# Scripture Burrito 0.1 RFC

**Mark Howe**

**May 28, 2019**

# CONTENTS:

CHAPTER

ONE

# THE ROOT ELEMENT

## 1.1 PROPOSED CHANGES FOR 0.1

- Root element name becomes **BurritoMetadata**, in the null namespace.
- @version must be 0.1
- @id regex should be expanded to include a prefix and to allow other formats of id.
- @revision should be expanded to allow Mercurial and Git commit ids as well as DBL-style positive integers.
- Zero or one id/revision may be placed in the root element, with the others in systemId

## 1.2 ISSUES TO CONSIDER FOR 0.2

None

CHAPTER

TWO

# IDS

## 2.1 PROPOSED CHANGES FOR 0.1

### 2.1.1 Required idServer declarations

These appear as the first children of the root element, eg

```
<idServer prefix="dbl">https://thedigitalbiblelibrary.org</idServer>
<idServer default="true">http://atlantisbibleconsortium.net</idServer>
<idServer prefix="myServer" local="true">http://localhost::8080</idServer>
```

- At least one idServer element is required
- The element must contain either a 'prefix' or a 'default="true"' attribute. (It may contain both, and only one idServer may be the default.) The default, if present, will be assumed to apply to namespaces with no prefix.
- The element may contain a 'local' attribute. When true, this signifies that the ids are for internal use only, and that they should be stripped before export.
- The enclosed text is a URI. In schema this means "pretty much any string", but using URLs that resolve to a server endpoint would help with discoverability.

If there is no default idServer, all ids in the document must be prefixed.

### 2.1.2 ID Syntax

(<prefix>::)?<id>

where

- "prefix" is a NCName (an XML name with no colon)
- "id" matches

```
[0-9A-Za-z]([0-9A-Za-z_-]{0,30}[0-9A-Za-z])?
```

ie a string starting and ending with an alphanumeric character and containing alphanumeric characters, hyphens and underscores.

IDs in this format can be tested for prefixedness (?!) by searching for "::", a substring which seems unlikely to occur in any existing id schemes.

### 2.1.3 Revision Syntax

The non-prefixed ID regex above ought to allow DBL (numeric) and PT/uW (UUID) revision/commit identifiers.

### 2.1.4 Expose User ID

This should happen anywhere that the metadata refers to a person by name. It probably needs to be optional since it may not be possible to recover this information retrospectively.

### 2.1.5 Expose License ids

This would be part of the new license subsection.

## 2.2 ISSUES TO CONSIDER FOR 0.2

None

CHAPTER

THREE

# IDENTIFICATION

## 3.1 TOP-LEVEL FIELDS

### 3.1.1 PROPOSED CHANGES FOR 0.1

#### 3.1.1.1 Replace english/Local fields with elements qualified by language

eg

```
<name lang="en">...</name>
<name lang="fr">...</name>
```

Affected fields include

- name/nameLocal
- description/descriptionLocal
- abbreviation/abbreviationLocal

#### 3.1.1.2 basedOn

This would uniquely identify the snapshot on which the entry is based, which might be from a different ecosystem. This information is potentially useful for forensics. It also provides a mechanism for 3-way diffing of documents when the two deltas are from different ecosystems.

```
<basedOn type="dbl">
   <id>482ddd53705278cc</id>
   <revision>1</revision>
</basedOn>
```

### 3.1.2 ISSUES TO CONSIDER FOR 0.2

#### 3.1.2.1 Evaluate uses of scope

Scope does not have enough options to describle all projects. In addition, it is unclear whether the scope describes the books actually present (impossible with an enum for incremental publishing) or the intended final scope of the project (which is a somewhat existential concept). Something like the canonicalContent section in publications, for a whole entry, would provide scope information in a more flexible and transparent way.

## 3.2 systemId

### 3.2.1 IN DBL METADATA 2.2

**/DBLMetadata/identification/systemId** (Exactly 1)

- **/DBLMetadata/identification/systemId[@type='gbc']** (0 or 1)
  - **id** (Exactly 1 string)
    - The GBC id (24 hex characters)
      - regex: [0-9a-f]{24}
- **/DBLMetadata/identification/systemId[@type='paratext']** (0 or 1)
  - **id** (Exactly 1 string)
    - The paratext id for this entry (40 hex characters)
      - regex: [0-9a-f]{40}
  - **name** (Exactly 1 string)
    - The Name for this ID
      - regex: S.*S
  - **fullName** (Exactly 1 string)
    - The Full Name for this ID
      - regex: S.*S
  - **csetId** (0 or 1 string)
    - The CSet id for this ID
      - regex: S.*S
- **/DBLMetadata/identification/systemId[@type='ptreg']** (0 or 1)
  - **id** (Exactly 1 string)
    - The Paratext Repository id (17 hex characters)
      - regex: [0-9a-zA-Z]{17}
- **/DBLMetadata/identification/systemId[@type='tms']** (0 or 1)
  - **id** (Exactly 1 string)
    - The TMS id for this entry (an UUID)
      - regex: [0-9a-f]{8}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]{12}
- **/DBLMetadata/identification/systemId[@type='reap']** (0 or 1)
  - **id** (Exactly 1 string)
    - The REAP id for this entry (an UUID)
      - regex: [^ ]+
- **/DBLMetadata/identification/systemId[@type='biblica']** (0 or 1)
  - **id** (Exactly 1 integer)
    - The Biblica ID (a number)

        - max: 99999
- **/DBLMetadata/identification/systemId[@type='dbp']** (0 or 1)
  - **id** (Exactly 1 string)
    - The DBP id for this entry (10 hex characters)
      - regex: [A-Z0-9]{10}

### 3.2.2 PROPOSED CHANGES FOR 0.1

#### 3.2.2.1 Add DBL type

This is required to make the document structure orthogonal.

#### 3.2.2.2 Add other known organizations

- Unfolding Word
- Vachan Online

#### 3.2.2.3 Support x-* types

The systemId type mechanism was created when DBL needed to work with a small number of large ecosystems. Future ecosystems may be small – maybe a national denomination or even one church. It may not always make sense to add such organizations to the schema and, when it does, this will take some time. Some architectures involve local servers (on a VPN, an intranet or even localhost), and testing sometimes requires server changes. Supporting types matching

provides a way to introduce new or private ecosystems without rewriting schema:

```
<idServer prefix="mvah">https://markspersonaltranslationproject.fr</idServer>
...
<systemId type="x-mvah">
   <id>idInMyPersonalFormat</id>
   <myDetail>something-that-interests-me</myDetail>
</systemId>
```

The type of all x-* systemIds should correspond to an idServer declaration.

### 3.2.3 ISSUES TO CONSIDER FOR 0.2

None.

## 3.3 canonSpec

### 3.3.1 IN DBL METADATA 2.2

This feature was added as a more flexible and transparent alternative to the scope and tradition values, and as a first step towards hierarchical publication structures. It is based on analysis of the Canons.xml used by Paratext. It is currently not used by Paratext, but is central to Nathanael's workflow.

