

OpenDocument Metadata Use Cases and Requirements

September 25, 2006

editor

Bruce D'Arcus

contributors

Patrick Durusau

Gary Edwards

John Madden

Florian Reuter

Svante Schubert

Rob Weir

1 Introduction

While existing OpenDocument metadata offers support for the widely-used Dublin Core vocabulary, it is limited in the following ways:

1. It provides no rules for extension.
2. It only allows descriptions of the document as a whole.

This document was prepared by the OpenDocument Metadata Subcommittee, and includes representative use cases and requirements for enhancing this metadata support.

1.1 *What is Metadata?*

The SVG metadata specification defines metadata as “structured data about data.” This structured data “describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource” [NISO-META]. It is the sort of content one typically uses to describe documents, images, contacts, events, and so forth.

1.2 *Why is Metadata Important?*

Metadata allows the document author to encode more of his/her human intelligence into the document. Document metadata in electronic documents is the natural evolution of the marginal comment or the footnote in paper documents. Moreover, this enhanced metadata can also enable powerful automated solutions. When customers and developers ask for support for custom content functionality, then, it is often in order to provide domain-specific metadata support.

2 Use Cases

The following use cases have been consolidated from a more comprehensive list collected by the OpenDocument Metadata Subcommittee.

2.1 *Accessibility*

Enhanced metadata about document content can be particularly useful for accessibility purposes. This use case imagines a blind user who is having software read back an ODF document to them. Upon getting to an image, the software could optionally present descriptive information about the image, such as its creator, title, and so forth. The same would apply to any included metadata descriptions (see below).

2.2 *Bibliographies and Citations*

Citations can be thought of as dynamic fields whose formatting is generated from separate metadata, based on the requirements of a given citation style. A common use case is collaborative editing, where each author may use a different editing application and/or bibliographic database. ODF metadata should make it possible for this sort of collaborative roundtrip editing scenario to work, and for the citation fields to remain "live" across applications. This might involve also the ability for database applications to store their own application-specific metadata along with the more generic bibliographic metadata.

2.3 *Content Tagging*

User wants to add richer metadata to in-document content. In the simple case, they highlight text and tag it with some term. In a more complex case, they associate more structured metadata with content; for example a contact record for a participant in a meeting. A similar example might involve a doctor who wishes to record diagnostic information about a patient in a structured way. In this case, they might invoke a contextual menu item that brings up a custom inline field that allows them to choose the patient and the diagnosis. The displayed text is then human readable, but

behind it lies more structured metadata that allows a user to access additional information about the patient, or for tools to later extract the information for other purposes.

2.4 *Enhanced Search*

Effective document search requires richer description than simply plain-text content. While basic default metadata such as title or creator can be helpful, there is need for the possibility of richer, custom, metadata. A legal office may want to include case information about a document, information about the plaintiffs, and so forth, all of which can facilitate enhanced search. They may also want to track relations among documents; for example, that between a final version and its draft original.

2.5 *Workflow*

Workflow and collaboration solutions depend on portable metadata embedded in files. This may include information about who edited a document, its status, the rights associated with its use, its relationships to other documents or to a collection of documents, and so forth.

3 Design Goals and Principles

The following design goals and principles have been extracted from the use cases, and are used to frame the requirements.

3.1 *Balance Expressiveness and Ease-of-Implementation*

Enhanced metadata support should allow for complex metadata description where necessary, but should be designed with ease-of-implementation in mind for both third-party and primary application developers.

3.2 *Enable Extensibility*

Enhanced metadata should provide a strategic innovation opportunity for OpenDocument by allowing developers to build compelling solutions independent of the OpenDocument standardization process.

3.3 *Ensure Interoperability*

Enhanced metadata should provide an infrastructure that assures round-tripping interoperability among compliant OpenDocument applications.

4 Requirements

The OpenDocument enhanced metadata proposal shall address the following requirements:

4.1 *Standard Model*

Beyond standard core support, there should be clear rules for custom extension metadata. These rules must be based on a clearly specified formal model, such that metadata extensions:

1. may be at least displayed without any particular programming logic
2. may reference other resource descriptions¹, including across partitions of user functionality (a document description or bibliographic citation, say, to reference a contact record)

This model may involve adopting all or part of [RDF](#) for a general extensible description framework, and may include compatibility or interoperability with similar RDF-based metadata formats such as [RSS 1.0](#), [XHTML 2.0 metainformation](#) or [XMP](#).

This requirement facilitates the following goals: Enable Extensibility and Ensure Interoperability.

4.2 *Default Vocabularies*

OpenDocument should provide a rich set of default properties and classes for describing resources that go beyond the current support. The proposal should seek to reuse or define mappings to existing vocabularies, which may include deepening support for [Dublin Core](#) and/or adding support for [vCard](#) and/or [Creative Commons](#) to describe specific content of general applicability to office documents.

This requirement facilitates the following goals: Balance Expressiveness and Ease-of-Implementation and Ensure Interoperability.

4.3 *Mechanism to Associate Content with Metadata*

Proposal should define valid document objects that can be described using enhanced metadata (for example, words, sentences, paragraphs, tables, pictures, etc.), as well as a mechanism to associate those objects with their descriptions.

This requirement facilitates the following goals: Enable Extensibility and Ensure Interoperability.

¹ A “resource description” is a collection of statements about an information resource: document, image, table, and so forth.

4.4 Consistent Identification Scheme

IRIs provide a well-developed infrastructure for globally unique identification which also facilitates integration with network-based solutions, and so should be used to identify resources. These identifiers can then be used to link document content with descriptions, or resource descriptions with other resource descriptions.

This requirement facilitates the following goals: Enable Extensibility and Ensure Interoperability.

4.5 Separate Processing

Metadata must be able to be processed, extracted, removed and so forth independently of the document content.

This requirement facilitates the following goals: Balance Expressiveness and Ease-of-Implementation, Enable Extensibility and Ensure Interoperability.

4.6 Levels of Conformance

The enhanced metadata proposal should define levels of conformance so that users and developers have expectations about the level of support to expect in a given application. At minimum, all OpenDocument applications must preserve foreign content, but other levels of conformance might relate to display and/or editing of such foreign content.

This requirement facilitates the following goals: Ensure Interoperability.

4.7 Ability to Add Multiple Annotations to the Same Content

This requirement falls out of the need to support collaborative editing, where the same document can be edited by multiple users, either synchronously or asynchronously. Since these users may also be applying metadata, the same words, paragraphs, etc., may see the same class of metadata applied multiple times. A practical consideration is that XML does not allow the same attribute name to be applied multiple times to the same element. So a metadata scheme which relied exclusively on metadata via attributes would have difficulty meeting this requirement.

This requirement facilitates the following goals: Enable Extensibility.

Bibliography

[NISO-META] NISO (2004) *Understanding Metadata*, NISO Press: Bethesda, MD, <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>