OASIS

Electronic Trial Master File (eTMF) Specification Version 1.0

Committee Specification Draft 02 / Public Review Draft 02

08 August 2016

Specification URIs

This version:
>http://docs.oasis-open.org/etmf/etmf/v1.0/csprd02/etmf-v1.0-csprd02.docx (Authoritative)
>http://docs.oasis-open.org/etmf/etmf/v1.0/csprd02/etmf-v1.0-csprd02.html
>http://docs.oasis-open.org/etmf/etmf/v1.0/csprd02/etmf-v1.0-csprd02.pdf

Previous version:
>http://docs.oasis-open.org/etmf/etmf/v1.0/csprd01/etmf-v1.0-csprd01.doc
>http://docs.oasis-open.org/etmf/etmf/v1.0/csprd01/etmf-v1.0-csprd01.html
>http://docs.oasis-open.org/etmf/etmf/v1.0/csprd01/etmf-v1.0-csprd01.pdf (Authoritative)

Latest version:
>http://docs.oasis-open.org/etmf/etmf/v1.0/etmf-v1.0.docx (Authoritative)
>http://docs.oasis-open.org/etmf/etmf/v1.0/etmf-v1.0.html
>http://docs.oasis-open.org/etmf/etmf/v1.0/etmf-v1.0.pdf

Technical Committee:
>OASIS Electronic Trial Master File (eTMF) Standard TC

Chair:
>Zack Schmidt (zs@sureclinical.net), SureClinical

Editors:
>Aliaa Badr (abadr@carelex.org), CareLex
>Jennifer Alpert Palchak (jalpert@carelex.org), CareLex
>Rich Lustig (rich.lustig@oracle.com), Oracle
>Catherine Schmidt (cms@SterlingBio.com), SterlingBio
>Zack Schmidt (zs@sureclinical.net), SureClinical
>Airat Sadreev,(airat@sureclinical.net), SureClinical
>Troy Jacobson,(troy.jacobson@forteresearch.com), Forte Research Systems
>Prabhat Vatsal (pvatsal@nextdocs.com), Next Docs

Additional artifacts:
>This prose specification is one component of a Work Product that also includes:
>
>* Schema document: http://docs.oasis-open.org/etmf/etmf/v1.0/csprd02/schema/OASIS-eTMF-V3-Schema-201607.xlsx
>* Metadata vocabulary: http://docs.oasis-open.org/etmf/etmf/v1.0/csprd02/vocabulary/OASIS-eTMF-V3-NCI-Vocab-201607.xlsx

Abstract:
>The OASIS eTMF Specification publishes details of an interoperable content classification system with rules, policies and procedures for how electronic content can be shared and customized for clinical trials. Machine readable, open standards-based technologies are used in a vendor neutral approach.

Instructions on converting the eTMF schema spreadsheet to an OWL RDF/XML ontology are available from IEEE [20].

Status:

This document was last revised or approved by the OASIS Electronic Trial Master File (eTMF) Standard TC on the above date. The level of approval is also listed above. Check the "Latest version" location noted above for possible later revisions of this document. Any other numbered Versions and other technical work produced by the Technical Committee (TC) are listed at https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=etmf#technical.

TC members should send comments on this specification to the TC's email list. Others should send comments to the TC's public comment list, after subscribing to it by following the instructions at the "Send A Comment" button on the TC's web page at https://www.oasis-open.org/committees/etmf/.

For information on whether any patents have been disclosed that may be essential to implementing this specification, and any offers of patent licensing terms, please refer to the Intellectual Property Rights section of the TC's web page (https://www.oasis-open.org/committees/etmf/ipr.php).

Citation format:

When referencing this specification the following citation format should be used:

[eTMF-v1.0]

Electronic Trial Master File (eTMF) Specification Version 1.0. Edited by Aliaa Badr, Jennifer Alpert Palchak, Rich Lustig, Catherine Schmidt, Zack Schmidt, Airat Sadreev, Troy Jacobson, and Prabhat Vatsal. 08 August 2016. OASIS Committee Specification Draft 02 / Public Review Draft 02. http://docs.oasis-open.org/etmf/etmf/v1.0/csprd02/etmf-v1.0-csprd02.html. Latest version: http://docs.oasis-open.org/etmf/etmf/v1.0/etmf-v1.0.html.

Notices

Copyright © OASIS Open 2016. All Rights Reserved.

All capitalized terms in the following text have the meanings assigned to them in the OASIS Intellectual Property Rights Policy (the "OASIS IPR Policy"). The full Policy may be found at the OASIS website.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published, and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this section are included on all such copies and derivative works. However, this document itself may not be modified in any way, including by removing the copyright notice or references to OASIS, except as needed for the purpose of developing any document or deliverable produced by an OASIS Technical Committee (in which case the rules applicable to copyrights, as set forth in the OASIS IPR Policy, must be followed) or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by OASIS or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and OASIS DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY OWNERSHIP RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

OASIS requests that any OASIS Party or any other party that believes it has patent claims that would necessarily be infringed by implementations of this OASIS Committee Specification or OASIS Standard, to notify OASIS TC Administrator and provide an indication of its willingness to grant patent licenses to such patent claims in a manner consistent with the IPR Mode of the OASIS Technical Committee that produced this specification.

OASIS invites any party to contact the OASIS TC Administrator if it is aware of a claim of ownership of any patent claims that would necessarily be infringed by implementations of this specification by a patent holder that is not willing to provide a license to such patent claims in a manner consistent with the IPR Mode of the OASIS Technical Committee that produced this specification. OASIS may include such claims on its website, but disclaims any obligation to do so.

OASIS takes no position regarding the validity or scope of any intellectual property or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; neither does it represent that it has made any effort to identify any such rights. Information on OASIS' procedures with respect to rights in any document or deliverable produced by an OASIS Technical Committee can be found on the OASIS website. Copies of claims of rights made available for publication and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this OASIS Committee Specification or OASIS Standard, can be obtained from the OASIS TC Administrator. OASIS makes no representation that any information or list of intellectual property rights will at any time be complete, or that any claims in such list are, in fact, Essential Claims.

The name "OASIS" is a trademark of OASIS, the owner and developer of this specification, and should be used only to refer to the organization and its official outputs. OASIS welcomes reference to, and implementation and use of, specifications, while reserving the right to enforce its marks against misleading uses. Please see https://www.oasis-open.org/policies-guidelines/trademark for above guidance.

Table of Contents

1 Introduction

[All text is normative unless otherwise labeled]

This document provides a specification for content classification and content interoperability in the clinical trial domain. This specification is designed for use by application developers who wish to design interoperable systems for the eTMF domain. However, the specification is built on global internet standards and the foundational components of the specification support other clinical trial domains.

[Non normative text]

Many organizations in the health sciences industry – BioPharma and Healthcare – use Electronic Document Management Systems (EDMS) to manage and archive clinical trial documents and records. Although many organizations coordinate and share the same documents, organizations lack a standards-based metadata vocabulary and method to classify and share electronic clinical trial documents, electronic medical images and related records. Additionally, it is difficult to efficiently search, report, and audit sets of clinical trial documents and their associated records due to a lack of a common metadata vocabulary. For example, if an organization wishes to search for a set of documents from the country 'France', unless each document is tagged with the metadata term 'Country', it would be very difficult to find such documents among distributed sets of clinical trial data. Information is often difficult to locate, unless it is indexed with a common published set of metadata vocabulary terms. This lack of interoperability among digital content repository resources, due to vocabulary and schema differences, makes rapid secure information discovery, retrieval, exchange, and sharing difficult for organizations. Systems may depend on different content models. Due to the lack of system to system interoperability and standard data formatting (e.g., CRO and Sponsor systems), such systems may find difficulty in sharing, exchanging, searching, and retrieving content.

Interoperability standard facilitates the information exchange in disconnected applications in a less error prone way. Less data cleansing and custom coding is required if both the sender and receiver comply with the standards, thus resulting in time savings. For example, a Pharma company employs eTMF system A but CRO's use system B, C and D. To get the TMF data in system A, system B, C and D must send the data in a format that A can understand. Consider this scenario for another Pharma company that uses system X for eTMF management. Now system B, C, and D must also send that data in format X can understand. If all applications A, B, C, D, and X comply with interoperability standards, only one interface is required for all these applications to exchange data/information.

Central to our vision is the belief that organizations that create document repositories should have the flexibility to classify, name, and organize documents in a way that meets their business needs and yet have interoperability, i.e., the ability to rapidly search and share repository resource information with other organizations in a standard format that is based on open systems standards.

1.1 Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119.

1.2 References

[1]   "[Normative] RDF 1.1 XML Syntax," 25 February 2014. [Online]. Available: http://www.w3.org/TR/rdf-syntax-grammar/.

[2]   "[Non-Normative] Universal Decimal Classification," [Online]. Available: http://www.udcc.org/udcsummary/php/index.php.

[3]   "[Normative] OWL2-SYNTAX OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax (Second Edition)," W3C Recommendation, 11 December 2012. [Online]. Available: http://www.w3.org/TR/2012/REC-owl2-syntax-20121211/.

[4]   [Non-Normative] A. Jentzsch, O. H. Zadeh, C. Bizer, B. Andersson and a. S. Stephens, "Enabling Tailored Therapeutics with Linked Data," 2009.

[5]   "[Non-Normative] European Medical Agency Policy on Digital Signatures," [Online]. Available: http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2013/07/news_detai l_001864.jsp&mid=WC0b01ac058004d5c1.

[6]   "[Non-Normative] EMA and FDA Digital Signature Policies on eSubmissions," [Online]. Available: http://esubmission.ema.europa.eu/doc/esignature/FAQs%20for%20EMA%20eSignature%20Capab ilities%20V3.0%20January%202014-Digitally%20Signed.pdf.

[7]   "[Non-Normative] National Cancer Institute (NCI)," February 2012. [Online]. Available: http://ncit.nci.nih.gov/.

[8]   "[Non-Normative] NCI EVS," [Online]. Available: http://evs.nci.nih.gov/.

[9]   "[Normative] Media Types IANA," [Online]. Available: http://www.iana.org/assignments/media-types/media-types.xhtml.