**/DBLMetadata/identification/canonSpec** (0 or 1)

- **@type** (Exactly 1 string)
  - The overall structure and order of this canon. (OT+ here means canonical and deuterocanonical OT books interleaved within the same section, like most Catholic Bibles)
    - Enum:
      - OT
      - OT+
      - DC
      - NT
      - OT, NT
      - OT+, NT
      - OT, NT, DC
      - OT, DC, NT
- **component** (1 or more string)
  - The components of this canon, which should match the canon type chosen above. eg, if the canon type is "OT, NT", there should be one OT and one NT component here.
    - Enum:
      - armenianApostolicDC
      - armenianApostolicOT
      - armenianApostolicOT2
      - armenianClassicalOT
      - armenianNT
      - catholicAndAnglicanDC
      - catholicLxxDC
      - catholicLxxOT
      - catholicLxxSeparatedDC
      - catholicPlusLutheranDC
      - catholicVulgateDC
      - catholicVulgateOT
      - catholicVulgateSeparatedDC
      - czechKralickaDC
      - danishLutheranDC
      - ethiopianOrthodoxDC
      - ethiopianOrthodoxNT
      - ethiopianOrthodoxOT
      - ethiopianProtestantNT
      - ethiopianProtestantOT

- georgianOrthodoxDC
- georgianOrthodoxOT
- georgianOrthodoxOT2
- georgianSynodalDC
- germanLutheranDC
- greekOrthodoxDC
- greekOrthodoxOT
- kjvDC
- kjvNonDC
- lutheranNT
- romanianOrthodoxDC
- romanianOrthodoxOT
- russianNT
- russianOrthodoxDC
- russianOrthodoxOT
- russianProtestantOT
- russianSynodalDC
- syriacNT
- syriacOT
- tanakhOT
- turkishInterconfessionalDC
- vulgateCatholicBible
- westernInterconfessionalDC
- westernInterconfessionalDC2
- westernNT
- westernOT

### 3.3.2 PROPOSED CHANGES FOR 0.1

#### 3.3.2.1 Remove the .*2 components

These variants of three components correspond to longstanding inconsistencies in the Canons.xml file, caused by inconsistent use of DAN/DAG and EST/ESG in canons of different scope for the same tradition (eg the OT part of the Armenian Bible canon does not match the Armenian OT canon). Also, there is no JER in the Greek Orthodox canon.

### 3.3.3 ISSUES TO CONSIDER FOR 0.2

#### 3.3.3.1 Develop canonSpec

One day, canonSpecs should be able to use custom components, which begs the question of where and how those components would be defined.

CHAPTER

FOUR

# RELATIONSHIPS

## 4.1 IN DBL METADATA 2.2

**/DBLMetadata/relationships** (Exactly 1)

**/DBLMetadata/relationships/relation** (0 or more)

- **@id** (Exactly 1 string key)
  - The DBL id of the related entry
    - regex: [a-f0-9]{16}
- **@revision** (Exactly 1 integer)
  - The revision of the related entry
    - min: 1
- **@relationType** (Exactly 1 string)
  - The role of the related entry with respect to this entry
    - Enum:
      - source
      - expression
- **@type** (Exactly 1 string)
  - The medium of the related entry
    - Enum:
      - text
      - audio
      - print
      - video
      - braille
- **@publicationId** (0 or 1 string)
  - The publication in the related text entry on which to base the braille
  - regex:
    - `[A-Za-z][A-Za-z0-9_\-]{0,31}`

## 4.2 PROPOSED CHANGES FOR 0.1

### 4.2.1 Expand @relationType enum

This mechanism could be used to represent other entry-to-entry relationships, eg between Bible text and related para-biblical material.

## 4.3 ISSUES TO CONSIDER FOR 0.2

There has been some discussion about resource-to-resource relationships. The relationships section probably isn't the best place to address this.

CHAPTER

FIVE

# AGENCIES

## 5.1 IN DBL METADATA 2.2

**/DBLMetadata/agencies** (Exactly 1)

- **/DBLMetadata/agencies/rightsHolder** (1 or more)
  - **uid** (Exactly 1 string key)
    - The id of this rights holder
      - regex: [a-f0-9]{24}
  - **name** (Exactly 1 string)
    - The name in English of this rights holder
      - regex: S.*S
  - **nameLocal** (0 or 1 string)
    - The local name of this rights holder
      - regex: S.*S
  - **abbr** (Exactly 1 string)
    - The abbreviation of this rights holder
      - regex: S.*S
  - **url** (0 or 1 string)
    - The URL of this rights holder
      - regex: S.*S
- **/DBLMetadata/agencies/contributor** (1 or more)
  - **uid** (Exactly 1 string key)
    - The id of this contributor
      - regex: [a-f0-9]{24}
  - **name** (Exactly 1 string)
    - The name of this contributor
      - regex: S.*S
  - **content** (Exactly 1 boolean)
    - Contributes to Content?

  - **finance** (0 or 1 boolean)
    - Contributes to Finance?
  - **management** (0 or 1 boolean)
    - Contributes to Management?
  - **qa** (0 or 1 boolean)
    - Contributes to Quality Assurance?
  - **publication** (0 or 1 boolean)
    - Contributes to publication?
- **/DBLMetadata/agencies/rightsAdmin** (0 or 1)
  - **uid** (Exactly 1 string key)
    - The id of this rights administrator (24 chars of hex)
      - regex: [a-f0-9]{24}
  - **name** (Exactly 1 string)
    - The name of this rights administrator
      - regex: S.*S
  - **url** (0 or 1 string)
    - The URL of this rights administrator
      - regex: S.*S

## 5.2 PROPOSED CHANGES FOR 0.1

### 5.2.1 Multiple *Local elements

- allow multiple instances of nameLocal? (Currently no nameLocal.)

### 5.2.2 Less compulsory fields for upload variant

Some background... DBL Metadata is used as the basis of the job spec at the heart of uploading. In this variant more values are optional - notably @revision since this will be overwritten by the server in any case.

Right now, all the denormalized fields in the agencies section are required. However, the uids in the agencies section may only be changed via the DBL website, ie revisions attempting to change ownership will be rejected. This means that clients need to generate a lot of boilerplate, some of which has to be identical to the information on the server, and some of which is not validated for coherence by the server. So, eg, it is currently possible to change the name and url of a rightsHolder but not the uid, which is a recipe for utter confusion.

The proposal is to make many or maybe all fields optional when revision > 1, to reduce boilerplate and to not create false expectations about what may be changed via the client.

### 5.2.3 contributor/content should be optional

This is the only role-type field that is required for contributors, and was probably a typo in the schema.

## 5.3 ISSUES TO CONSIDER FOR 0.2

Proper support for public licenses may have implications here.

