
type="code">eng</mods:languageTerm>
     </mods:language>
     <mods:physicalDescription>
      <mods:form authority="marcform">electronic</mods:form>
      <mods:extent>5.375 x 8.625 in.</mods:extent>
      <mods:digitalOrigin>reformatted digital</mods:digitalOrigin>
     </mods:physicalDescription>
     <mods:subject>
```

```
      <mods:hierarchicalGeographic>
       <mods:country>USA</mods:country>
       <mods:state>MO</mods:state>
       <mods:county>Marion, Ralls</mods:county>
      </mods:hierarchicalGeographic>
     </mods:subject>
     <mods:subject authority="LCSH">
      <mods:topic>Hannibal (Mo.)</mods:topic>
     </mods:subject>
     <mods:subject authority="LCSH">
      <mods:topic>Directories.</mods:topic>
     </mods:subject>
     <mods:subject authority="LCSH">
      <mods:topic>History.</mods:topic>
     </mods:subject>
     <mods:location>
      <mods:url>hannibal/HF000001/00000001</mods:url>
     </mods:location>
     <mods:identifier type="local"> hannibal
/HF000001/00000001</mods:identifier>
    </mods:mods>
</xmlData>
  </mdWrap>
 </dmdSec>
 <amdSec ID="AMD0001">
  <techMD ID="TMD0001">
   <mdWrap MIMETYPE="text/xml" MDTYPE="NISOIMG">
    <xmlData xmlns:mix="http://www.loc.gov/mix/"
xsi:schemaLocation="http://www.loc.gov/mix/
http://www.loc.gov/mix/mix.xsd">
     <mix:mix>
      <mix:BasicImageParameters>
       <mix:Format>
        <mix:MIMEType>image/tiff</mix:MIMEType>
        <mix:ByteOrder>little-endian</mix:ByteOrder>
        <mix:Compression>
         <mix:CompressionScheme>4</mix:CompressionScheme>
        </mix:Compression>
        <mix:PhotometricInterpretation>
         <mix:ColorSpace>0</mix:ColorSpace>
        </mix:PhotometricInterpretation>
       </mix:Format>
      </mix:BasicImageParameters>
     </mix:mix>
    </xmlData>
   </mdWrap>
  </techMD>
 </amdSec>
 <fileSec>
  <fileGrp ID="FG0001">
   <file ID="FID0001" MIMETYPE="image/tiff" SEQ="1" SIZE="457500"
CHECKSUM="71591e6c8616bb7b4eaedeb17283db909578daa0" CHECKSUMTYPE="SHA-
1" GROUPID="FG0001">
    <FLocat LOCTYPE="URL" xlink:type="simple"
xlink:href="file://./0001000A.TIF"/>
   </file>
```

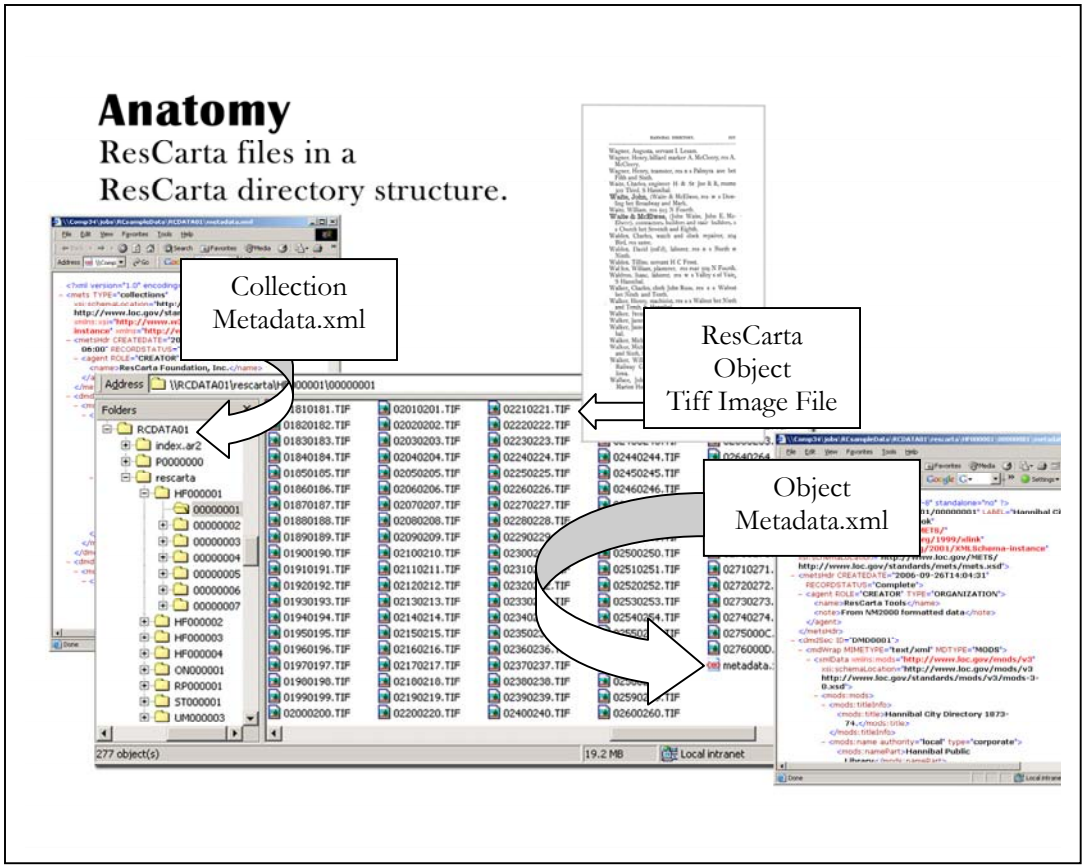
```
   <file ID="FID0002" MIMETYPE="image/tiff" SEQ="2" SIZE="51683"
CHECKSUM="4944c9481c4c0a63220cb703dea3dbfa63eb24ad" CHECKSUMTYPE="SHA-1"
GROUPID="FG0001">
    <FLocat LOCTYPE="URL" xlink:type="simple"
xlink:href="file://./0002000B.TIF"/>
   </file>
```


…Continues in a like manner…