[10]   "[Non-Normative] Dublin Core Metadata," [Online]. Available: http://dublincore.org/.

[11]   "[Non-Normative] CarLex Content Models for Health Science," [Online]. Available: http://bioportal.bioontology.org/ontologies/3008/?p=terms.

[12]   "[Non-Normative] Object Management Group, Business Process Model and Notation (BPMN) Version 2.0, OMG," 3 January 2011. [Online]. Available: http://www.omg.org/spec/BPMN/2.0/.

[13]   "[Non-Normative] The Protégé Ontology Editor and Knowledge Acquisition System, Protégé ontology editing tools and open source community," Stanford Center for Biomedical Informatics Research(BMIR), Stanford University. [Online]. [Accessed http://protege.stanford.edu February 2012].

[14]   [Non-Normative] National Center for Biomedical Ontology (NCBO), February 2012. [Online]. Available: http://www.bioontology.org/.

[15]   [Non-Normative] P.R.Alexander, C.I.Nyulas, T.Tudorache, T.Whetzel, N.F.Noy and a. M.A.Musen, "[Non-Normative] Semantic Infrastructure to Enable Collaboration in Ontology Development," Philadelphia, 2011.

[16]   [Non-Normative] Y. G. Varun Ratnakar, "A Comparison of Semantic Markup Languages," Pensacola, Florida, 2002.

[17]   "[Non-Normative] TMF Reference Model," [Online]. Available: http://www.diahome.org/en/News-and-Publications/Publications-and-Research/EDM-Corner.aspx.

[18]   "[Non-Normative] Zip File Format Specification Version: 6.3.4," 1 October 2014. [Online]. Available: https://pkware.cachefly.net/webdocs/casestudies/APPNOTE.TXT.

[19]

[20] "[Non-Normative] National Cancer Institute, Enterprise Vocabulary Services, electronic Trial Master File Standard (eTMF) controlled vocabulary, published 21 July 2016 [Online]. Available: http://evs.nci.nih.gov/ftp1/CareLex/

"[Non-Normative] Institute of Electrical and Electronics Engineers (IEEE), "Ontology Based Data Conversion from Spreadsheet to OWL", 2012 Seventh ChinaGrid Annual Conference, 20-23 Sept. 2012, Xiaohui Zhang ; Sch. of Comput. Sci., Beijing Univ. of Technol., Beijing, China ; Ruihua Di ; Xiaochen Feng; DOI: 10.1109/ChinaGrid.2012.17; [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6337279

2 Problem Definition

[All text is non-normative unless otherwise labeled]

As clinical trial stakeholder organizations seek to move from paper-based record-keeping to electronic approaches, information interoperability, information standards and agency compliance are key factors in accelerating the safe delivery of therapies to patients.

In order to move clinical trial content (In this specification, the scope of the content is limited to the eTMF domain) from paper-based approaches to electronic Document Management Systems in the cloud, on-premises (in network) or offline, a standardized machine readable content classification system, with a web standards-based controlled metadata vocabulary, is needed. For those with access to Electronic Document Management Systems (EDMSs), a method to exchange content between systems is needed. For those without access to EDMSs, a method to exchange, view, and navigate content offline is needed. Ideally anyone with a web browser and proper permissions should be able to view the records and documents exported from an EDMS.

In the clinical trial domain, documents, medical images, and other electronic content are typically stored in a digital archive known as the electronic Trial Master File (eTMF). The eTMF serves as a central repository to store and manage essential clinical trial documents and content for possible use in regulatory submissions. Today, there is no standard that defines how eTMF documents and records should be for electronic export and exchange between systems. To maximize interoperability, it is important to adopt an open systems approach that is standards-based, operating system independent, software application independent, and computer language independent. Throughout this document, the use of the term *Digital Archive* does not assume the use of any specific archive storage format and should not be conflicted with the EU regulatory definition. Archive storage format is beyond the scope of this specification and it is up to the implementing organization or regulatory agency to specify such format.

Finally, any eTMF system should support government agency requirements for exported electronic records. The use of a standards-based, agency supported electronic document export formats will help raise the effectiveness, efficiency and safety of clinical trials and will help organizations share higher quality information more efficiently.

3 Objective

[All text is non-normative unless otherwise labeled]

The purpose of the OASIS eTMF Standard Specification is to define machine readable formats for clinical trial electronic Trial Master File (eTMF) content interoperability and data exchange, a metadata vocabulary, and a classification system that has a set of defined policies and rules. This goal is achieved by specifying:

a) An eTMF content classification model, which is comprised of a standards-based metadata vocabulary and a content classification ontology;

b) A set of eTMF content classification rules and policies;

c) An eTMF Data Model.

Features supported in the OASIS eTMF Standard Specification are divided into the following categories:

1. Core Technology Architecture

2. Content Classification System

3. Core Metadata and Content Type Term Sources

4. Content Model

5. Data Model

6. eTMF Metadata Vocabulary for Content Classification

7. eTMF Metadata Vocabulary for Content Tagging

The benefits of implementing interoperable systems that enable data sharing among clinical trial stakeholders are obvious and evident. Everyone who has used the internet experiences the benefits of interoperability – the ability to open and view a web page in a browser is premised on interoperability capabilities and web standards. Regardless of who authored or who hosted the content, users are able to view the content in a web browser.

The high level benefits of a standard for interoperable clinical trial information exchange that is based on web standards are summarized below:

- Accelerate clinical trial development timelines with interoperable data exchange
- Streamline agency compliance with standards-based exports and eSubmissions
- Enable sharing between clinical trial stakeholders with language independent taxonomies
- Enhance clinical trial safety and efficacy with serious adverse event data exchange

4 Core Technology Architecture

[All text is normative unless otherwise labeled]

The key OASIS eTMF foundational layers, as illustrated in Figure 1, include a Content Classification System (CCS) layer to automate content classification; a Vocabulary for Content Management Layer to describe classifications and documents through published vocabulary; and a Web Standard Technology Core Layer, which includes W3C standards for information discovery and exchange in addition to support for electronic and digital signatures and business process models that reduce paper handling processes.
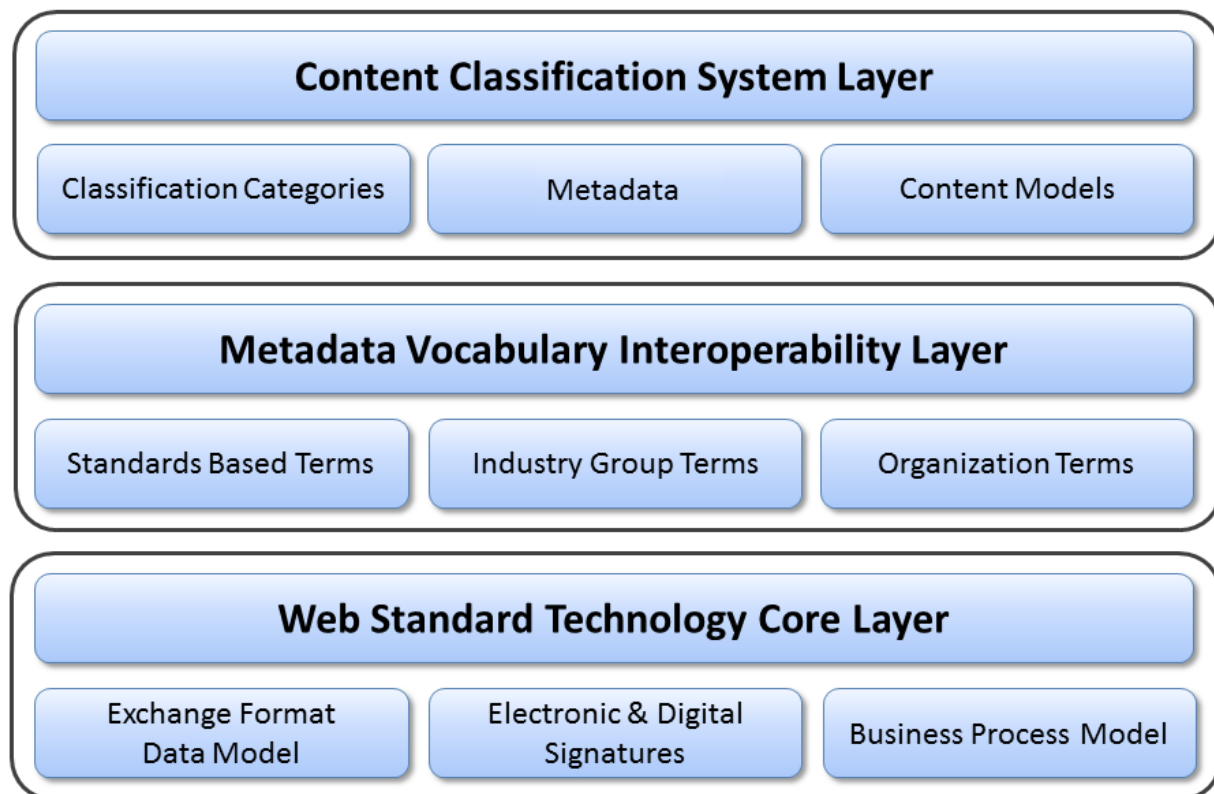
Content Classification System Layer

| Classification Categories | Metadata | Content Models |

Metadata Vocabulary Interoperability Layer

| Standards Based Terms | Industry Group Terms | Organization Terms |

Web Standard Technology Core Layer

| Exchange Format Data Model | Electronic & Digital Signatures | Business Process Model |

Figure 1: eTMF Foundational Layers

The **first layer** is the Content Classification System (CCS) layer, which includes three components: Classification Categories Component, Metadata component, and Content Model (Taxonomy) component. Details of this layer are discussed in Section 5.

The **second layer** is the Metadata Vocabulary Interoperability Layer. The OASIS eTMF model utilizes a controlled vocabulary for content management that is based on terms curated by the National Cancer Institute's NCI Thesaurus Enterprise Vocabulary Services (NCI EVS) [3]. The NCI thesaurus repository is the authoritative source of terms for this specification.