CHAPTER

SIX

# TYPE

## 6.1 IN DBL METADATA 2.2

- **medium** (exactly 1 element)
  - The "type" of the bundle
  - enum:
    - text
    - audio
    - video
    - print
    - braille
- **hasCharacters** (exactly 1 boolean)
  - Does this medium have characters? (Used mainly to control other schema features)
- **isTranslation** (exactly 1 boolean)
  - Is this entry the definitive source, ie does it correspond to a translation project?
- **isExpression** (exactly 1 boolean)
  - Does this entry include one or more publication scenario?
- **isConfidential** (exactly 1 boolean)
  - Is this entry confidential?

For translations only:

- **translationType** (exactly 1 string)
  - The type of translation
  - enum:
    - First
    - New
    - Revision
    - Study / Help Material
- **audience** (exactly 1 string)
  - The audience for whom the translation is intended

  - enum:
    * Basic
    * Common
    * Common - Literary
    * Literary
    * Liturgical
    * Children

For text translations only:

- **projectType** (exactly 1 element)
- enum:
  - Standard
  - Daughter
  - StudyBible
  - StudyBibleAdditions
  - BackTranslation
  - Auxiliary
  - TransliterationManual
  - TransliterationWithEncoder
  - Unknown

For audio expressions only:

- **dramatization** (exactly 1 element)
- enum:
  - Dramatized
  - Non-Dramatized
  - Single-Voice

## 6.2 PROPOSED CHANGES FOR 0.1

### 6.2.1 Private Licences

DBL has a rich model for "private" licenses. In addition to determining whether an entry can be downloaded by a given Library Card Holder, they are also used to filter information such as as publications, cross-references and footnotes. It would be useful to be able to tell if the metadata document has been filtered.

```
<license type="dbl">
    <id>326</id>
    <uri>https://app.thedigitalbiblelibrary.org/api/license/326</uri>
    <serviceOptions>
        <option type="footnotes">false</option>
        <option type="publications">protestant catholic</option>
```

(continues on next page)

(continued from previous page)

```
    </serviceOptions>
</license>
```

### 6.2.2 Public Licenses

DBL doesn't really have a model for this...

### 6.2.3 REAP-compatible isConfidential

Apparently REAP uses more than three states to represent the degree of confidentiality of a project. It would make sense for DBL to use the same system. It will be quite important to make sure that migration does not make previously confidential projects visible.

## 6.3 ISSUES TO CONSIDER FOR 0.2

None.

CHAPTER

SEVEN

# FORMAT - OVERVIEW

The format section is highly medium-dependent. The details for each of the five existing media are provided in subsequent chapters.

## 7.1 PROPOSED CHANGES FOR 0.1

### 7.1.1 Conventions

Conventions are intended to provide a mechanism for subtyping media, allowing greater flexibility along with improved server-side checking and bundle-consumer visibility.

```
<convention type="structure" version="1.0">usx_dirs</convention>
```

Zero or more convention elements would be allowed for each entry. In the absence of a convention, the entry may or not comply, ie *caveat emptor*.

**@type** is one of

- **structure** ie the "directory" structure, eg "usx-dirs"
- **content-format** ie the standard to which specific resources comply, eg "usx3"
- **content** ie the actual content of resources, eg, "usx-refs"

**@version** is required since more complex conventions are likely to evolve.

The enum of conventions will need to be defined in consultation with stakeholders. x-.* should be supported for emerging conventions.

## 7.2 ISSUES TO CONSIDER FOR 0.2

None.

CHAPTER

EIGHT

# TEXT FORMAT

## 8.1 IN DBL METADATA 2.2

- **versedParagraphs** (Exactly 1 boolean)
  - Should the USX be rendered one paragraph per verse?

## 8.2 PROPOSED CHANGES FOR 0.1

### 8.2.1 Remove versedParagraphs

This can be replaced with a convention, which accurately describes the semantics of this field.

### 8.2.2 Syntax

This would provide the USFM and USX version.

```
<usfmVersion>3.0</usfmVersion>
<usxVersion>3.0</usxVersion>
```

## 8.3 ISSUES TO CONSIDER FOR 0.2

None.

CHAPTER

NINE

# AUDIO FORMAT

## 9.1 IN DBL METADATA 2.2

- **compression** (0 or 1 string)
  - The compression method used for the audio files in the entry
    - Enum:
    - MP3
    - FLAC
    - WAV
    - AAC
- **trackConfiguration** (0 or 1 string)
  - The track configuration of the audio files in the entry
    - Enum:
    - 1/0 (Mono)
    - Dual mono
    - 2/0 (Stereo)
    - 5.1 Surround
- **bitRate** (0 or 1 integer)
  - The bit rate of the audio files in the entry
    - min: 1
- **bitDepth** (0 or 1 integer)
  - The bit depth of the audio files in the entry
    - min: 1
- **samplingRate** (0 or 1 integer)
  - The sampling rate of the audio files in the entry
    - min: 1

## 9.2 PROPOSED CHANGES FOR 0.1

### 9.2.1 Conventions

- whole-chapter
- book-dirs

### 9.2.2 Roles

- book-introduction
- audio-timing

## 9.3 ISSUES TO CONSIDER FOR 0.2

None.

CHAPTER

TEN

# VIDEO FORMAT

## 10.1 IN DBL METADATA 2.2

- **container** (Exactly 1 string)
  - The video container (something like 'MP4')
- **/DBLMetadata/format/videoStream** (Exactly 1)
  - **codec** (Exactly 1 string)
    - The codec used to encode the video
  - **bitRate** (Exactly 1 integer)
    - The bit rate of the video
    - min: 1
  - **frameRate** (Exactly 1 decimal)
    - The frame rate of the video
    - min: 1
  - **screenResolution** (Exactly 1 string)
    - The screen resolution of the video (WxH)
    - regex: [1-9][0-9]{0,3}x[1-9][0-9]{0,3}
- **/DBLMetadata/format/audioStream** (0 or 1)
  - **compression** (0 or 1 string)
    - The compression method used for the audio stream of the video in the entry
    - Enum:
      - MP3
      - FLAC
      - WAV
      - AAC
  - **trackConfiguration** (0 or 1 string)
    - The track configuration of the audio stream of the video in the entry
    - Enum:
      - 1/0 (Mono)

        - Dual mono
        - 2/0 (Stereo)
        - 5.1 Surround
    - **bitRate** (0 or 1 integer)
        - The bit rate of the audio stream of the video in the entry
        - min: 1
    - **bitDepth** (0 or 1 integer)
        - The bit depth of the audio stream of the video in the entry
        - min: 1
    - **samplingRate** (0 or 1 integer)
        - The sampling rate of the audio stream of the video in the entry
        - min: 1

## 10.2 PROPOSED CHANGES FOR 0.1

### 10.2.1 Conventions

- whole-chapter
- book-dirs
- roles-in-uris (a Nathanael wizard convention for encoding roles in filenames)

### 10.2.2 Roles for non-canonical video files

It may be possible to reuse some USFM peripheral "roles", and the list will probably need to be extended after consultation with sign language stakeholders.