```
<file ID="FID0275" MIMETYPE="image/tiff" SEQ="275" SIZE="93410"
CHECKSUM="29e8707337b1b2ead795495aeeec0c98662240b8" CHECKSUMTYPE="SHA-1"
GROUPID="FG0001">
    <FLocat LOCTYPE="URL" xlink:type="simple"
xlink:href="file://./0275000C.TIF"/>
   </file>
   <file ID="FID0276" MIMETYPE="image/tiff" SEQ="276" SIZE="514957"
CHECKSUM="22a10881295e45fbb03ccc1d3310824766d17d32" CHECKSUMTYPE="SHA-1"
GROUPID="FG0001">
    <FLocat LOCTYPE="URL" xlink:type="simple"
xlink:href="file://./0276000D.TIF"/>
   </file>
  </fileGrp>
 </fileSec>
 <structMap TYPE="LOGICAL">
  <div ORDER="1" DMDID="DMD0001" TYPE="serial">
   <div ORDER="1" LABEL="1873" TYPE="volume">
    <div ORDER="1" LABEL="000" TYPE="issue">
     <div ORDER="1" LABEL="000A" ADMID="AMD0001" TYPE="page">
      <fptr FILEID="FID0001"/>
     </div>
     <div ORDER="2" LABEL="000B" ADMID="AMD0001" TYPE="page">
      <fptr FILEID="FID0002"/>
     </div>
     <div ORDER="3" LABEL="0007" ADMID="AMD0001" TYPE="page">
      <fptr FILEID="FID0003"/>
     </div>
```


…Continues in a like manner…


```
      <div ORDER="273" LABEL="0273" ADMID="AMD0001" TYPE="page">
      <fptr FILEID="FID0273"/>
     </div>
     <div ORDER="274" LABEL="0274" ADMID="AMD0001" TYPE="page">
      <fptr FILEID="FID0274"/>
     </div>
     <div ORDER="275" LABEL="000C" ADMID="AMD0001" TYPE="page">
      <fptr FILEID="FID0275"/>
     </div>
     <div ORDER="276" LABEL="000D" ADMID="AMD0001" TYPE="page">
      <fptr FILEID="FID0276"/>
     </div>
    </div>
   </div>
  </div>
 </structMap>
</mets>
```



This document can be referred to as version RC2007.bk.1.000.8.1

For further information see…

Digital Object Identifiers
http://www.doi.org

Handle System
http://www.handle.net

METS
http://www.loc.gov/standards/mets/

MODS
http://www.loc.gov/standards/mods/

ResCarta
http://www.rescarta.org/

TIFF
http://www.alternatiff.com/resources/TIFF6.pdf