Non-normative text]

NCI EVS' term database is used globally, and contains terms used by healthcare and life sciences organizations such as HL7, CDISC, FDA, NIH and others. As part of its term curation effort, NCI manages the semantic relationships between terms and publishes its term database in a machine readable format known as RDF/XML; a web standard that enables interoperable data exchange between systems. Just as the internet uses the HTML to exchange data on websites, the OASIS eTMF Standard's metadata vocabulary layer is based on the RDF/XML web standard so that any language or term names can be used in the presentation layer (a typical layer that handles the presentation of data through user interface in n-tier architecture). Data interoperability is maintained through the use of RDF/XML in the metadata vocabulary layer, and is separate from the term label that end users see.
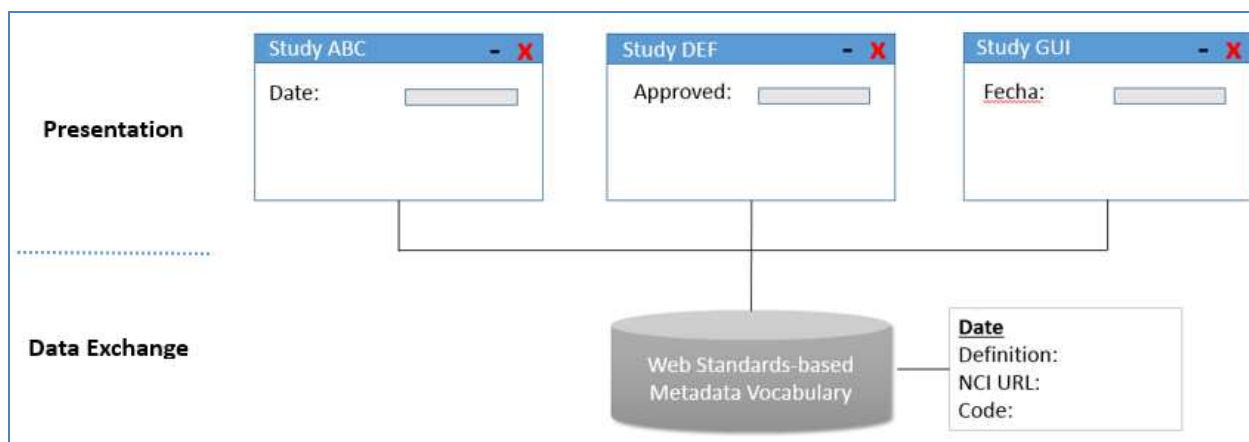
Figure 2: Presentation flexibility and data exchange interoperability with web standards-based metadata vocabulary

In addition to providing organizations with the ability to localize and customize any term label through the use of the Display Name metadata attribute, the metadata vocabulary layer provides organizations with the ability to create cross-referenced taxonomies that have a common interoperable data. As an example, a sponsor and a CRO organization might use a different display name to represent the same content in an eTMF. The use of display names enables these organizations to share data seamlessly regardless of the names or even language used to display the terms.
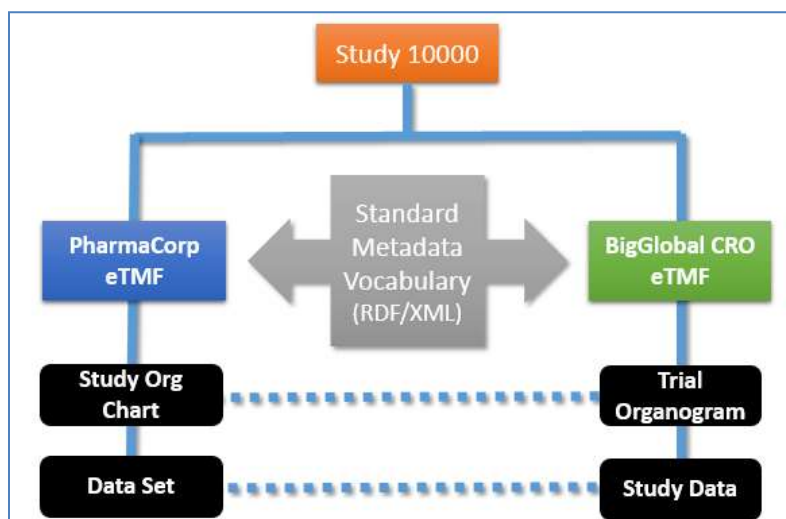

Figure 3: Example for multiple study taxonomies that can be used with different names yet the same data.

The metadata vocabulary layer is flexible and can be extended using a set of metadata vocabulary policies. The metadata vocabulary layer contains standards-based terms, terms sourced from industry groups, and organization-specific internal terms. The OASIS eTMF model metadata vocabulary layer defines a standard set of metadata vocabulary terms that are present in all OASIS eTMF Standard content models, enabling interoperable exchange of taxonomies between sponsors, contract research organizations, investigators and other stakeholders in the clinical trial ecosystem. Support for the TMF Reference Model (TMF RM) is provided in the OASIS eTMF Model through a cross-reference mapping of terms. The TMF RM provides a set of terms and taxonomy for the industry through its published spreadsheet. The OASIS eTMF Model provides a cross-reference mapping of the NCI-based OASIS metadata vocabulary terms to TMF RM terms, providing a path forward to a global ISO standard under the OASIS eTMF Standards effort.

The **third layer** is the Web Standard Technology Core layer. This layer is based upon the W3C RDF/XML, which represents web-based resources and can be easily searched, used, and shared through web-based applications.

5 Content Classification System

[All text is normative unless otherwise labeled]

The CCS Layer includes Classification Categorization (through a classification hierarchy or Taxonomy), a Metadata component, which characterizes content, and a Content Model component, which includes a published set of classification metadata for a specific domain (e.g., eTMF). These components are described in the following sub-sections.

5.1 Classification Categorization

[Normative and non-normative text]

Similar to how the Dewey Decimal system is used to classify books by category in a library, the OASIS eTMF model classification categories component utilizes Categories and Content Types with a decimal numbering format to classify documents or content items. The classification categories component format is based on the Universal Decimal Classification System (UDC) [1], which is widely adopted and used by libraries in over 170 countries worldwide. This component is designed for both human and machine readability. It allows for automated sorting of content classifications and documents in addition to offering a flexible and infinitely expandable hierarchical system that can use any vocabulary or text-based terms.
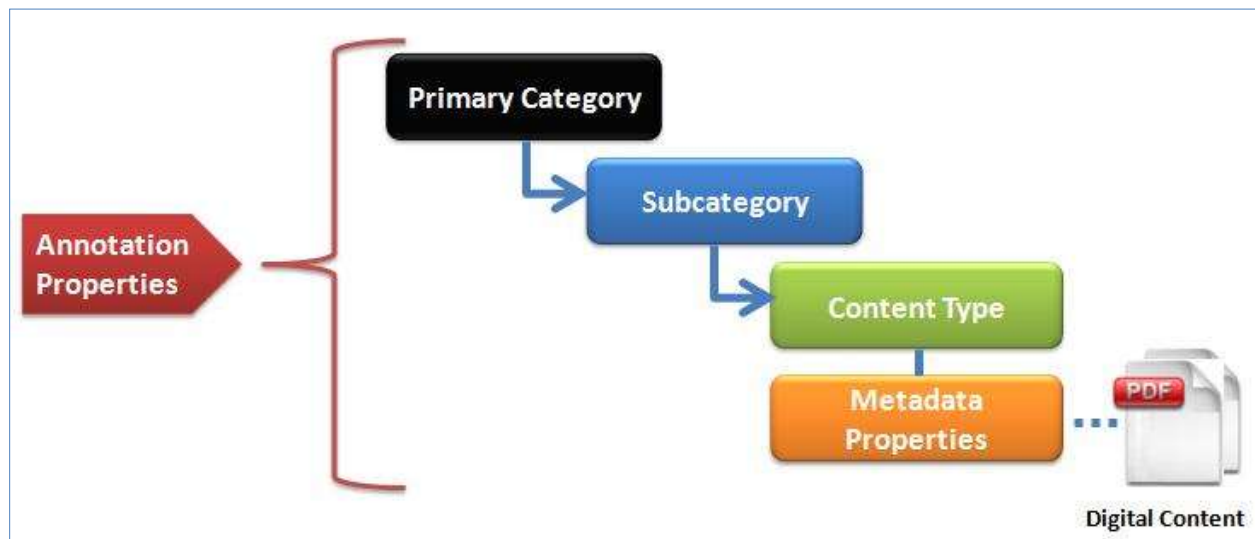

Figure 4: The OASIS eTMF Classification Categories Hierarchy

[Normative and non-normative text]

The classification categories component uses a hierarchy of classifications and a numbering scheme to classify documents and content in unique categories. To maximize machine readability, the classification and numbering scheme is based on the W3C XML naming conventions. In this naming convention, only simple text is allowed for category naming and numbering, while special characters, such as ( ) *$ @ ! and others are prohibited. Classification numbers use a digital dot notation where leading zeros are prohibited to conform to the W3C XML naming conventions.

The classification categories component contains classification entities, such as **Categories, Sub-Categories**, and **Content Types**. **Content Items**, such as documents or images, are linked to Content Types. **Metadata Properties** are associated with Content Types to provide information about Content Items. **Annotation Properties** provide information about Classification entities and Metadata properties. An example for the Classification Categories hierarchy is shown in Figure 4. Each digital Content Item is classified by a Primary Category, linked to a Sub-Category, with a link to a single specific Content Type. Annotation Properties provide annotations for Categories, Sub-Categories, and Content Types. As an example, the Annotation Property 'Definition' describes briefly the meaning of the title assigned to a Category, a Sub-Category, or a Content Type. The W3C OWL2 specification provides a general

description of Annotation Properties. Further details about Annotation Properties are provided in the Oasis eTMF Metadata Vocabulary Spreadsheet.