- bible-menu
- book-menu
- frontmatter
- backmatter
- copyright
- book-introduction

## 10.3 ISSUES TO CONSIDER FOR 0.2

None.

CHAPTER

# ELEVEN

# PRINT FORMAT

## 11.1 IN DBL METADATA 2.2

**pod** (Exactly 1 boolean)

- Is this entry suitable for POD?

**width** (Exactly 1 string)

- The width of the page in millimetres
  - regex: [1-9][0-9]{1,4}mm

**height** (Exactly 1 string)

- The height of the page in millimetres
  - regex: [1-9][0-9]{1,4}mm

**scale** (Exactly 1 string)

- The scale of the page as a percentage
  - regex: [1-9][0-9]{1,3}%

**orientation** (Exactly 1 string)

- The orientation of the pages
  - Enum:
  - portrait
  - landscape

**pageCount** (Exactly 1 integer)

- The number of printed pages
  - min: 1

**color** (Exactly 1 string)

- The color model of the document
  - Enum:
  - RGB
  - CMYK

**/DBLMetadata/format/fonts** (Exactly 1)

- **/DBLMetadata/format/fonts/font** (0 or more)

  - **text()** (Exactly 1 string key)
  - The name of the font
    - regex: .+
  - **@type** (Exactly 1 string)
  - The type of font
    - Enum:
      - Bitmap
      - OpenType
      - TrueType
      - Type1
      - Other

**/DBLMetadata/format/edgeSpace** (Exactly 1)

- **top** (Exactly 1 string)
  - The space at the top of the pages
  - regex: [1-9][0-9]{0,2}mm
- **bottom** (Exactly 1 string)
  - The space at the bottom of the pages
  - regex: [1-9][0-9]{0,2}mm
- **inside** (Exactly 1 string)
  - The space on the inside of each page
  - regex: [1-9][0-9]{0,2}mm
- **outside** (Exactly 1 string)
  - The space on the outside of each page
  - regex: [1-9][0-9]{0,2}mm

## 11.2 PROPOSED CHANGES FOR 0.1

### 11.2.1 Add print-oriented roles

- printBody
- printCover
- printThumbnail

### 11.2.2 Metadata for thumbnail JPEG

- width
- height
- colorModel

### 11.2.3 fonts element should be optional and never empty

This is a schema error, ie it should be

not

### 11.2.4 Support Biblica's Tagged Text Toolbox format

This variant of print content is essentially InDesign XML. We'd need a convention for this, and we might need to rework the metadata fields too.

### 11.2.5 Enforce exactly one publication

Print entries are always an expression and it is hard to imagine a multiple-publication scenario.

## 11.3 ISSUES TO CONSIDER FOR 0.2

None.

CHAPTER

TWELVE

# BRAILLE FORMAT

## 12.1 IN DBL METADATA 2.2

- **isContracted** (Exactly 1 boolean)
  - Does the braille transliteration process include contractions?
- **/DBLMetadata/format/liblouis** (Exactly 1)
  - **version** (Exactly 1 string)
    - The version of Liblouis used for Braille conversion
  - **/DBLMetadata/format/liblouis/table** (Exactly 1)
    - **source** (Exactly 1 string)
    - The URI or other identifier of the table
    - **name** (Exactly 1 string)
    - The human-readable name of the table
- **/DBLMetadata/format/hyphenationDictionary** (0 or 1)
  - **@src** (Exactly 1 string)
    - The URI or other identifier of the dictionary
  - **name** (Exactly 1 string)
    - The human-readable name of the table
- **/DBLMetadata/format/numberSign** (Exactly 1)
  - **character** (Exactly 1 string)
    - The symbol used to prefix numbers
    - regex: ([<braille>])*
  - **useInMargin** (Exactly 1 boolean)
    - Set this field if the number sign should be used before numbers in the margin
- **/DBLMetadata/format/continuousPoetry** (0 or 1)
  - **startIndicator** (0 or 1 string)
    - Character(s) to indicate the start of a section of continuous poetry
    - regex: ([<braille>])*
  - **lineIndicator** (0 or 1 string)

        * Character(s) to indicate the end of a line within a section of continuous poetry
        * regex: ([<braille>])*
    - **lineIndicatorSpaced** (Exactly 1 string)
        * Choose whether line indicator is: never spaced from last word in line; always spaced; or only spaced if the line doesn't end in punctuation
        * Enum:
            · never
            · always
            · sometimes
    - **endIndicator** (0 or 1 string)
        * Character(s) to indicate the end of a section of poetry
        * regex: ([<braille>])*
- **/DBLMetadata/format/content** (Exactly 1)
    - **chapterNumberStyle** (Exactly 1 string)
        * Select "upper" to display chapter numbers normally or "lower" to use lower 4 dots of braille cell
        * Enum:
            · upper
            · lower
    - **chapterHeadingsNumberFirst** (Exactly 1 boolean)
        * Set to display the chapter number before the word for "chapter"
    - **versedParagraphs** (Exactly 1 boolean)
        * Set to display each verse as a new paragraph
    - **verseSeparator** (Exactly 1 string)
        * Character(s) to insert between verses if one verse follows another on the same line
        * regex: ([<braille>])*
    - **includeIntros** (Exactly 1 boolean)
        * Set to display book introductions
    - **/DBLMetadata/format/content/characterStyles** (0 or 1)
        * **emphasizedWord** (0 or 1 string)
        * The symbol used to prefix emphasized words
            · regex: ([<braille>])*
        * **emphasizedPassageStart** (0 or 1 string)
        * The symbol used to start an emphasized passage
            · regex: ([<braille>])*
        * **emphasizedPassageEnd** (0 or 1 string)
        * The symbol used to end an emphasized passage

            - regex: ([<braille>])*
        - **/DBLMetadata/format/content/footnotes** (0 or 1)
            - **callerSymbol** (Exactly 1 string)
            - Character(s) to use to show a footnote
                - regex: ([<braille>])*
        - **/DBLMetadata/format/content/crossReferences** (0 or 1)
            - **callerSymbol** (Exactly 1 string)
            - Character(s) to use to show a cross reference
                - regex: ([<braille>])*
- **/DBLMetadata/format/page** (Exactly 1)
    - **charsPerLine** (Exactly 1 integer)
        - The maximum number of characters on a line
        - min: 1
    - **linesPerPage** (Exactly 1 integer)
        - The maximum number of lines on a page
        - min: 1
    - **defaultMarginWidth** (Exactly 1 integer)
        - The default margin width in number of characters
        - min: 1
    - **versoLastLineBlank** (Exactly 1 boolean)
        - Set to leave last line of verso pages blank
    - **carryLines** (Exactly 1 integer)
        - A heading starting within this number of lines of the bottom of page will be carried over to next page
        - min: 1

## 12.2 PROPOSED CHANGES FOR 0.1

### 12.2.1 liblouis => brailleConvertor

LibLouis is **almost** ubiquitous for braille transcription in 2019, but other options do exist.

