

Livingstone Online Downloadable Archival Packet READ_THIS_FILE
Adrian S. Wisnicki, Project Director (18 September 2016)

Contents

- A) Overview
- B) Introduction
- C) File Naming
- D) Archival Packet Contents
- E) Metadata

A) Overview

This file describes the contents of the archival packets of David Livingstone manuscript and contextual images and transcriptions published by Livingstone Online (<http://livingstoneonline.org>). Each packet corresponds to one item in our digital collection (e.g., diary, letter, map, illustration, etc.) or, in the case of items with spectral images, one page of one item and will contain a variety of file types: docx, jpg, md5, pdf, txt, xlsx, xml, and xmp.

1) General users should take note of the following files in particular:

a) 1_Livingstone_Online_Digital_Catalogue_18_Sept_2016.xlsx – This file presents our digital catalogue in spreadsheet form and so provides full information about each Livingstone item published in our digital collection. For context, we also include full information about a large variety of items that we do not currently publish. Users can sort and filter this table according to their needs.

Given our standardized file naming practices (see below), the first column of the spreadsheet ("Livingstone Online Project File Number") enables easy identification of the item that corresponds to the downloaded archival packet;

b) the "copyright_information.txt" file – This file describes the terms under which a given item is published and can be reused by our users. Livingstone Online takes the rights of our collaborating individuals and archives very seriously, and we cite copyright information wherever possible. We ask that our users also respect these rights. We have made the copyright information available in this separate, additional form for easy access;

c) the "reading_copy.pdf" file (if present) – This file offers a user-friendly, reading copy transcription of the given collection item; and

d) the .jpg images – These files present the individual pages of the given collection item.

2) Specialist users will find a detailed description of the archival packets in the text that constitutes the rest of this file.

B) Introduction

Whenever possible and permitted by our stakeholders, Livingstone Online allows users to download images, transcriptions, and metadata associated with the items in our digital collection. We provide all the files associated with the given item in a single .zip file, as part of our broader educational mission to promote the widest possible, non-commercial use of our core source materials. The present file, which we include in each .zip file, defines the contents of this archival packet and sets out the guidelines by which the data in the packet has been created and assembled.

C) File naming

We assign each item in our digital collection a base file name. This base file name consists of only a "liv" prefix plus a unique six-digit item number, as in the following examples:

liv_000455

liv_013013

The base file name for each item documented in our digital catalogue can be found in the first column ("Livingstone Online Project File Number") of the 1_Livingstone_Online_Digital_Catalogue_1_Apr_2016.xlsx file included in our archival packets.

The base file number has no semantic value, and we make it a practice not to put any metadata or bibliographic information into our file names. Rather we keep all such information in our metadata files.

Images receive an additional four digit segment that identifies the specific image in the item sequence. For instance, the following are the third and fourth images of the given item:

liv_000005_0003.jpg

liv_000005_0004.jpg

MODS and TEI files for an item have the relevant acronyms added to the base file name:

liv_000149_MODS.xml

liv_002010_TEI.xml

We also provide a TXT file with basic copyright information related to the given item and a PDF-based reading copy transcription of the item:

liv_014080_copyright_information.txt

liv_001753_reading_copy.pdf

Consequently, the base file names enable the easy association and sequential organization of all files related to an item, while additions to this name and/or file suffixes allow differentiation among the files, as in the following example:

liv_000063 – base file name

liv_000063_0001.jpg – JPEG image file generated from the original TIFF images held by Livingstone Online

liv_000063_0001.jpg.md5 – MD5 data integrity verification file for the given JPEG image

liv_000063_0001.tif.txt – TXT file of Dublin Core and image capture and processing metadata extracted from the original TIFF image

liv_000063_0001.tif.xmp – XMP sidecar file of metadata extracted from the original TIFF image

liv_000063_copyright_information.txt – TXT file with item title, date, and copyright information

liv_000063_MODS.xml – XML MODS metadata file

liv_000063_TEI.xml – XML TEI P5 transcription file (if available)

liv_000063_reading_copy.pdf – User-friendly transcription generated from the TEI file (if available)

D) Archival Packet contents

Each archival packet will contain in its main directory some combination of the foregoing types of files, named and organized as described in the previous section. Each archival packet will also contain the present file, which we include in in DOCX, PDF, and TXT versions, and 1_Livingstone_Online_Digital_Catalogue_1_Apr_2016.xlsx, which provides detailed metadata about each item in our collection in an easy-to-consult format.

Users should, however, take note of the following:

1) MODS XML metadata and copyright information TXT files: always included. These files document each item in our collection, so there will always be one of each of these in the main directory of the archival packet. Importantly, these files contain the use and reproduction rights for any given item, so users should always consult these rights before reusing images from our digital collection elsewhere.

2) TEI P5 XML transcription and reading copy transcription PDF files: included if available. If we have transcribed and critically encoded a given item in a TEI P5 XML file, that file will be included in the main directory of the archival packet. If a TEI transcription exists, then users will also find a user-friendly, reading copy transcription PDF file generated through a modified version of oXygen's default TEI P5 PDF transformation.