The Primary Categories in the eTMF domain are numbered from 100 – 199, providing 100 primary category divisions.   The use of three digits for primary category numbering allows for additional categories for future growth in other health science domain areas such as legal, administrative, research and development or other domain areas, using numbers from 200-999. An example of how content might be classified, named, and numbered using the OASIS eTMF content classification model is shown below.
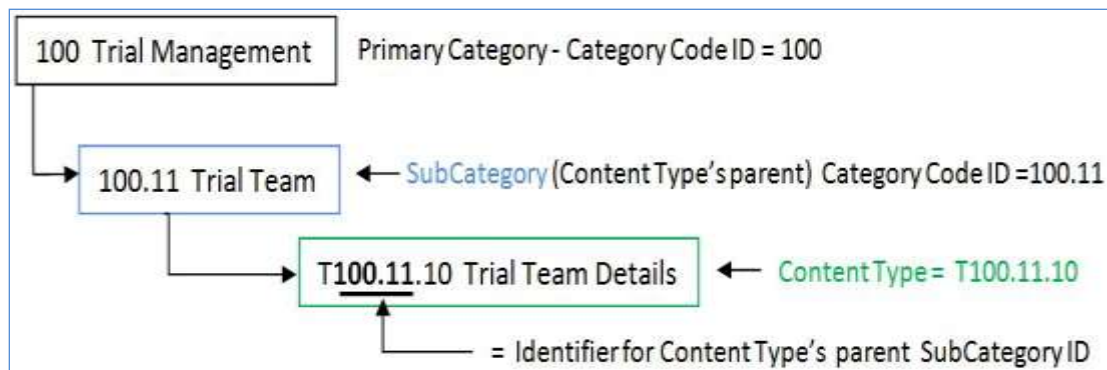

Figure 5: Example for OASIS eTMF Classification and Numbering

Classification categories terms are uniquely numbered using a hierarchical numbering scheme with a digital dot notation. The classification category numbers, known as **Category Codes**, allow content to be automatically classified in the content tree hierarchy, as shown in Figure 5. Each classification number forms a unique identifier that describes the position of the content in the hierarchy, making documents or content items easily sorted and searched. Design and naming of classification categories, in addition to rules for Category creation and modification, are all discussed in the following section.

Content type is a different entity than category or subcategory. Content type is associated with content and is like a file folder with documents under it. The use of 'T' as a prefix for content type category code uniquely identifies a content type. This prefix is not added to Category Codes for Categories and Sub-Categories.

Every Category, Sub-Category, and Content Type contains Annotation Properties. Content Types have both annotation properties and metadata properties assigned to them, as illustrated in Figure 6. Each level in the classification hierarchy contains at least the following annotation properties: Category Name, Category Code, and Term Type. Further details about Annotation Properties are provided in the Oasis eTMF Metadata Vocabulary Spreadsheet.

Figure 6: Example for Classification Hierarchy Relationships, Annotation Properties, and Metadata

5.1.1 Classification Categories Design

Primary Categories and Sub-Categories are assigned unique decimal numbers for identification and classification, called **Category Code**. They correspond to one and only one **Category Name**. Content Types are linked to a parent Sub-Category and are used to link related documents or content items. Additionally, the Category Code for any Content Type has a unique numeric code prefixed by the letter "T" for identification and classification. The Category Code for a Content Type is known as **Content Type ID**. A Content Type ID always contains its parent Sub-Category's Category Code number in the numeric prefix of the Content Type ID, as illustrated in Figure 7.

Figure 7: Classification hierarchy example showing categories and classification hierarchy names

5.1.2 Classification Categories Naming Scheme

The Classification System follows a naming scheme that combines the classification hierarchy name (i.e., Category Code, which is designed to automate document classification and locate the category in the content model hierarchy) and the simple text-based name. An example is the name "100 Trial Management" assigned to a Primary Category, where the first part "100" reflects the Category Code, while the second part "Trial Management" represents the Category Name.

Generally, all text-based names, or Category Names, should be unique and should not start with a number or a special character to be compatible with XML naming standards. In the following, we summarize the scheme used in generating the classification hierarchy name:

- **Primary Category:** Has a 3-digit number starting from 100: 100, 101, 102, etc. up to 999. The maximum number of Primary Categories is 900. Primary Categories act as headings, which include Sub-Categories. Primary Categories may have zero or more Sub-Categories as their children.

- **Sub-Category:** The Sub-Category classification hierarchy number is based on the number assigned to the Sub-Category's parent category. It is a sequence number delimited with a period, to indicate the Sub-Category division. Numbers for sub-categories start from 10: 10, 11, 12, etc., up to 99. The maximum number of nested sub-category items per parent category is 90. Examples are: 100.10, 100.11, 100.12, and 142.23.67. The maximum number of Sub-Category divisions is 5 excluding the 3-digits for the Primary Category (i.e., aaa.bb.cc.dd.ee.ff). Every Sub-Category must have one and only one Primary Category or Sub-Category as its parent. Sub-Categories may have zero or more Content Types as their children.
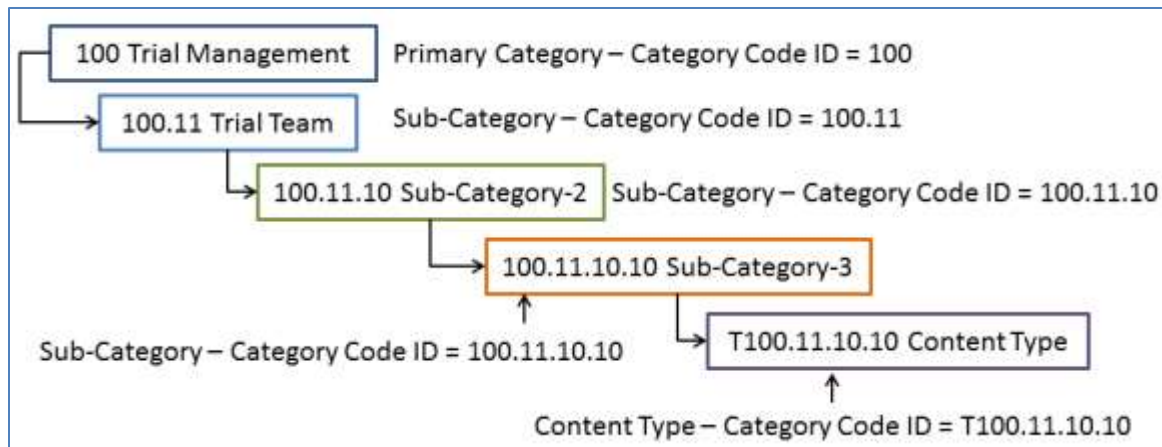
Figure 8: Sub-Category Nested Divisions Example

- **Content Type:** This is a two digit number sequenced from 10-99. It is based on the Content Type's parent Sub-Category and a sequence number delimited with a period to indicate the Content Type ID, as illustrated in Figure 9. The Content Type ID is always prefixed with the letter "T" to indicate that it is a Content Type entity. Content Types have one and only one Sub-Category as their parent. Typically, a Content Type ID number includes its hierarchical position in the Content Classification hierarchy.

The TMF RM and OASIS eTMF spec hierarchy initially map between Category = Zone, Subcategory = section, Content Type = Artifact. The OASIS eTMF's nested subcategories entity allows additional detailed classification, if desired. It is important to prefix all content types with the letter 'T' as per the specification to avoid duplicate naming.

Figure 9: Content Type ID Structure Example

5.1.3 Classification Categories Numbering Example

An implementation example for a content repository is shown in Figure 10. Within a repository, a single archive would typically contain a collection of related content, such as those relevant to a specific clinical trial study. This figure shows the archive of a clinical study labeled "Study 102880." Examples for Category numbering (e.g., "100 Trial Management Category"), Sub-Category numbering (e.g., "100.10 Trial Oversight Subcategory"), and Content Type numbering (e.g., "T100.10.10 Trial Master File Plan") are illustrated in the figure in different colors.
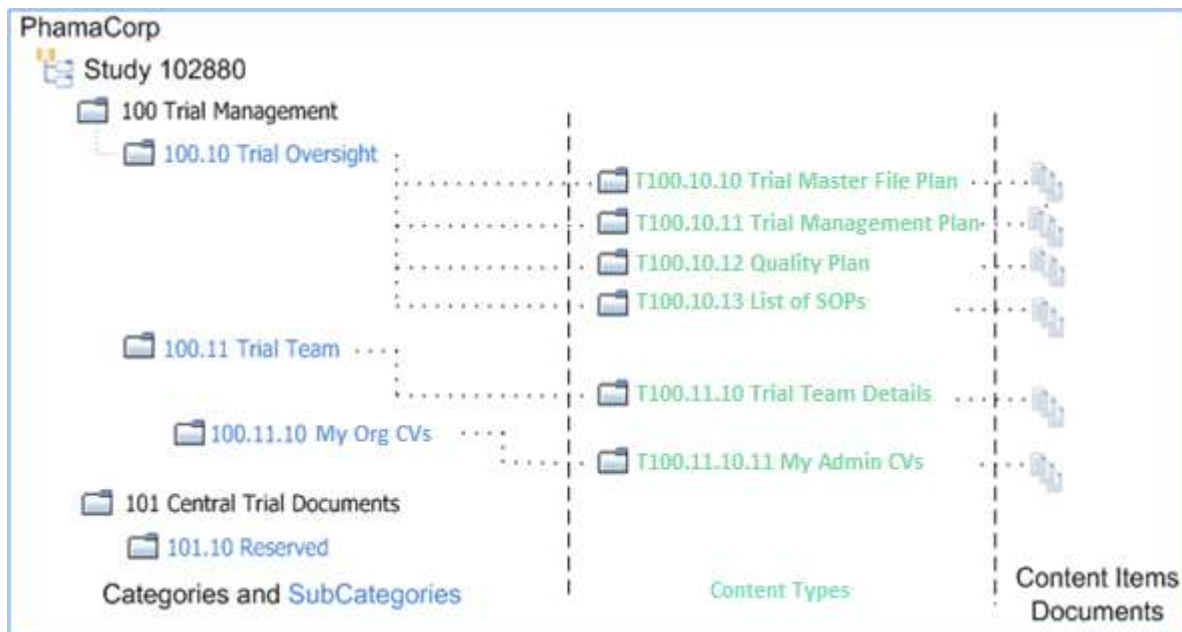
Figure 10: Content Classification Scheme for a Clinical Trial Example

5.1.4 Rules to Modify/Create Classification Categories Entities

Often organizations use abbreviated names, organization specific names, or localized language names for some or all of the Classification Entities. The eTMF model allows applications to use an abbreviation as the Classification Entity Name and use an optional internal label. For interoperability, when using an abbreviation or internal name with any Entity, make sure to retain the original Code for Categories and Content Types.