<brailleConvertor>liblouis-3.7.1</brailleConvertor>

### 12.2.2 table/source => table/src

This is just for consistency.

### 12.2.3 Enforce exactly one publication

Braille entries are always an expression and it is hard to imagine a multiple-publication scenario.

## 12.3 ISSUES TO CONSIDER FOR 0.2

None.

CHAPTER

THIRTEEN

# LANGUAGES

## 13.1 IN DBL METADATA 2.2

- **iso** (Exactly 1 string)
  - The language's 3-character ISO 639-3 code
    - regex: [a-z][a-z][a-z]
- **name** (Exactly 1 string)
  - The name, in English, of the language
    - regex: S.*S
- **nameLocal** (0 or 1 string)
  - The localized name of the language
    - regex: S.*S
- **scriptCode** (Exactly 1 string)
  - The ISO 15924 script code used in this entry
    - Enum:
    - Adlm
    - Afak
    - Aghb
    - Ahom
    - Arab
    - Aran
    - Armi
    - Armn
    - Avst
    - Bali
    - Bamu
    - Bass
    - Batk
    - Beng

* Blis
* Bopo
* Brah
* Brai
* Bugi
* Buhd
* Cakm
* Cans
* Cari
* Cham
* Cher
* Cirt
* Copt
* Cprt
* Cyrl
* Cyrs
* Deva
* Dsrt
* Dupl
* Egyd
* Egyh
* Egyp
* Elba
* Ethi
* Geok
* Geor
* Glag
* Goth
* Gran
* Grek
* Gujr
* Guru
* Hang
* Hani
* Hano
* Hans

* Hant
* Hatr
* Hebr
* Hira
* Hluw
* Hmng
* Hrkt
* Hung
* Inds
* Ital
* Java
* Jpan
* Jurc
* Kali
* Kana
* Khar
* Khmr
* Khoj
* Kitl
* Kits
* Knda
* Kore
* Kpel
* Kthi
* Lana
* Laoo
* Latn
* Latf
* Latg
* Lepc
* Limb
* Lina
* Linb
* Lisu
* Loma
* Lyci

* Lydi
* Mahj
* Mand
* Mani
* Marc
* Maya
* Mend
* Merc
* Mero
* Mlym
* Modi
* Mong
* Moon
* Mroo
* Mtei
* Mult
* Mymr
* Narb
* Nbat
* Nkgb
* Nkoo
* Nshu
* Ogam
* Olck
* Orkh
* Orya
* Osge
* Osma
* Palm
* Pauc
* Perm
* Phag
* Phli
* Phlp
* Phlv
* Phnx

* Plrd
* Prti
* Rjng
* Roro
* Runr
* Samr
* Sara
* Sarb
* Saur
* Sgnw
* Shaw
* Shrd
* Sidd
* Sind
* Sinh
* Sora
* Sund
* Sylo
* Syrc
* Syre
* Syrj
* Syrn
* Tagb
* Takr
* Tale
* Talu
* Taml
* Tang
* Tavt
* Telu
* Teng
* Tfng
* Tglg
* Thaa
* Thai
* Tibt

        * Tirh
        * Ugar
        * Vaii
        * Visp
        * Wara
        * Wole
        * Xpeo
        * Xsux
        * Yiii
        * Zinh
        * Zmth
        * Zsym
        * Zxxx
        * Zyyy
        * Zzzz
- **script** (Exactly 1 string)
    - The name of the script used in this entry
        * regex: S.*S
- **scriptDirection** (Exactly 1 string)
    - The direction of the script used in this entry
        * Enum:
        * LTR
        * RTL
- **numerals** (0 or 1 string)
    - The numerals system used in this entry
        * Enum:
        * Arabic
        * Bengali
        * Burmese
        * Chinese
        * Cyrillic
        * Devanagari
        * Ethiopic
        * Farsi
        * Gujarati
        * Gurmukhi

        * Hebrew
        * Hindi
        * Kannada
        * Khmer
        * Malayalam
        * Oriya
        * Roman
        * Tamil
        * Telugu
        * Thai
        * Tibetan
- **ldml** (0 or 1 string)
    - The LDML of the language
        * regex: [A-Za-z]{2,3}([-_][A-Za-z0-9]+){0,4}
- **rod** (0 or 1 string)
    - The ROD of the language
        * regex: [0-9]{5}

## 13.2 PROPOSED CHANGES FOR 0.1

### 13.2.1 BCP 47

This is the currently preferred way to store language information, as it can include most other standards and can handle minority languages and dialects that are unlikely to be supported by the older standards such as ISO 639-3.

It seems that the field currently called "ldml" is actually closer to BCP 47. This should be renamed and we should check that it handles all BCP 47 permutations.

We also need to decide whether to store the components of BCP 47 separately, since decomposing BCP 47 is not trivial. (For example, there is no 639-3 code if a 639-1 code exists.) At that point, another option would be to only store the components, which reduces duplication but pushes the onus of constructing BCP 47 onto the consumer.

### 13.2.2 Multiple Language Support

In the current schema there is exactly one language element. To support multiple languages we should adopt a structure similar to the current countries structure, ie a "languages" wrapper with one or more "language" child element. Exactly one of those languages should be marked as the default.

Here, as elsewhere, name and nameLocal need rethinking to support multiple languages. In addition to languages of the scriptural content, there may be localization languages. So, eg, an English-French diglot might contain localization strings in Spanish, Arabic or Hindi in order to localize interfaces.

In a world where English is no longer always the default language, it might make sense to abandon the name/nameLocal distinction altogether and, instead, to require one or more name, each in a specified language. The

downside of this is that every consumer then needs to implement some form of language negotiation to handle the case where, say, the preferred language is French and the options in the metadata are Chinese and Swahili.

## 13.3 ISSUES TO CONSIDER FOR 0.2

None.

CHAPTER

# FOURTEEN

# COUNTRIES

## 14.1 IN DBL METADATA 2.2

- **/DBLMetadata/countries/country** (1 or more)
  - **iso** (Exactly 1 string key)
    - The country's 2-character country code
    - regex: [A-Z][A-Z]
  - **name** (Exactly 1 string)
    - The country's name in English
    - regex: S.*S
  - **nameLocal** (0 or 1 string)
    - The country's localized name
    - regex: S.*S

## 14.2 PROPOSED CHANGES FOR 0.1

### 14.2.1 Multiple Language Support

See the languages section.

## 14.3 ISSUES TO CONSIDER FOR 0.2

None.

CHAPTER

# FIFTEEN

# NAMES

## 15.1 IN DBL METADATA 2.2

- **/DBLMetadata/names/name** (0 or more)
  - **@id** (Exactly 1 string key)
    - The id of this name
    - `regex: [A-Za-z][-A-Za-z0-9_]+`
  - **short** (Exactly 1 string)
    - The short label for this name, which is required and will be used as a default for the other labels if necessary
    - regex: S.(.{0,253}S)?
  - **abbr** (0 or 1 string)
    - The abbreviation for this name
    - regex: S(.{0,253}S)?
  - **long** (0 or 1 string)
    - The long label for this name
    - regex: S(.{0,1022}S)?

