

FHMWG Tech Discussion 1

Dr. Luther Tychonievich
FHMWG Board
University of Virginia

Outline

- Stage 1 goals
- Stage 1 process
- Common threads
- Specific recommendations
- Open questions

Stage 1 Goals

- Use existing metadata standards
- Document core metadata fields
- Clarify use in Family History space

Stage 1 Summary

Topic	Recommendation
Album	Name & ID of albums image is in
Caption	Free-text catch-all
Date	Datetime of depicted scene
Event	Free-text description
Location	GPS and free-text
Person	Name and image region
Object	Title and image region

- Sources: option suited to genealogy not found
- Rights: IPTC just text, PLUS too complex

Stage 1 Process

- Board identifies key areas
- Teams review options, draft recommendations
- **Technical feedback**
- Revised recommendation
- Stakeholder feedback
- Release

Common Threads

- Piecemeal support OK
 - ... as long as unsupported parts not removed
- One place to write
 - Described in detail
 - If present, read this too
- Optional places to read
 - Mentioned with links
 - Can be used as fallbacks
- Write XMP
 - ... each team's conclusion; IPTC IIM, EXIF also reviewed

Albums

- From `metadatabworkinggroup.com` - now defunct
- Each image has `mwg-coll:Collections` containing any number of albums it belongs to
- Each album has a name and/or URI
- Not stage 1:
 - Ordinal in album, album nesting, album metadata

Caption

- From `dublincore.org`, endorsed by XMP and IPTC
- Each image has a `dc:description`
- Not stage 1:
 - Text markup, links to other metadata, caption metadata

Date

- From `http://ns.adobe.com/photoshop/`, endorsed by IPTC
- Each image has a `photoshop:DateCreated` with date of the depicted scene
- Not stage 1:
 - Approximate dates, date intervals, dates BCE
 - Dates of any other image-related event

Event

- From IPTC Extension schema
- Each image has a `Iptc4xmpExt:Event` with an description
- Not stage 1:
 - Event metadata

Location

- From IPTC Extension schema, referencing EXIF
- Each image has a `Iptc4xmpExt:LocationShown` containing any number of locations
- Each location has a GPS coordinate and/or name and/or URI
- Not stage 1:
 - Jurisdiction hierarchy (City/State etc inadequate)
 - GPS regions and/or precision

Objects and People

- Decision: use IPTC 2019.1 image regions
- Limit complexity:
 - Require regions (`Iptc4xmpExt:ImageRegion`)
 - One person or object per region
 - `Iptc4xmpExt:PersonInImageWDetails` with name, description, and/or URI
 - `Iptc4xmpExt:ArtworkOrObject` with title
- Use “whole-image” region if region unknown

Objects and People

- Not stage 1:
 - Other person metadata
 - Recommendation re non-person non-objects, like animals
 - Distinction between “unknown location” and “relevant but not depicted”

Outline

- Stage 1 goals
- Stage 1 process
- Common threads
- Specific recommendations
- **Open questions**

Use of URI

- Albums and People both admit URI
 - Unique identifier to allow matching
 - URL to allow linking to DB
- Should we recommend using these?
 - Pro: more functionality
 - Con: possibility of malicious links
 - Possible: sanitize to hash-based UUID
 - e.g., `https://www.familysearch.org/tree/person/LC32-HZ6`
becomes `urn:uuid:da3fac6c-a2e5-59d7-936b-1124707fd3ef`

One or Several?

- Can tools handle one image with
 - multiple image regions
 - multiple people per region
 - multiple locations
 - multiple albums
 - multiple languages (see next slide)
- If some cannot, should we recommend
 - store only one
 - first/last takes precedence
 - custom “primary” attribute

Handling AltLang

- IPTC text fields have AltLang values:

- Multiple (language, text) pairs; e.g.

```
<rdf:Alt>
```

```
  <rdf:li xml:lang='en'>Gordon Clarke</rdf:li>
```

```
  <rdf:li xml:lang='ja-Latn'>Kurāku Gōdon</rdf:li>
```

```
</rdf:Alt>
```

- Should we recommend
 - Full support for AltLang
 - Handling one language per field
 - Handling one language per image file

Free-form vs canonical RDF

- XMP is encoded as RDF/XML
 - can represent arbitrary property graph
 - many ways to serialize same data
- All common XMP vocabularies are trees
 - Tree-structured RDF's XML has nested-element normalized form
- Do we say “use nested form” or “use RDF toolchain”?
 - Nested: much less code to implement
 - RDF: guaranteed to read any tool's output

How much should we standardize?

- Handling secondary fields
 - “shall read these other fields”
 - “shall keep these 2 fields in sync”
 - “if main field missing, default to this other”
 - “...with this conversion algorithm”
- UI guidelines
 - “call it this”
 - “...in each of these human languages”
 - “present it in this order”
- Techniques for censoring/sanitizing/dropping metadata

Stage 1 Summary

Topic	Recommendation
Album	Name & ID of albums image is in
Caption	Free-text catch-all
Date	Datetime of depicted scene
Event	Free-text description
Location	GPS and free-text
Person	Name and image region
Object	Title and image region

- Stage 2: Sources, Rights
- Questions: URI, Multiplicity, AltLang, RDF, secondary fields, UI/UX, sanitizing