In an OASIS eTMF model implementation, a new Category, Sub-Category, or Content Type could be added to the classification hierarchy, as long as the classification numbering format is followed and the new entity does not conflict with any existing item. Addition of Categories and Content Types has two possibilities: the new Category or Content Type can either be **Organization-specific** or **Domain-specific**. The former is a Classification Category or Content Type created by an organization to meet their needs, while the latter is a Classification Category or Content Type related to a Content Model in a specific domain area, e.g., the health sciences domain.

In the first case, the details of the **Organization-specific** Category or Content Type are entered by the user and a Category Code and a machine-readable unique Code are generated locally in a way that ensures the classification numbering format is followed and no conflicts exist in the classification hierarchy. Category Codes and Codes are for internal use. Category Codes are generated following the published Category Code format, while Codes should be generated with 'Z' prefix, as illustrated in Table 1. Note that, Organization-specific Classification Categories and Content Types could also be imported to a content model (e.g., one organization publishes its own Classification Categories and Content Types for public use and another organization uses them). However, the importing party must check for conflicts, as Organization-specific Classification Categories and Content Types are not interoperable. Organization-specific entities are for the internal use by an organization, while Domain-specific entities are part of a controlled vocabulary.

Table 1: Vocabulary Term Code Prefix

Term Source	Code Prefix	Code Example
NCI Thesaurus	**C**	**C**12345
CareLex Industry Expert Panel (prior to submission to NCI)	**X**	**X**12345

Provisional Term – Pending Review	**Y**	**Y**12345
Organization Internal Use Term	**Z**	**Z**12345

In the second case, the **Domain-specific** Category or Content Type is imported (through looking up the term) from a published Content Model. A content model is a collection of content classifications, relationships, and metadata, and is further discussed in Section 5.3. In this case, the Category Code could be modified to enable placing the new Category in the classification hierarchy. However, the assigned Code should not be changed under any circumstances, so as not to violate interoperability. Generally, new Categories, Sub-Categories, and Content Types could only be imported from existing Content Model Categories, Sub-Categories, and Content Types, respectively (i.e., no mixing is allowed, such as importing an existing Content Model's Content Type as a new Sub-Category).

Restrictions apply to which Annotation Properties can be modified for different types of Classification Categories and Content Types. Except for annotation properties of Organization-specific classification Categories and Content Types, only Display Name, Definition, Abbreviation, Term Source URL, and Requirement Annotation Properties can be modified. All Oasis vocabulary terms (i.e., the domain-specific terms) must have a URL with either NCI or CareLex for the Term Source URL annotation property. Table 2 illustrates rules for modifying/creating annotation properties for different types of Classification Categories and Content Types. Please refer to the Oasis eTMF Metadata Vocabulary Spreadsheet for detailed information regarding Annotation Properties. It is worth mentioning that domain specific classifications are considered to be the core for the content model instance in use. Additionally, a content model could use more than one domain.

Table 2: Rules for Creating and Editing Annotation Properties of Categories and Content Types

Type of Term/ Create or Edit Annotation Property	Display Name	Preferred Name	Requirement	Definition	Value Set	Abbreviation	Data Type	Term Source URL
Domain-specific	Yes	No	Yes	Yes	No	Yes	No	Yes
Organization-specific	Yes	Yes (1)	Yes	Yes	Yes	Yes	Yes	Yes (2)
(1) Preferred Name can be created but not edited. (2) Organization-specific can use any URL as its source.								

Domain-specific Classification Categories and Content Types cannot be deleted from the content model. Instead of deleting a classification item that is unused, the item's name is marked as 'Reserved', to denote that the item is not used in the content model. This marking enables interoperability and the future use of the item, if needed. However, Organization-specific Categories and Content Types can be deleted from the Content Model. Table 3 provides a summary of the OASIS eTMF modification rules discussed in this section.

(Continued)

Table 3: Rules for Addition, Modification, and Delete of Classification Categories and Content Types

Type of Classification Category or Content Type/Action	Import Term	Generate Code	Add / Modify Term	Delete/Reserve Term
Domain-specific	Yes	No	Yes/Yes	No/ Yes
Organization-specific	Yes	Yes	Yes/Yes	Yes/No

Finally, it is worth mentioning that if organizations decide to create organization specific classification categories and content types, then organizations involved should mutually agree upon these terms. However, it is not required that any receiving organizations accept such terms.

5.2 Metadata Definitions

Metadata is used to give additional information about digital content items and classification categories. The OASIS eTMF model includes two types of metadata (illustrated in Figure 11):

- Metadata Properties: Describes the Content Type, e.g., Study ID, Site ID, Org, etc. It is also called Data Properties.
- Annotation Properties: Describes attributes of Content Classifications and attributes of Metadata Properties.

Each of these metadata types are discussed in details in the following sections.
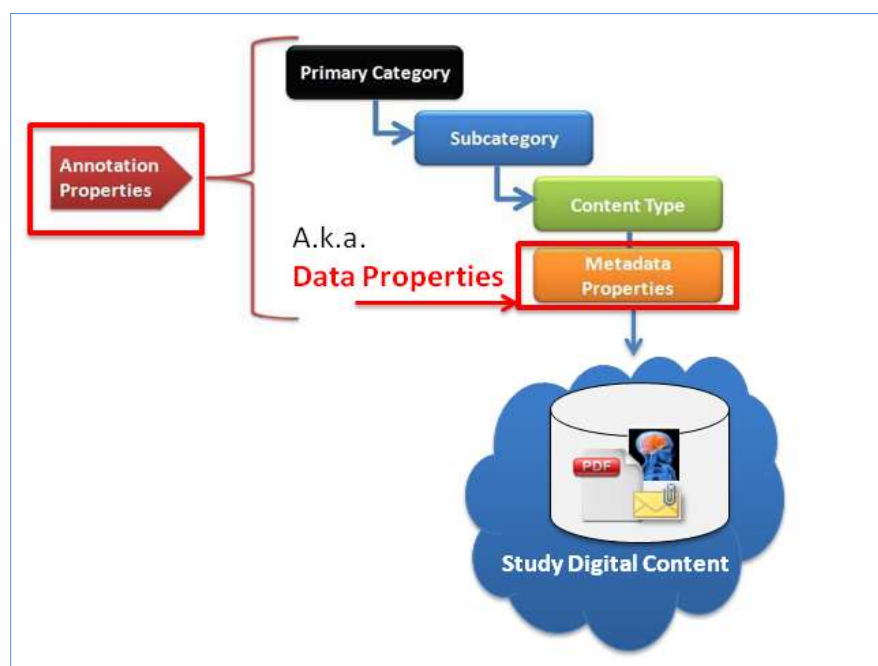
Figure 11: OASIS eTMF Metadata Definitions

5.2.1 Metadata Properties

[Non-normative text]

What is metadata and why is it needed? Metadata, or information about data, is used to tag or index digital content items such as documents. In the context of an EDMS, metadata is used to help organizations automate the classification and search, report, and exchange of digital content items. Every content item has some basic core metadata. For example, every content item contains metadata such as the 'file type.' That metadata facilitates file exchange and enables applications to automatically process files. Figure 12 shows an example for file metadata, displayed upon right-clicking on it.

Figure 12: File Metadata Example
[Non-normative text]

If every organization uses different metadata terms (as shown in Figure 13), it is impossible to enable efficient global search, reporting, and classification of documents within and outside of an organization.
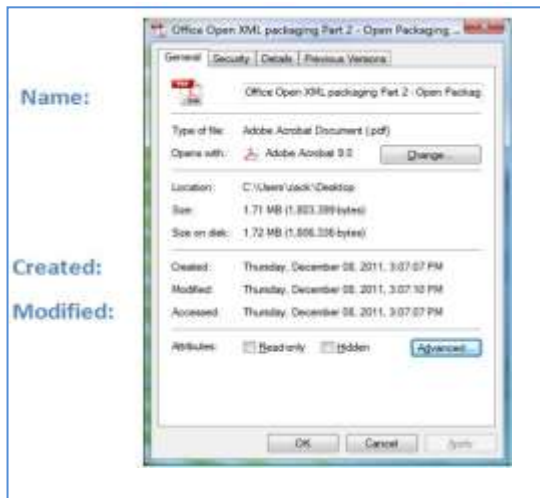
As illustrated in the figure, the use of dissimilar metadata terms inhibits efficient search and decreases interoperability, while the use of a controlled vocabulary enhances search efficiency and business productivity.
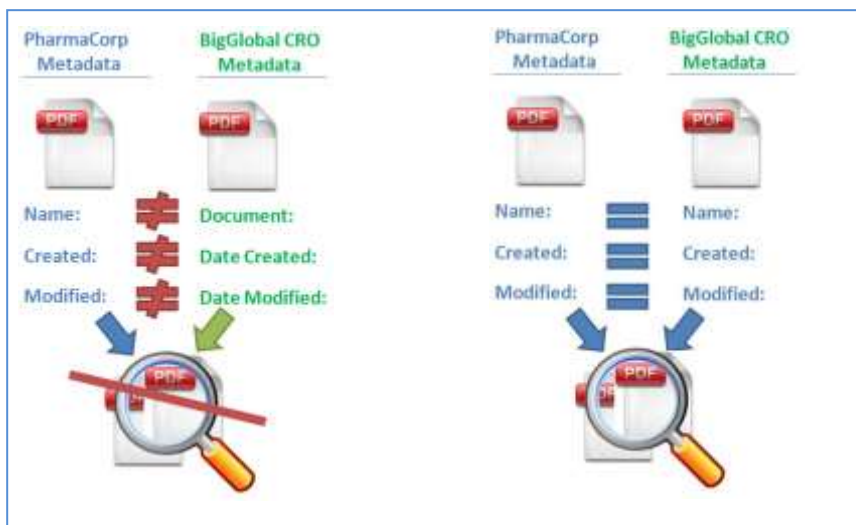
Figure 13: Similar vs. Dissimilar Metadata
[Normative and non-normative text]

The OASIS eTMF model metadata is also known as **Data Properties** (both names are used interchangeably throughout this document). Data Properties are similar to XML attributes. They define the attributes of Content Types. For example, 'Country' is a Data Property, which describes the country for a Content Item. Data properties are associated with each Content Type instance. Data Properties are simple Content Item tags without explicit relationships defined within the model. For a general description of a Data Property, see the W3C OWL2 specification [2]. Data properties have a number of annotation properties that provide additional information.