## 15.2 PROPOSED CHANGES FOR 0.1

This might mean allowing multiple abbr, short and long, distinguished by language attribute.

## 15.3 ISSUES TO CONSIDER FOR 0.2

None

CHAPTER

SIXTEEN

# MANIFEST

## 16.1 IN DBL METADATA 2.2

- **manifest** (exactly 1 string)
  - contains one or more **container** or **resource**
- **container**
  - **@uri** (exactly 1 string)
    - The URI of this container, relative to any uris specified in ancestor containers
    - type xsd:anyUri
  - contains one or more **container** or **resource**
- **resource**
  - **@uri** (exactly 1 string)
    - The URI of this container, relative to any uris specified in ancestor containers
    - type xsd:anyUri
  - **size** (exactly one positive integer)
    - The size of the resource in bytes
  - **checksum** (0 or 1 string)
    - The MD5 checksum of the resource
    - `regex: [a-f0-9]{32}(\-\d+)?"`
  - **mimeType** (0 or 1 string)
    - The mimetype of the resource
    - `regex: [\-a-z0-9]+/[\-a-z0-9+]+"`
  - **progress** (0 or 1 positive integer)
    - An alternative, unused way to track progress in DBL
    - range: 1-999

## 16.2 PROPOSED CHANGES FOR 0.1

### 16.2.1 Drop progress attribute

### 16.2.2 Tighten checksum regex

The regex currently allows S3 part suffixes which should never have been present in any metadata.

### 16.2.3 Consider dropping containers

Containers provide a hierarchical alternative to a flat list of resources with fully-qualified uris. The flat version is easier for most processing scenarios, and both the DBL server and Nathanael coerce the manifest data into this format. Supporting containers means that all compliant technologies need to be able to handle both the flat and hierarchical variants. However, the hierarchical form is much easier for humans to create and maintain.

## 16.3 ISSUES TO CONSIDER FOR 0.2

None

CHAPTER

# SEVENTEEN

# PUBLICATIONS

## 17.1 IN DBL METADATA 2.2

**/DBLMetadata/publications** (Exactly 1)

- **/DBLMetadata/publications/publication** (1 or more)
  - **@id** (Exactly 1 string key)
    - The id of the publication
    - `regex: [A-Za-z][A-Za-z0-9_\-]{0,31}`
  - **name** (0 or 1 string)
    - The publication's name, in English
    - regex: S.*S
  - **nameLocal** (0 or 1 string)
    - The publication's localized name
    - regex: S.*S
  - **abbreviation** (0 or 1 string)
    - The publication's abbreviation, in English (no exotic characters)
    - regex: [-A-Za-z0-9]{2,12}
  - **abbreviationLocal** (0 or 1 string)
    - The publication's localized abbreviation
    - regex: S.{0,10}S
  - **description** (0 or 1 string)
    - The publication's description, in English
    - regex: S.*S
  - **descriptionLocal** (0 or 1 string)
    - The publication's localized description
    - regex: S.*S
  - **scope** (0 or 1 string)
    - The publication's scope

        * Enum:
            · Bible
            · Bible with Deuterocanon
            · New Testament
            · New Testament+
            · Old Testament
            · Old Testament + Deuterocanon
            · Old Testament+
            · Portions
            · Selections
            · Shorter Bible

Each publication must include:

- **canonicalContent** (exactly 1 element)
  - **book** (1 or more element)
    * **@code**
      · The USFM book code
- **structure** (exactly 1 element)
  - Contains at least one division and/or content element
  - ** division**
    * A “section” (eg “OT” or “Pentateuch”)
    * **@name** (exactly 1 string)
      · A name id for this division
    * **@role** (0 or 1 string)
      · The canonical or peripheral function of the division

```
((GEN|EXO|LEV|NUM|DEU|JOS|JDG|RUT|1SA|2SA|1KI|2KI|1CH|2CH|EZR|
NEH|EST|JOB|PSA|PRO|ECC|SNG|ISA|JER|LAM|EZK|DAN|HOS|JOL|AMO|
OBA|JON|MIC|NAM|HAB|ZEP|HAG|ZEC|MAL|MAT|MRK|LUK|JHN|ACT|ROM|
1CO|2CO|GAL|EPH|PHP|COL|1TH|2TH|1TI|2TI|TIT|PHM|HEB|JAS|1PE|2PE
1JN|2JN|3JN|JUD|REV|TOB|JDT|ESG|WIS|SIR|BAR|LJE|S3Y|SUS|BEL|
1MA|2MA|3MA|4MA|1ES|2ES|MAN|PS2|ODA|PSS|JSA|JDB|TBS|SST|DNT|BLT|
EZA|5EZ|6EZ|DAG|PS3|2BA|LBA|JUB|ENO|1MQ|2MQ|3MQ|REP|4BA|LAO)
(( [1-9][0-9]{0,2}(-[1-9][0-9]{0,2})?((,([1-9][0-9]{0,2}
(-[1-9][0-9]{0,2})?))*))|
( [1-9][0-9]{0,2}:[1-9][0-9]{0,2}(-[1-9][0-9]{0,2}(:[1-9][0-9]{0,2})?)?
(,(([1-9][0-9]{0,2}(-[1-9][0-9]{0,2})?)|
([1-9][0-9]{0,2}:[1-9][0-9]{0,2}(-[1-9][0-9]{0,2}
(:[1-9][0-9]{0,2})?)?)))*))?)(;
(((GEN|EXO|LEV|NUM|DEU|JOS|JDG|RUT|1SA|2SA|1KI|2KI|1CH|2CH|EZR|
NEH|EST|JOB|PSA|PRO|ECC|SNG|ISA|JER|LAM|EZK|DAN|HOS|JOL|AMO|
OBA|JON|MIC|NAM|HAB|ZEP|HAG|ZEC|MAL|MAT|MRK|LUK|JHN|ACT|ROM|
1CO|2CO|GAL|EPH|PHP|COL|1TH|2TH|1TI|2TI|TIT|PHM|HEB|JAS|1PE|2PE|
1JN|2JN|3JN|JUD|REV|TOB|JDT|ESG|WIS|SIR|BAR|LJE|S3Y|SUS|BEL|
```

(continues on next page)

