Gary Griswold
gary@shortsands.com
513-508-6127
August 17, 2020
Revision 2

DPB v4 Data Dictionary

Descriptions of DBP data affected by Bible, Fileset load

This document describes the content of the DBP database for each table and column that is effected by the new load of the DBP. Additional tables will be added to this report as the logic to update those tables nears deployment.

bibles Table

All of the columns of the bibles table are returned by the /languages API query, while the /bibles API query returns bible_id, language_id, date, and copyright.

Often, there is one LPTS record for each Bible, but very frequently, there is more than one. For example, ENGKJV has stock numbers: C1KJV, C2KJV, N1KJV, N2KJV, and P2KJV. In those cases where there are multiple records, if each record has the same value, then the program uses that value. If they do not have the same value, then in most cases, the program displays a warning to notify the data administrator might that a correction might be necessary, the program arbitrarily picks one of the values to update the database.

1,2,3 linkage

Most data fields in the LPTS occur no more than once in each record. However, there are a number of fields that have a suffix on their name of 1, 2, or 3. These variants exist because some languages can be written in more than one alphabet, such as Arabic, Cyrillic, and Latin. In cases where multiple alphabets do exist, the data fields will use the 1, 2, 3 suffix to tie together

data fields that are related to the same alphabet. The words script and orthography are synonyms of alphabet.

id - primary key

Bible id is often 6 characters and sometimes as many as 8. Often the first 3 characters match the ISO language code, and often the 3 following characters are an abbreviation of the version. The LPTS source for this field is DBP_Equivalent, DBP_Equivalent2, and DBP_Equivalent3. A single LPTS record can have at most one of each of these.

language_id

language_id is the primary key of the languages table. It is primarily found by doing a lookup in the languages_translations table using the LangName and ISO from the LPTS record. This query usually finds the correct record, but if the LangName is absent or does not match anything in the languages_translations table, then a second lookup is done in the languages table using only the ISO code. If the ISO code is missing or cannot be found in the languages table, an error is produced. This error prevents the bibles table from being updated.

versification

The versification field contains information about the sequence of the books of the Bible in this specific Bible. It is returned by the /languages API query, it is not known whether the App uses this information. Old records updated by DBS (Digital Bible Society) contained a few different values for this field, but LPTS was not the source. Later, this field was defaulted to 'protestant' for all records, which is incorrect. At this writing, the new load is populating this field with a concatenation of LPTS field OTOrder,NTOrder.

The filenames stored in the bible_files table do contain an explicit book sequence field for most filesets. If this information were parsed from the filename and made part of the the bible_files table, or the bible_books table, it would become an explicit source of information about the sequence of books.

numeral_system_id

The numeral_system_id is the primary key of the numeral_systems table, which together with the numeral_system_glyphs table and the alphabet_numeral_systems table provides information of the numeric digits of the language of this Bible.

The exact meaning of this field is unclear, because each alphabet (alias script, orthography) has its own unicode space for numeric digits. So, knowing the alphabet should be sufficient to knowing the numeral_system_id. The only purpose, I can think of for setting this field is when the numeral_system_id to be used by a specific Bible is different than the alphabet of that Bible.

This field is sourced from LPTS Numerals field, which should have a 1, 2, 3 linkage, but does not.

date

The date column is returned in the API by most requests that return any kind of Bible data. This field is source is a 4 digit year field parsed from the Copyrightc (text copyright) field. If there is no year in the Copyrightc field, or the Copyrightc field is absent, the attempts to parse a 4 digit year from the Volumne_Name field.

scope

The scope field is only returned by the /languages API query as part of a bibles record. The source of this field is the set_size_code for all of the filesets of this Bible. It combines these doing a logical union of the values.

script

The script code (alias alphabet, orthography) is returned by the /languages and /bibles API queries. It is sourced from the the LPTS fields: _x0031_Orthography, _x0032_Orthography, _x0033_Orthography. These orthography fields contain a name, and not the standard ISO 15925 standard script codes. So, these names are looked up in the lpts_script_codes table to retrieve the correct script_id. If no script code is found, the data administrator should add the correct record to the lpts_script_codes table.

A program should be written that opens Bible html or USX files and determines the range of unicode values found in a sample of text. From that information the program could identify the script perfectly.

derived

The derived column is returned by the /languages API query. The DBP database currently has this data for only 97 row. LPTS has no source to populate this field, so the update program does not update this field

copyright

The copyright column is returned by the /languages and /bibles API query. It is the bible_fileset_copyrights table that contains the full copyright message that is included in the App. This column will either be populated with the Copyrightc field (text copyright) from LPTS.

priority

The priority column is only returned by the /languages query. There are only 45 rows where this field has a non-zero value. For the rows where the value is set, it appears to refer to a sequence that should prioritize versions of a language. This field is set manually, and is not updated by the load program.

reviewed

The reviewed column is returned by the /languages query. About 95% of this column have a value of 1 and 5% have a value of 0. The meaning of these values is not known, and this field is not updated by the load program.

notes

The notes column is returned by the languages query. The current DBP database has only 4 rows where this column contains data, and has no known purpose. This field is not updated by the load.

bible_fileset_copyrights table

The bible_fileset_copyrights table only appears in the /bible/filesets/{fileset_id}/copyright API request. The copyright data in this table is displayed in the App.