The Oasis eTMF model supports a comprehensive list of metadata vocabulary that satisfies many requirements in the clinical trials domain (please refer to the Oasis eTMF Metadata Vocabulary spreadsheet for further details). However, in order to allow for a flexible model that organizations can easily adapt to existing business terms and business processes, Organization-specific metadata properties can be used with any content type. Organization specific metadata terms, unless widely published, may not be interoperable with content models in use by other organizations or entities. Due to its lack of interoperability, Organization-specific metadata is not published in the Oasis eTMF model.

Metadata properties are associated with a Content Type and specify the content item filling requirement. Each content type in the Oasis eTMF content model specifies two sets of metadata properties: required (i.e., should be filled for each content item) and optional. This is achieved through the Content Type annotation properties **Required Metadata** and **Optional Metadata**; a comma separated list of required and optional metadata properties, respectively. The Oasis eTMF model provides a default set of required and optional metadata properties for each Content Type (please refer to the Oasis eTMF Metadata Vocabulary spreadsheet for details).

The Oasis eTMF default set of required metadata properties per content type cannot be modified. However, Organizations can extend this set through associating their Organization-specific metadata properties with Content Types as required metadata, whenever necessary.

To ensure interoperability in the OASIS eTMF model, a number of rules must be followed when adding, modifying, or deleting metadata terms, as summarized in Table 4 and Table 5 . Metadata terms could be added to the content model through looking up (i.e., importing) or enabling the user to enter its details, as shown in Table 4. The former case applies to Metadata properties imported from other content models as their details are already available in other published content models or in resources for public vocabularies, while the second case applies to Organization-specific Metadata terms. The Oasis eTMF content model Metadata properties are part of the published content model and cannot be added.

Generally, interoperability is satisfied through the **Code** annotation property included with every Metadata property. The Code annotation property provides a unique number assigned to every Metadata property, Category, Sub-Category, and Content Type. To ensure interoperability, this unique code value cannot be modified once created.

In case of Organization-specific metadata terms, the details of these terms (i.e., values of annotation properties) are user-defined with exception to Code annotation property, which is generated locally. Metadata properties could be imported through looking up existing sources, such as the NCI Thesaurus and Dublin core. When importing metadata terms, the Code annotation property should be retained unmodified. Note that, some Metadata properties may not include the Code annotation property. To ensure interoperability, these codes should be generated locally whenever required. The local generation of Codes should ensure the uniqueness of the generated code such that no conflicts exist with any other codes in the content model. Rules for Code generation and the import of metadata properties are summarized in Table 4. This table also presents rules for addition and modification of metadata properties .

Metadata Type/Action	Import Term	Generate Code	Add / Modify Term	Delete/Reserve Term
Metadata Properties from External Resources	Yes	No(1)	Yes/Yes	Yes/No
Organization-specific	No	Yes	Yes/Yes	Yes/No
Oasis eTMF Metadata Properties	No	No	No/Yes	No/No
(1) Only generate Code if a code is not available from the term source (e.g., Dublin Core metadata)				

Table 4: Rules for Addition, Modification, Import, and Delete of Metadata Properties

Modification of metadata terms is allowed; however, not all annotation properties could be modified. Table 5 illustrates annotation properties that could be modified per metadata type.

Metadata/ Add or Edit	Display Name	Preferred Name	Requirement	Definition	Value Set	Abbreviation	Data Type	Term Source URL
Oasis eTMF Metadata Properties	Yes	No	No	Yes	No	Yes	No	No
Organization Specific	Yes	Yes (1)	Yes	Yes	Yes	Yes	Yes	Yes
Metadata Properties from External Resources	Yes	No	Yes	Yes	No	Yes	No	Yes(2)
(1) Preferred Name can be created but not edited. (2) Only allow editing of a URL if there is no URL available from the term source.								

Table 5: Rules to Edit/Create Metadata Annotation Properties

The Oasis eTMF metadata properties are not allowed to be reserved or deleted, as they provide basic information in any content model (as shown in Table 4). Organization-specific metadata properties, and those added from external resources, can be deleted if they are not actively used (referenced) in any content item.

5.2.2 Annotation Properties

Annotation properties provide additional information about different entities in the OASIS eTMF model, including classification categories (i.e., Categories and Sub-Categories), Content Types, and Metadata Properties. An example for annotation properties is Code, which uniquely identifies different terms in a content model. The OASIS eTMF model includes two types of annotation properties: **Core** and **Organization-specific**. The former are part of the published content model, while the latter are added to the content model according to an organization's specific needs.

As with Classification Categories, Content Types, and Metadata Properties, Annotation Properties also have rules for modification, addition, and deletion. Table 6 summarizes rules for addition, deletion, modification, and import of annotation properties. Generally, Core annotation properties are part of the published content model and they can be neither added nor deleted. Both core and organization specific annotation properties cannot be imported (i.e., looked up in public resources). Additionally, when an organization-specific annotation property is added, a unique code is generated locally. This code should not conflict with any other existing code.

Annotation Properties Type/Action	Import Term	Generate Code	Add / Modify Term	Delete/Reserve Term
Core	No	No	No/Yes	No/No
Organization-specific	No	Yes	Yes/Yes	Yes/No

Table 6: Rules for Addition, Modification, Import, and Delete of Annotation Properties

Annotation properties also have additional annotation properties that provide details about them. In case of modifying details of annotation properties, only a subset of the annotation properties of annotation properties could be modified for the Core type, while more annotation properties could be modified for the

Organization-specific type. Annotation properties that can be created or modified are illustrated in Table **7**.

Annotation Properties Type/ Create or Edit	Display Name	Preferred Name	Requirement	Definition	Value Set	Abbreviation	Data Type	Term Source URL
Core	Yes	No	Yes	Yes	No	Yes	No	No
Organization-specific	Yes	Yes (1)	Yes	Yes	Yes	Yes	Yes	Yes
(1) Preferred Name can be created but not edited.								

Table 7: Rules to Edit/Create Annotation Properties of Annotation Properties

5.3 Content Model

[Non-normative text]

Content models represent content classifications, relationships, and metadata in a semantic web taxonomy or 'Ontology'. Content models are technology agnostic; there is no particular software, computer operating system, or application required in order to use them. They are published using a Web standard format known as 'OWL'. This format allows browser-based discovery of information. To illustrate how the OASIS eTMF content model can be used in an organization to manage content items, consider the technology that nearly everyone has used - management and organization of music MP3 files.

[Non-normative text]

All MP3 files today contain standard tags or metadata, such as Artist, Title, Album, and Genre that are embedded in the MP3 file to enable rapid electronic classification and search. These tags describe the music or album. A user typically classifies the MP3 files by Genre or category, such as 'Rock', 'Jazz', etc. Because of standard MP3 tags/terms and the MP3 genres or classifications, it is possible to easily search for MP3 songs on the internet or in a file system. Additionally, many software applications can read an MP3 library collection, import it, and automatically classify and sort MP3s based on the metadata tags. Similarly, by using both a standardized set of metadata terms for tagging and a set of published content classification categories, health science content items (such as documents, images, photos, and other digital media) can be tagged, organized, and more efficiently searched.

As in the MP3 content classification example, the OASIS eTMF content model uses a published set of content classification categories in a hierarchy, as shown in Figure 14. Each content classification primary category contains Sub-Categories and Content Types. A Content Type is linked to a reusable collection of metadata for a category of items or documents. Content Data Properties, or Metadata, describe the document or content item's attributes.
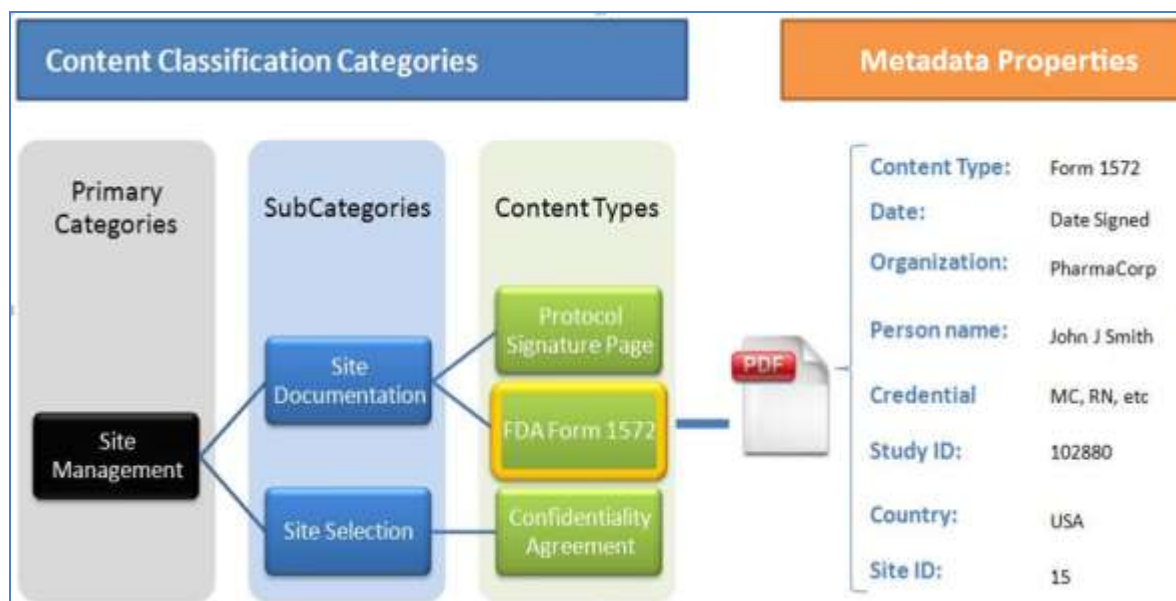
Figure 14: OASIS standard eTMF Content Model Example

While the OASIS eTMF content model hierarchy is flexible and interoperable, it is also useful for organizations that want to save time and resources and share content models with others.