(continued from previous page)

```
1MA|2MA|3MA|4MA|1ES|2ES|MAN|PS2|ODA|PSS|JSA|JDB|TBS|SST|DNT|BLT|
EZA|5EZ|6EZ|DAG|PS3|2BA|LBA|JUB|ENO|1MQ|2MQ|3MQ|REP|4BA|LAO)
(( [1-9][0-9]{0,2}(-[1-9][0-9]{0,2})?((,([1-9][0-9]{0,2}
(-[1-9][0-9]{0,2})?))*))|
( [1-9][0-9]{0,2}:[1-9][0-9]{0,2}(-[1-9][0-9]{0,2}(:[1-9][0-9]{0,2})?)?
(,(([1-9][0-9]{0,2}(-[1-9][0-9]{0,2})?)|
([1-9][0-9]{0,2}:[1-9][0-9]{0,2}(-[1-9][0-9]{0,2}
(:[1-9][0-9]{0,2})?)?)))*))?)))*" }
```

```
"sourceZip" | "sourceFile" | "lds" | "ldml" | "styles" | "vrs"
```

```
Regex: X-[A-Za-z0-9][A-Za-z0-9\-]*
```

```
"XXA" | # Extra A, e.g. a hymnal
"XXB" | # Extra B
"XXC" | # Extra C
"XXD" | # Extra D
"XXE" | # Extra E
"XXF" | # Extra F
"XXG" | # Extra G
"FRT" | # Front Matter
"BAK" | # Back Matter
"OTH" | # Other Matter
"INT" | # Introduction
"CNC" | # Concordance
"GLO" | # Glossary
"TDX" | # Topical Index
"NDX"   # Names Index
```

- **content**
  - **@src** (exactly 1 string)
    - The URI of the manifest entry corresponding to this content

```
[A-Za-z0-9][\-A-Za-z0-9_\.]+(/[A-Za-z0-9][\-A-Za-z0-9_\.]+)*
```

  - **@srcPart** (0 or 1 string)
    - Optional in-document identifier

```
regex: (((zip|xpath)://)|#).+
```

  - **@name** (0 or 1 string)
    - A name id for this division
  - **@role** (0 or 1 string)
    - The canonical or peripheral function of the division

Each publication may also include

- **/DBLMetadata/publications/publication/countries** (0 or 1)
  - **/DBLMetadata/publications/publication/countries/country** (1 or more)
  - same format as /DBLMetadata/countries/country
- **/DBLMetadata/publications/publication/canonSpec** (0 or 1)

- same format as /DBLMetadata/identification/canonSpec

## 17.2 PROPOSED CHANGES FOR 0.1

### 17.2.1 Multiple Language Support

See the languages section.

### 17.2.2 Consider dropping scope

An enum can never accurately capture all the possible permutations of content, and the options that are currently present are routinely abused by archivists. Statistics should be based on the canonicalContent which should correspond to what is actually present in the Burrito.

### 17.2.3 metaContent

This would allow content elements to have child elements for supporting content. The first concrete use case is for timing files, which are closely related to audio or video files, but which appear as separate entries within the manifest:

```
<content src="MAT.usx" name="book-mat" role="MAT">
    <metaContent src="timing/MAT.xml"/>
</content>
<content src="MRK.usx" name="book-mrk" role="MRK">
    <metaContent src="timing/MRK_1-6.xml" role="MRK 1-6"/>
    <metaContent src="timing/MRK_7-16.xml" role="MRK 7-16"/>
</content>
```

### 17.2.4 Peripherals ids in metadata role enum

This enables the tagging of extra-canonical content without relying on well-known file names. The list, from the USFM 3 spec, would be

- **abbreviations**: Table of abbreviations
- **alphacontents**: Alphanumeric Contents
- **chron**: Chronology
- **cnc**: Concordance
- **contents**: Table of Contents
- **cover**: Cover
- **foreword**: Foreword
- **glo**: Glossary
- **halftitle**: Half Title Page
- **imprimatur**: Imprimatur
- **intbible**: Introduction to the Bible
- **intdc**: Deuterocanon Introduction

- **intepistles**: Introduction to Epistles
- **intgospels**: Introduction to Gospels
- **inthist**: Introduction to History
- **intnt**: Introduction to New Testament
- **intot**: Introduction to the Old Testament
- **intpent**: Introduction to the Pentateuch
- **intpoetry**: Introduction to Poetry
- **intprophesy**: Introduction to Prophecy
- **lxxquotes**: Quotes from LXX in NT
- **maps**: Map Index
- **measures**: Weights and Measures
- **ndx**: Names Index
- **preface**: Preface
- **promo**: Promotional Page
- **pubdata**: Publication Data
- **spine**: Spine
- **tdx**: Topical Index
- **title**: Title Page

This enum may well need extending and adapting in order to meet the needs of all media, notably sign language video.

## 17.3 ISSUES TO CONSIDER FOR 0.2

None

CHAPTER

EIGHTEEN

# SOURCE

## 18.1 IN DBL METADATA 2.2

- **/DBLMetadata/source** (exactly 1 element)
  - Must contain canonicalContent and structure
- **canonicalContent** (0 or 1 element)
  - **book** (1 or more element)
    - **@code**
      - The USFM book code
- **structure** (0 or 1 element)
  - Contains at least one division and/or content element
  - ** division**
    - A “section” (eg “OT” or “Pentateuch”)
    - **@src**
      - The URI of the manifest entry corresponding to this content
      - ```
        [A-Za-z0-9][\-A-Za-z0-9_\.]+(/[A-Za-z0-9][\-A-Za-z0-9_\.]+)*
        ```
    - **@srcPart** (0 or 1 string)
      - Optional in-document identifier
      - ```
        regex: (((zip|xpath)://)|#).+
        ```
    - **@role**
      - The canonical or peripheral function of the division
      - ```
        ((GEN|EXO|LEV|NUM|DEU|JOS|JDG|RUT|1SA|2SA|1KI|2KI|1CH|2CH|EZR|
        NEH|EST|JOB|PSA|PRO|ECC|SNG|ISA|JER|LAM|EZK|DAN|HOS|JOL|AMO|
        OBA|JON|MIC|NAM|HAB|ZEP|HAG|ZEC|MAL|MAT|MRK|LUK|JHN|ACT|ROM|
        1CO|2CO|GAL|EPH|PHP|COL|1TH|2TH|1TI|2TI|TIT|PHM|HEB|JAS|1PE|2PE
        1JN|2JN|3JN|JUD|REV|TOB|JDT|ESG|WIS|SIR|BAR|LJE|S3Y|SUS|BEL|
        1MA|2MA|3MA|4MA|1ES|2ES|MAN|PS2|ODA|PSS|JSA|JDB|TBS|SST|DNT|BLT|
        EZA|5EZ|6EZ|DAG|PS3|2BA|LBA|JUB|ENO|1MQ|2MQ|3MQ|REP|4BA|LAO)
        (( [1-9][0-9]{0,2}(-[1-9][0-9]{0,2})?((,([1-9][0-9]{0,2}
        (-[1-9][0-9]{0,2})?))*))|
        ```