3) JPEG images files: Livingstone Online does not own any of its core image data. Rather we have received this data thanks to the generosity of collaborating individuals and archives. In each case, these individuals and archives have made the data available to us under specific conditions. In some cases, we have been allowed to make the data available without restriction. In other cases, we are only allowed to provide lower-resolution and/or watermarked JPEG images to our users. So the kinds of image files we do and don't make available depend on the permission given us by our collaborators.

4) TIFF images files: not included. Livingstone Online archival packets do not presently include TIFF images. Users with legitimate research needs, however, may request access to TIFF images from the Livingstone Online project team. Provision of TIFF images will depend on the permissions provided by the holding institution.

5) XMP sidecar and Dublin Core and image capture and processing TXT metadata files: extracted from the original TIFF images. Our workflow to produce the Dublin Core and XMP sidecar metadata files has taken the following form:

a) Convert the information in the 1_Livingstone_Online_Digital_Catalogue_1_Apr_2016.xlsx file to CSV form;

b) Create a MODS metadata file for a given object from the CSV file using a Groovy script;

c) Crosswalk the MODS metadata (in XLSX form) to the Dublin Core metadata (in XLSX form);

d) Convert Dublin Core metadata from XLSX form to CSV form.

e) Use a Ruby script to embed Dublin Core metadata in TIFF image file headers alongside pre-existing image capture and processing metadata; and

f) Use the same Ruby script to extract Dublin Core metadata plus other pre-existing image metadata from image headers to create TXT metadata and XMP sidecar metadata files. (Note: Embedding the Dublin Core information in the image file headers and extracting the image metadata into TXT and XMP files is done through a single process.)

As a result, our Dublin Core and XMP sidecar metadata files are partly derivative from (and indeed less comprehensive than) our MODS metadata files. Users are therefore directed to the MODS metadata files in the first instance, and we provide these derivative files only for an additional layer of image documentation.

6) MD5 files: included for each image in the archival packet. We provide these files as a means for our users to ensure that the JPEG image files downloaded from our site have not become corrupt.

E) Metadata

We build detailed metadata for each item in our digital collection in an XML file created according to Version 3 of the Metadata Object Description Schema (MODS) (<http://www.loc.gov/standards/mods/v3/>). Our MODS files include the following elements:

identifier – the base file name for the given item; the genre and number of the item as set out in the Clendennan and Cunningham catalogues of Livingstone documents (1979, 1985); where relevant, the shelfmark of a copy of the item held by the National Library of Scotland

titleInfo.title – the title of the item with and without the date

name.namePart and name.description – the authority name and birth and death dates of the creator(s) and, if a letter, the authority name of the addressee(s); a short set of biographical facts related to the addressee; the authority name of the repository as set out, whenever possible, in the Library of Congress Name Authority File (NAF) file (<http://id.loc.gov/authorities/names.html>)

genre – the genre of the item as drawn from the Getty Research Institute's Art and Architecture Thesaurus Online (<http://www.getty.edu/research/tools/vocabularies/aat/>)

originInfo.dateCreated – the date of the item in written day-month-year form; the date as expressed according to the ISO 8601 format (<http://www.iso.org/iso/home/standards/iso8601.htm>)

originInfo.place.placeTerm – the place where an item was created or composed, as specified by Livingstone himself or, if not specified, then as supplied by the Livingstone Online project team based on contextual inference; the authority name in the Library of Congress Name Authority File (NAF) file (<http://id.loc.gov/authorities/names.html>) of the place where an item was created or composed

subject.cartographics - the approximate latitude and longitude of the place where an item was created or composed

physicalDescription.note and physicalDescription.extent – physical details relating to an item, including whether it takes the form of a manuscript, photocopy, typescript, newspaper item, or other printed format; the page length of the item; the size of the item in millimeters

location.shelfLocator – the repository shelfmark or identifier

accessCondition – the terms by which an item is available for use and reuse

relatedItem.identifier – the bibliographical details or URL for any previous full or partial publications of the item

As noted above, the MODS metadata also forms the basis for the Dublin Core metadata that we add to the image header of each image and for the derivative TXT and XMP standalone metadata files that we provide with each image. We have crosswalked the MODS metadata to Dublin Core metadata using the following equivalencies:

mods:identifier = dc:identifier

mods:title = dc:title (Note: We used an alternative version of mods:title with the date.)

mods:name and mods:role = dc:creator or dc:contributor

mods:name plus mods:shelfLocator = dc:description

mods:genre = dc:type

mods:physicalDescription = dc:format (Note: We crosswalked only a portion of our mods:physicalDescription information.)

mods:dateCreated = dc:date (Note: We used the ISO 8601 format of the date and, for date ranges, crosswalked only the first half of the range.)

mods:publisher = dc:publisher

mods:accessCondition = dc:rights

As a result, the Dublin Core (and XMP sidecar metadata) files only contain a sampling of the rich metadata found in our MODS files.