5.3.1 Content Model Format

[Non-normative text]

The OASIS eTMF content models are created and published as ontologies based on the W3C's OWL 2.0 syntax [1] and RDF/XML. Semantic web allows seamless sharing, linking, and search of data across domains. Additionally, the W3C's OWL format is used in describing the content model in order to retain compatibility with the technology that NCI, WHO, and other leading health science vocabulary standards groups are using to model information relationships [3]. The OWL format can be used to represent hierarchies, complex relationships, data properties, and values. It also allows search engine discovery and presentation in a web browser. Furthermore, many Organizations are moving towards a semantic web model, which enables interoperability between content models. For example, the CTMS model follows BRIDG semantic web model; hence, the OASIS eTMF content model can be more readily integrated with the CTMS for interoperability given they are both based upon the semantic web model.

[Non-normative text]

The best way to view, understand, and edit the OASIS eTMF content model hierarchy is to open the content model in the Protégé OWL editing application. Using the NCBO BioPortal/OASIS sites, download the OASIS eTMF content model OWL file and open it in Protégé. In Protégé, users can add new Categories, Content Type, and Data Properties, and perform editing operations to change labels or other minor changes.

[Non-normative text]

Protégé is a very sophisticated application. However, the OASIS eTMF content model only uses a subset of the features of the OWL syntax and Protégé. The areas in Protégé that are most relevant are Classes, Individuals, and Data Properties.show a screenshot of the Protégé editor. This example uses the Stanford University's Protégé application, freely available for download [4].
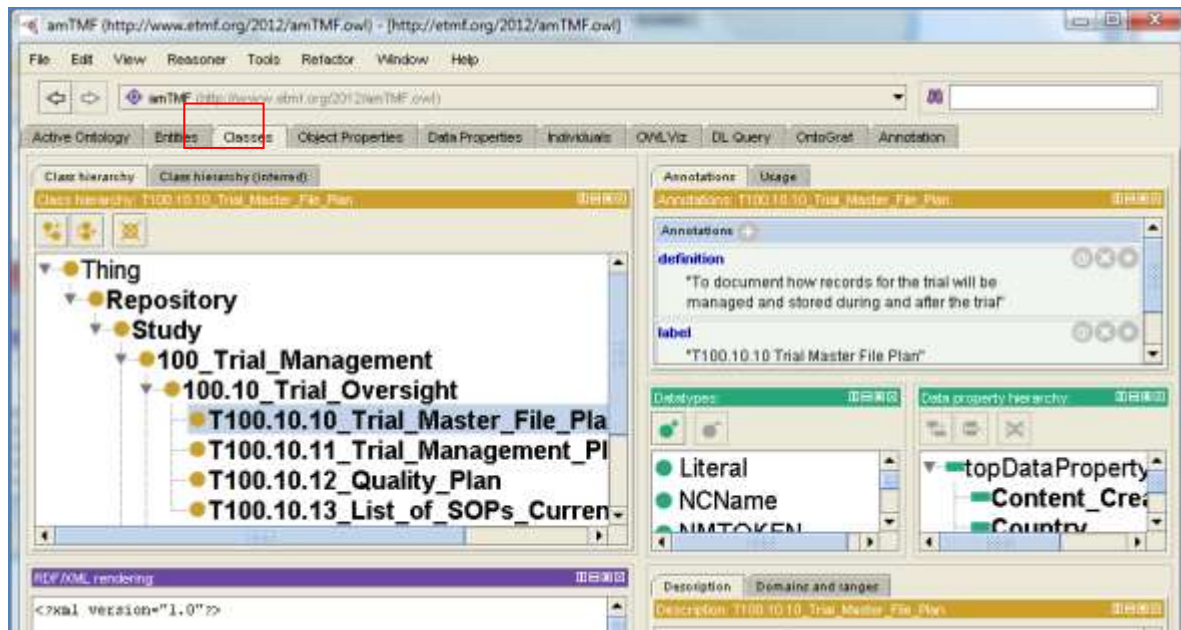
Figure 15: Using Protégé OWL Editor to Modify OASIS eTMF content Models

[Non-normative text]

An example for a W3C RDF/XML file is illustrated in Figure 16. The W3C RDF/XML is used as the syntax for content model representation and exchange. The file contains RDF and OWL in XML. Additionally, it includes reference to content model **ontology** for the OASIS eTMF Standard**.** Furthermore, the RDF/XML file contains the content model ontology for a study instance.

[Non-normative text]

The **.owl** filename extension is used for the RDF/XML files. Filenames for content model exchange shall be similar to IETF URL naming as follows: Alphanumeric characters and the hyphen '-' (special character) can be used to ensure future compatibility.

```
XML V1.0 ──▶   <?xml version="1.0"?>

              <!DOCTYPE rdf:RDF [
                  <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
                  <!ENTITY swrl "http://www.w3.org/2003/11/swrl#" >
                  <!ENTITY swrlb "http://www.w3.org/2003/11/swrlb#" >
                  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
                  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
                  <!ENTITY etmf "purl.bioontology.org/ontology/etmf#">
                  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
                  <!ENTITY protege "http://protege.stanford.edu/plugins/owl/protege#" >
                  <!ENTITY xsp "http://www.owl-ontologies.com/2005/08/07/xsp.owl#" >
RDF/OWL ──▶   ]>

              <rdf:RDF xmlns="http://purl.bioontology.org/ontology/etmf#"

              ▬ ▬ ▬ ▬ ▬ ▬ ▬ ▬ ▬ ▬ ▬ ▬ ▬ ▬ ▬ ▬ ▬ ▬ ▬ ▬

                  <!-- http://purl.bioontology.org/ontology/etmf#Abbreviation -->

                  <owl:AnnotationProperty rdf:about="&carelex;Abbreviation">
                      <rdfs:label rdf:datatype="&xsd;string">Abbreviation</rdfs:label>
                      <Value_Set rdf:datatype="&xsd;string"></Value_Set>
                      <Abbreviation rdf:datatype="&xsd;string">Abbrev</Abbreviation>
                      <Term_Type rdf:datatype="&xsd;string">Annotation Property
              </Term_Type>
                      <Code rdf:datatype="&xsd;string">C42610</Code>
                      <Metadata_Type rdf:datatype="&xsd;string">Core</Metadata_Type>
                      <Term_Source rdf:datatype="&xsd;string">NCI</Term_Source>
                      <Requirement rdf:datatype="&xsd;string">Optional</Requirement>
                      <Definition rdf:datatype="&xsd;string">Preferred Abbreviation of
              term name</Definition>
                      <DataType rdf:datatype="&xsd;string">String</DataType>
                      <Content_Model rdf:datatype="&xsd;string">eTMF</Content_Model>
                      <Term_Source_URL rdf:datatype="&xsd;string">
              http://ncit.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI%
              20Thesaurus&amp;code=C42610</Term_Source_URL>
                      <Display_Name>Abbreviation</Display_Name>
                      <Preferred_Name>Abbreviation</Preferred_Name>
                  </owl:AnnotationProperty>
```
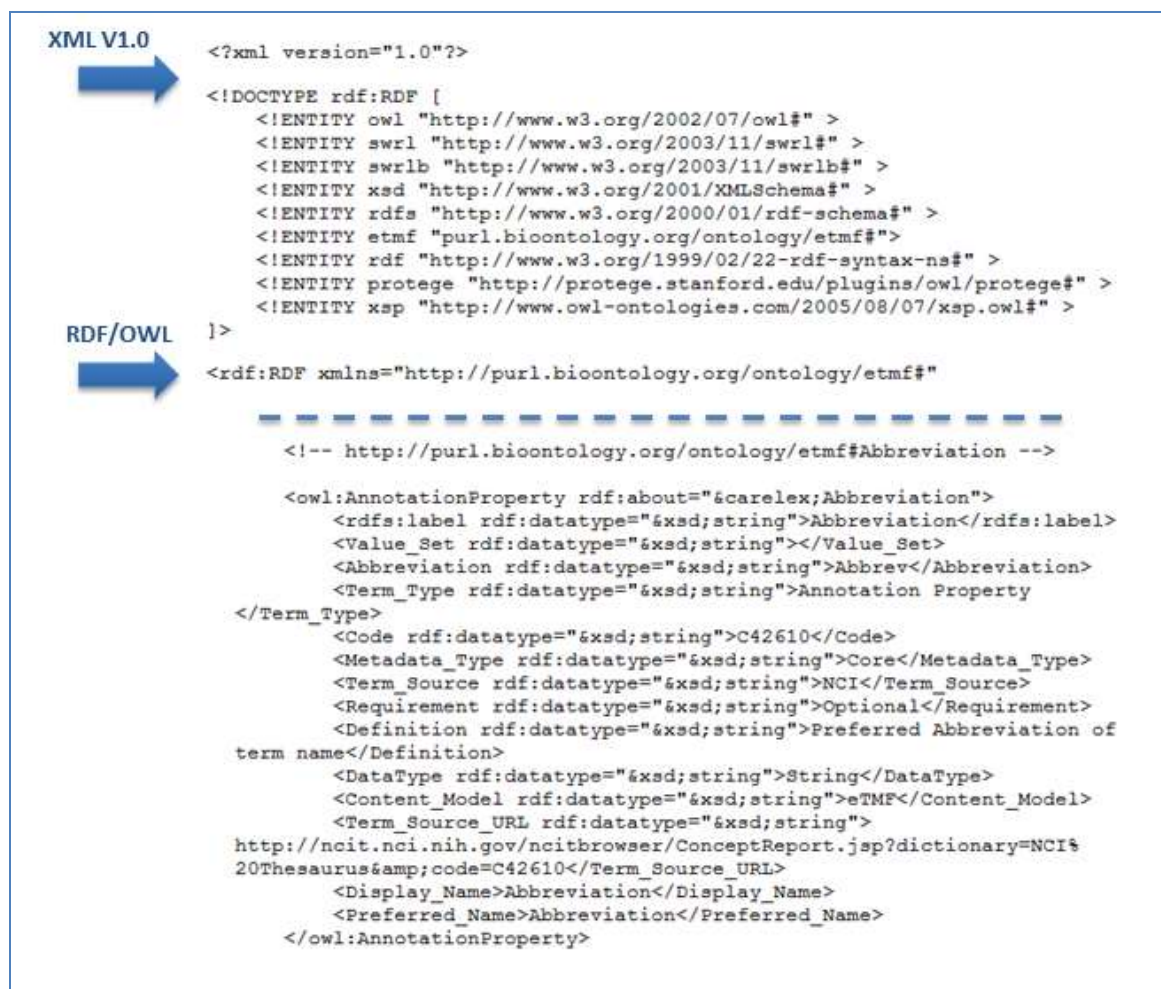
Figure 16: Example W3C RDF/XML File Snippet

5.3.2 Content Model Exchange

The OASIS eTMF content model ontology is represented as W3C OWL2 classes. In this way, content models can be easily edited and shared by anyone. Content Model instances are expressed as W3C RDF/XML (e.g., eTMF specific study).