(continues on next page)
        ```

(continued from previous page)

```
( [1-9][0-9]{0,2}:[1-9][0-9]{0,2}(-[1-9][0-9]{0,2}(:[1-9][0-9]{0,2})?)?
(,(([1-9][0-9]{0,2}(-[1-9][0-9]{0,2})?)|
([1-9][0-9]{0,2}:[1-9][0-9]{0,2}(-[1-9][0-9]{0,2}
(:[1-9][0-9]{0,2})?)?))))*))?)(;
(((GEN|EXO|LEV|NUM|DEU|JOS|JDG|RUT|1SA|2SA|1KI|2KI|1CH|2CH|EZR|
NEH|EST|JOB|PSA|PRO|ECC|SNG|ISA|JER|LAM|EZK|DAN|HOS|JOL|AMO|
OBA|JON|MIC|NAM|HAB|ZEP|HAG|ZEC|MAL|MAT|MRK|LUK|JHN|ACT|ROM|
1CO|2CO|GAL|EPH|PHP|COL|1TH|2TH|1TI|2TI|TIT|PHM|HEB|JAS|1PE|2PE|
1JN|2JN|3JN|JUD|REV|TOB|JDT|ESG|WIS|SIR|BAR|LJE|S3Y|SUS|BEL|
1MA|2MA|3MA|4MA|1ES|2ES|MAN|PS2|ODA|PSS|JSA|JDB|TBS|SST|DNT|BLT|
EZA|5EZ|6EZ|DAG|PS3|2BA|LBA|JUB|ENO|1MQ|2MQ|3MQ|REP|4BA|LAO)
(( [1-9][0-9]{0,2}(-[1-9][0-9]{0,2})?((,([1-9][0-9]{0,2}
(-[1-9][0-9]{0,2})?))*))|
( [1-9][0-9]{0,2}:[1-9][0-9]{0,2}(-[1-9][0-9]{0,2}(:[1-9][0-9]{0,2})?)?
(,(([1-9][0-9]{0,2}(-[1-9][0-9]{0,2})?)|
([1-9][0-9]{0,2}:[1-9][0-9]{0,2}(-[1-9][0-9]{0,2}
(:[1-9][0-9]{0,2})?)?))))*))?))*" }
```

- `"sourceZip" | "sourceFile" | "lds" | "ldml" | "styles" | "vrs"`

- `Regex: X-[A-Za-z0-9][A-Za-z0-9\-]*`

```
"XXA" | # Extra A, e.g. a hymnal
"XXB" | # Extra B
"XXC" | # Extra C
"XXD" | # Extra D
"XXE" | # Extra E
"XXF" | # Extra F
"XXG" | # Extra G
"FRT" | # Front Matter
"BAK" | # Back Matter
"OTH" | # Other Matter
"INT" | # Introduction
"CNC" | # Concordance
"GLO" | # Glossary
"TDX" | # Topical Index
"NDX"   # Names Index
```

- **content**
  - **@src** (exactly 1 string)
    - The URI of the manifest entry corresponding to this content
    - `[A-Za-z0-9][\-A-Za-z0-9_\.]+(/[A-Za-z0-9][\-A-Za-z0-9_\.]+)*`
  - **@name** (0 or 1 string)
    - A name id for this division
  - **@role** (0 or 1 string)
    - The canonical or peripheral function of the division

## 18.2 PROPOSED CHANGES FOR 0.1

### 18.2.1 Consider dropping @name

### 18.2.2 Consider unpacking PT sources

## 18.3 ISSUES TO CONSIDER FOR 0.2

None

CHAPTER

NINETEEN

# COPYRIGHT

## 19.1 IN DBL METADATA 2.2

This section can contain long and/or short versions of the copyright statement, either of which may be in plain text and/or xhtml. (Historically, most DBL entries have the long version in XHTML.)

**/DBLMetadata/copyright** (Exactly 1)

- **/DBLMetadata/copyright/fullStatement** (0 or 1)
  - **statementContent[@type='xhtml']/*** (0 or 1 xml)
    - The copyright statement in DBL's subset of XHTML (must be valid XML, ie tags must match.)
      - schema: dbl/2_2/dbl-xhtml
  - **statementContent[@type='plain']** (0 or 1 string)
    - The copyright statement in plain text

**/DBLMetadata/copyright/shortStatement** (0 or 1)

- **statementContent[@type='xhtml']** (0 or 1 xml)
  - The copyright statement in DBL's subset of XHTML (must be valid XML, ie tags must match.)
    - schema: dbl-xhtml
- **statementContent[@type='plain']** (0 or 1 string)
  - The copyright statement in plain text

## 19.2 PROPOSED CHANGES FOR 0.1

### 19.2.1 Language attribute for statementContent

To support multiple languages, we would need to add a language attribute (or subelement) to each statement. This could be optional if there is only one language in the languages section.

## 19.3 ISSUES TO CONSIDER FOR 0.2

Proper support for public licenses may have implications here.

CHAPTER

TWENTY

# PROMOTION

## 20.1 IN DBL METADATA 2.2

**/DBLMetadata/promotion** (Exactly 1)

- **/DBLMetadata/promotion/promoVersionInfo** (0 or 1)
  - **./*** (0 or 1 xml)
    - Promotional material in DBL's subset of XHTML (must be valid XML, ie tags must match.)
      - schema: dbl/2_2/dbl-xhtml

## 20.2 PROPOSED CHANGES FOR 0.1

### 20.2.1 Replace promoVersionInfo with statementContent

Right now the promotion section is similar to but confusingly different to the copyright section. A more coherent structure that also allows plain text promotional material would be

**/DBLMetadata/promotion** (Exactly 1)

- **statementContent[@type='xhtml']/*** (0 or 1 xml)
  - Promotional material in DBL's subset of XHTML (must be valid XML, ie tags must match.)
  - schema: dbl/2_2/dbl-xhtml
- **statementContent[@type='plain']** (0 or 1 string)
  - Promotional material in plain text

## 20.3 ISSUES TO CONSIDER FOR 0.2

None.

CHAPTER

# TWENTYONE

# PROGRESS

## 21.1 IN DBL METADATA 2.2

DBL Metadata currently supports two mechanisms for tracking translation progress. The first, which is supported by PT, uses a top-level section to list the progress for each book:

```
<progress>
    <book code="GEN" stage="4"/>
    <book code="EXO" stage="1"/>
    <book code="JOS" stage="2"/>
    <book code="LUK" stage="4"/>
</progress>
```

There are two issues with this:

- It means another potentially long list in the metadata
- More importantly, it can only record progress for books when, in reality, progress on introductions and other para-canonical content may also be important.

The alternative mechanism, which is defined in the schema, but which has probably never been used, is to record the progress against manifest entries.

```
<resource checksum="0e6c24ebcf1ca2e928578ab239b69687" mimeType="application/xml" size=
↪"296803" uri="release/USX_2/1CH.usx" progress=”37”/>
```

Progress can therefore be logged against any document in the entry, without bloating the metadata document. One possible argument against this approach is that project tracking and manifest information may be generated by very different routes. Also, PT currently duplicates most canonical content several times when multiple booklists are specified (but maybe we should fix the duplication of content).

## 21.2 PROPOSED CHANGES FOR 0.1

We should pick one of these options, or come up with a new one, ensure that it will be supported by Paratext, and remove the unused options.

## 21.3 ISSUES TO CONSIDER FOR 0.2

None.