

Livingstone Online Metadata

Fields

document_type

This field is used to record an abbreviation of the type of document the item is.

The majority of items in the Livingstone Online project are letters, abbreviated "LETT". But this is going to be changing, as from DLC and NLS we are bringing in Diaries, Maps, Notebooks and all kinds of other stuff. We should come up with a standardized abbreviation list.

type_number

This field contains a unique number for the item, which when paired with the document_type value becomes a unique identifier for the item. The numbering system began at one and was auto-incremented. The cleanliness of the number data is generally expected but not known. That is, it is possible there are numbers that were skipped. *I am reluctant to include this category. Right now this is the key to sorting out the online data, but it is also what makes working with that data confusing. I suggest we generate the unique identifier using a combo of document_type and cc_catalogue. In other words, I strongly urge that we remove this field after we've gleaned from this whatever information we need.*

cc_catalog

The "cc catalog" refers to a canonical, printed catalogue of documents related to David Livingstone (*Clendennen and Cunningham were the editors*). The cc_catalog number is an important identifier, but does not have a match with the numbers created under the type_number column. *(exactly and that's what makes type_number confusing. The cc_catalog number might be described as the canonical number and so much better to go with that for the unique identifier).*

author

Initials of the author of the item.

recipient

Name of the recipient of the letter in direct order. *Note that this field will be blank for other kinds of documents.*

recipientNorm

Ostensibly, a normalized form of the name of the recipient. However, the data reveals little or no modification of the form from the recipient field. Names in this field remain in direct order.

title

The title field appears to have been left blank for letters. *We could auto generate this using the document_type, recipient, and date_string columns. We need to discuss how we'll handle this field, if at all. I think we should but we would need to fill in the blanks, so to speak for letters.*

date_string

A text field in day month year format for the date written.

date

A numeric field for the date created using YYYY-MM-DD format.

place

The place the letter writer was when writing the letter. Note that this field will be blank for other kinds of documents.

placeNorm

Ostensibly, a normalized version of the place name from the “place” field, but it doesn’t appear to have much normalization; square brackets have been removed. We may want to keep it as Livingstone doesn’t always spell the names of places the same way, and, moreover, his usage may sometimes be different from modern day usage.

length

Presumably the length of the letter, but data values are unclear. This is taken from the C&C catalog. I can provide a scan of the relevant page of the catalog for this when necessary. And, yes, this is in a more complex manner the length of the letter, is it 4 pages, 5 pages, etc.

size1

Probably a length or width measurement. Not standardized. Data ranges from “lge” to “4to” (meaning quarto) to “400”. Plain numeric data is probably in millimeters.

size2

See “size1”.

type

Seemingly a further genre / manuscript description. Sample data includes “Pr copy” and “MS aut”. “Pr copy” probably means “Print copy” and “MS aut” probably means “manuscript autograph”. Yes, again taken directly from the C&C catalog.

n

unknown. The only value other than NULL I see for this is “n”. We can omit this.

holder

Abbreviation for the name of the holding institution.

shelfmark

Call number or location information for the item in its holding institution.

folio

unknown. Folio number, if assigned, by the holding institution.

sorting_folio

unknown. Never used as far as I can see. Cut.

location

unknown.

msnumber

the shelfmark of the corresponding copy of this document at the NLS. Valuable data and worth keeping.

sub

Not sure what this is. Nothing entered.

publication

The abbreviations here are explained in the C&C catalogue. If a document has been published, this is where it appears. Worth keeping.

The rest of the fields in MySQL, reading from left to right, don't contain anything or anything worth preserving as far as I can see. The only exception is images_online which appears to have either NULL or 1 depending on whether images are not or are online. Could be helpful to us when we sort the image TIFFdata to at least know which letters we should be finding data for.

Additional Notes

Document_type + cc_catalog= possible unique identifier

There's a lot of rich metadata in the MySQL database. Reviewing this reminds of that. It'd be great to autodisplay a lot of this when we show images of manuscripts on the site etc. as it would really be a benefit to scholars.