For 6 digit text filesets, there will sometimes be more than one 10 digit text damid, and sometimes more than one LPTS record. When this happens, the program looks for one that is live, and selects that one. Or, if there is more than one that is live, it will arbitrarily pick one. Or, if none are live, it will arbitrarily pick one of the records that has a damId that is not live, and use its copyright data.

hash_id - primary key

The hash_id is what links a record in this table to the corresponding record in bible_filesets. There should be exactly one row in this table for every row in bible_filesets, which means the fields in this table could instead be part of the bible_filesets table.

copyright_date

The process searches for a 4 digit year in the copyright field, and uses it to update the copyright_date. If the fileset_id is text, then Copyrightc is the source. If the fileset_id is audio, the Copyrightp is the source. If the fileset_id is video then Copyright_Video is the source.

copyright

This field or copyright_description are is the copyright field that is used within the App. When the fileset type is text, Copyrightc is used. When the fileset type is audio, the contents are produced as "Text: {Copyrightc}\nAudio: {Copyrightp}". When the fileset type is video, the contents are produced as "Text: {Copyrightc}\nAudio: {Copyrightp}\nVideo: {Copyright_Video}"

copyright_description

This field contains the identical contents as copyright column. If we knew for certain which of these two fields was being used by the App, the other could be set to null.

open_access

This field is not being set, it always contains the default value of 1. The meaning of this field is not known.

bible_fileset_tags table

The contents of the bible_fileset_tags table does not appear anywhere in the API results. For text filesets, there will sometimes be more than one 10 digit text damid that matches the 6 character fileset_id, and sometimes more than one LPTS record. When this happens, the program looks for one that is live, and selects that one. Or, if there is more than one that is live, it will arbitrarily pick one. Or, if none are live, it will arbitrarily pick one of the records that has a damId that is not live.

hash_id

The hash_id is what links a record in this table to the corresponding record in bible_filesets. The hash_id and name together are the primary key.

name

The name field is a constant that defines the content of the description field. The possible values are bitrate, stock_no, and volume. The volume name is slated for deletion. There are also records with the name, timing_est_err, but these are set by the AudioHLS process, and are not affected by the DBP load.

description

The contents of description varies depending upon the constant in the name field.

- bitrate - for audio filesets only, it contains 16kbps for filesets whose 10th-12th digits are 16, and otherwise contains 64kbps.
- stock_no - it contains the LPTS Reg_StockNumber of the record that data was extracted from.
- volume - it contains the LPTS Volume_Name. This field is slated for deletion.

admin_only

This is set to a constant value of 1. It has no know meaning.

notes

This field is not set by the load, and defaults to null.

iso

This is set to a constant value of ‘eng’.

language_id

This is set to a constant value 6414.

bible_filleset_copyright_organizations Table

The bible_fileset_copyright_organizations table only appears in the /bible/filesets/{fileset_id}/copyright API request. We think that the organization_id information in this table is used to retrieve the correct logos to display for a Bible fileset. Comments made under bible_fileset_copyrights about duplicate records found, because DBP uses 6 digit filesets, and LPTS uses 10 digit filesets apply to this table as well.

hash_id

The hash_id is what links a record in this table to the corresponding record in bible_filesets. The hash_id, organization_id, and organization_role together are the primary key.

organization_role

The organization_role is 1 for copyright holder, and 2 for licensor. When a record is set by LPTS Copyrightc, Copyrightp, or Copyright_Video, the organization_role is 1. When a record is set by LPTS Licensor or CoLicensor, the organization role is 2.

organization_id

The load program finds LPTS Copyrightc, Copyrightp, Copyright_Video, Licensor, and CoLicensor fields in the lpts_organizations table. This lookup returns the organization_id. When the copyright or licensor is not found, a correct lpts_organizations record should be added by the table, so that the organization_id will be found.

access_group_filesets Table

The access_group_filesets defines the allowed type of permissions for each Bible fileset.

hash_id

The hash_id links each access_group_filesets record to a single specific record in the bible_filesets table.

access_group_id

The LPTS fields listed in the second column are read and when the value is -1, the corresponding DBP access_group is assigned to this table.

- allow_text_NT_DBP | DBPText
- allow_text_OT_DBP | DBPTextOT
- allow_audio_DBP | DBPAudio
- allow_text_WEB | HubText
- allow_audio_WEB | DBPWebHub
- allow_video_WEB | WebHubVideo
- allow_text_API | APIDevText
- allow_audio_API | APIDevAudio
- allow_video_API | APIDevVideo
- allow_text_APP | MobileText
- allow_audio_APP | DBPMobile
- allow_video_APP | MobileVideo
- allow_text_GBA | GBN_Text
- allow_audio_GBA | GBN_Audio
- allow_video_GBA | GBN_Video
- allow_audio_RADIO | Streaming
- allow_video_RADIO | StreamingVideo
- allow_audio_ITUNES | ItunesPodcast
- allow_video_ITUNES | ItunesPodcastVideo
- allow_audio_SALES | FCBHStore
- allow_video_SALES | FCBHStoreVideo
- allow_audio_DOWNLOAD | Download