

# ODF Metadata and Interoperability

Rob Weir  
IBM  
2011-07-14, Berlin

<http://www.robweir.com>  
Twitter: @rcweir  
[rob@robweir.com](mailto:rob@robweir.com)

# Basic questions

- 1 What can we say about behavior?
- 2 What do we do about metadata we don't recognize?
- 3 Best practices for interoperable metadata
- 4 A metadata editing constraint language?

# Skin, but no bones

- By itself, ODF 1.2 metadata does not do anything
- It is a framework that can represent anything, but it means nothing.
  - Unless we agree on what it means
  - Extensibility versus Interoperability
- Are there a handful of commonly used vocabularies that we should try to implement broadly?

# Some hypothetical candidates

- FOAF: a person and their social network
- vCard: electronic business card
- iCalendar/RDF: events, alarms, to-dos
- ccRel: Creative Commons Rights Expression Language
- SCOVO: Statistical Core Vocabulary
- BasicGeo: geotagging

# How to make these interop?

- 1 Agree on a subset of vocabularies to support
- 2 Agree how the can be used in ODF.
- 3 Document this agreement as a “profile standard” in OASIS, e.g., “A Profile of ODF 1.2 + ccRel” or in ODF 1.3
- 4 Define not just encoding, but interactions with editing operations.
- 5 Test interop at Plugfests

# Unknown Metadata

- Your editor loads a document that contains metadata that you do not understand. What do you do?
  - Do you preserve it?
  - Do you allow it to be edited?
  - Do you allow it to be cut & paste?
  - Across documents?

# The Hidden Constraint Problem

- Metadata can have implicit constraints that are easily broken by editing
- When you have completed the required paper work, please submit the documentation to Roberto González for further processing

  - What happens when I copy that text to another document? To a document that already has an approver? What happens if I edit the document and change the name (but not the email address, which is hidden)?

# Examples of Hidden Constraints

- **Cardinality.** The given metadata attribute can only occur 1x, 2x, or N-times in a document.
- **Referential integrity.** The given metadata attribute must be consistent with another metadata attribute or with some other document content.
- **Volatility.** The metadata is based on an editing timestamp, transaction ID or a random number.
- **Security.** Should I be digitally signing a document that contains metadata that I cannot see?

# “Edit Safe” Metadata

- Behaves like a visual attribute:
  - `<p>This is an example of bold text.</p>`
  - Can be cut and pasted, in total or in part
  - Can be inserted into
  - Can be split
  - No dependency between attribute and content
  - Can be copied into another document with metadata preserved.
  - More thinking needed to define what “Safe” means.
- Go beyond this and metadata is not interoperable except by agreement.
- Net effect is that interoperable metadata is quite weak unless we agree on some specific vocabularies.
- Semantic web doesn't have this problem, since the web, for the most part, is not editable.

# Declarable Editing Constraints

- What if metadata declared its constraints with regards to cardinality, etc.?
- A set of atomic constraints, like:
  - no-copy, no-duplicate, no-split, no-insert, etc.
- Then editors, even if they do not understand the underlying semantics of the metadata can do a better job of preserving the metadata and constraining the editing operations where necessary.
- This might be a reasonable standard, useful not only for ODF but for other editable formats.

# Toward Interoperable Metadata

- If there is a widely adopted vocabulary that already exists, adopt that rather than reinventing.
  - Work with the “network effect”, not against it.
- See if we can agree on a set of generally useful and mature vocabularies and define their use in ODF via a Profile Standard
- If you must invent your own vocabularies (and if you don't, your customers surely will) then avoid hidden constraints, use “edit safe” metadata.
- Investigate possibilities of defining an editing constraint vocabulary for expressing these concerns in a generic way.