Generally, RDF/XML is used as the syntax for content model exchange. Content models can be exchanged using Serialized RDF/XML or RDF/XML as a file with .owl extension (e.g., etmf.owl). No specific exchange protocol is specified by RDF/XML nor is one required for the content model exchange (the protocol is application/implementation specific). Any protocol which supports exchange of RDF/XML files or serialized data can be used (e.g., W3C http/s, REST, SOAP, RPS, CMIS, etc.).

5.3.3 Content Model Versioning

Versioning of content models is supported through the W3C OWL Versioning Policies. The W3C OWL supports granular level of versioning. However, version management is considered to be an application-specific task. Being a W3C standard, an OWL file includes an element for content model versioning; the **owl:versionInfo**, which provides a hook suitable for use by version management systems.

The OASIS content model version numbering text follows the Major.Minor numbering format, where the Major part reflects the content model ontology version number. The Minor part should reflect an Organization-specific version of the content model (as illustrated in Figure 17). This Minor numbering may be enhanced with Organization-specific and application-specific numbering within the W3C OWL versioning policies. The element owl:versionInfo in RDF/XML should be used for Categories, Content

Types, Annotation Properties, and Metadata Properties. Finally, the first version of the OASIS eTMF content model would be published as Version 1.0.
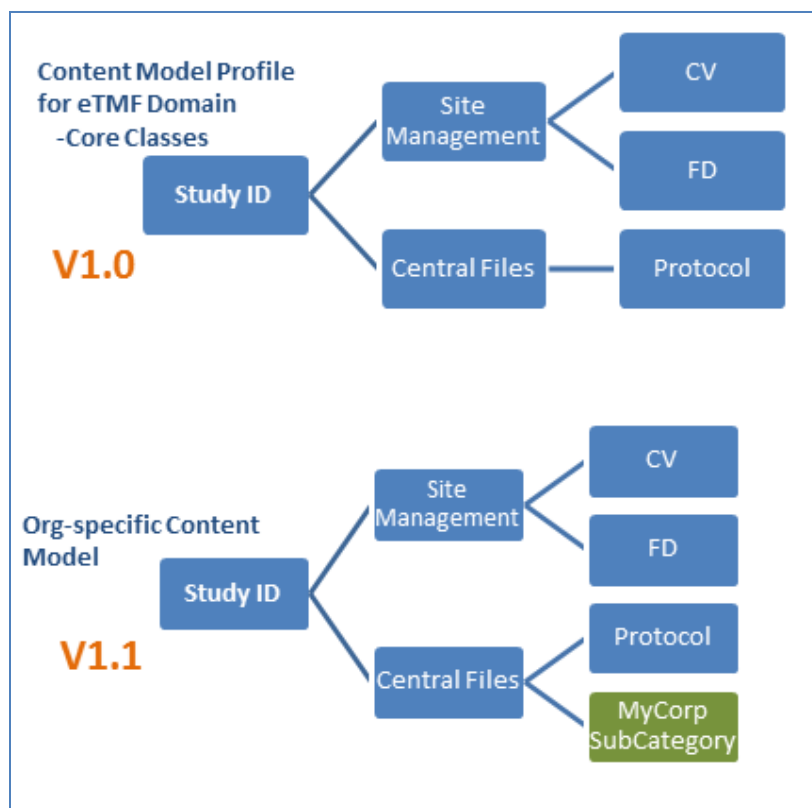

Figure 17: Content Model Versioning Example

6 Web Standard Technology Core

The Web Standard Technology Core Layer has three components: the Data Model, the Electronic and Digital Signatures technologies, and Business Process Model support.

6.1 OASIS eTMF Data Model

The OASIS eTMF data model represents a single instance of an eTMF content model for a single clinical trial. An eTMF data model instance contains data values for metadata properties and content items for this clinical trial study. It also includes the core and organization specific content model categories and content types. Note that, content item handling operations, such as viewing and opening, are application dependent.

The OASIS eTMF data model enables organizations to package, archive, and share clinical trial records with other systems or with regulatory agencies. To comply with regulatory authority guidelines and common industry practice, the eTMF records and content items can be exported using common file formats such as XML and PDF, or any published IANA media type. These electronic eTMF archives can be viewed in a simple web browser or exchanged using simple files and folders that are operating system and application independent. The information in all data models should contain the content model instance, content item URIs and metadata values, and should follow content classification rules and policies. The OASIS eTMF clinical trial data model can be exported to other systems and can be viewed in any web browser.

6.1.1 OASIS eTMF Data Model Exchange Format

The OASIS eTMF data model primary exchange file format is the RDF/XML. The file includes core and organization specific Categories and Content Types (reserved and in use), annotation properties, metadata properties, and links to instance resource content items offline and online (linked data). Content item name is unspecified and the content item file format is any supported IANA media type format [1]. Additionally, the Oasis eTMF content types should support the export of any IANA media type format. This format enables the interoperable exchange of content models, content items from cloud or physical media, metadata terms, and metadata values for clinical trial study instances between systems and applications.
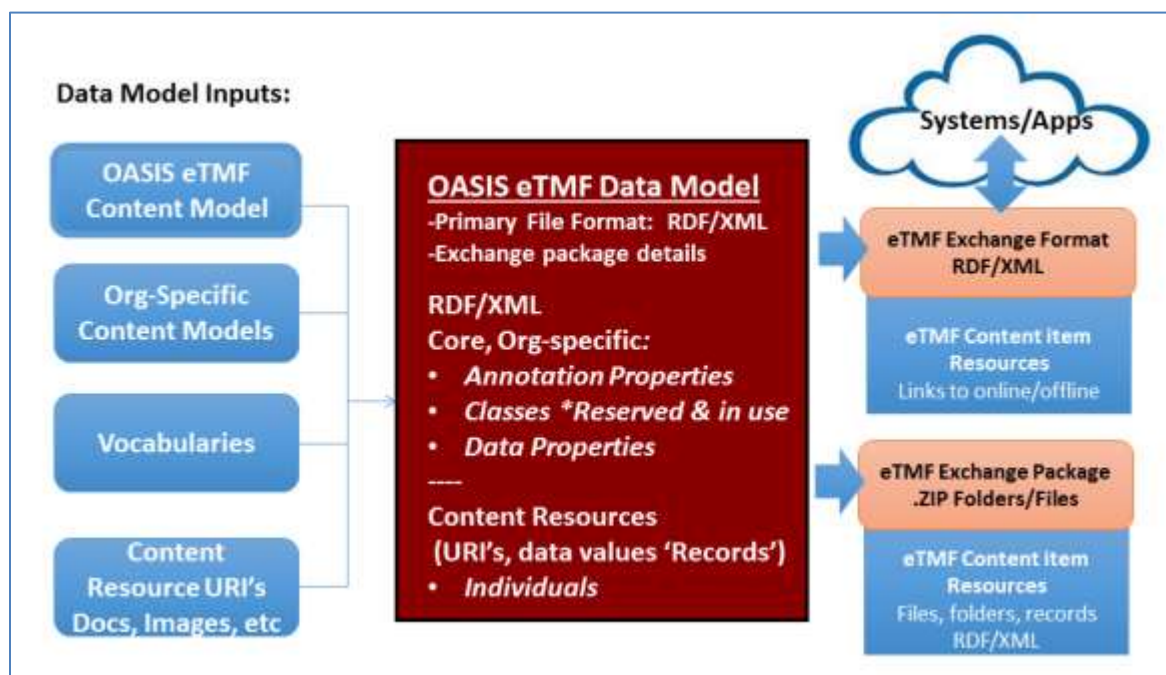

Figure 18: Data Model inputs, file format; eTMF Exchange format and package

6.1.2 OASIS eTMF Exchange Package

The folder taxonomy includes Categories, Sub-Categories, and Content Types as a collection of folders, where each content type folder includes its relevant content items (e.g., documents, images, etc). All Oasis eTMF content items can be exchanged as a collection of folders (following the folder taxonomy by default) and with links to content items and records using an RDF/XML format. The Oasis eTMF exchange package allows creating a data package to export content items and records in a Zip file, as shown in Figure 19.
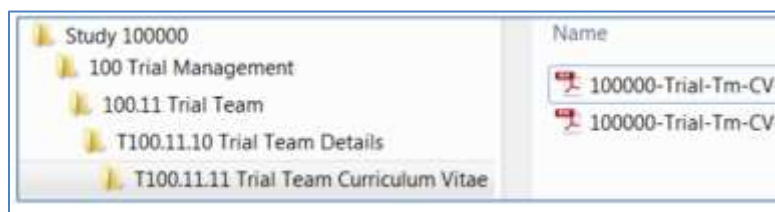
Figure 19: Oasis eTMF Exchange Package Example – Zip File with Folders and Content items

Folders and resources are packaged in a standard .ZIP file, with or without encryption. Additionally, the OASIS eTMF standard supports an **Alternate Taxonomy**. Alternate Taxonomy means that alternate names or languages can be used for classification category names and content type names for the exchange package. The Alternate Taxonomy is supported through the use of **Display Name** annotation property. The **Display Name**   annotation property allows the use of any language and names for the exported taxonomy. It can be used for any category and content type for exchange. Interoperability is enforced with RDF/XML. Figure 20 shows an example for the Alternate Taxonomy for the eTMF Exchange Package folder names.

The eTMF exchange file format has a default structure and naming and supports any IANA media type; thus, enabling broad flexibility. The eTMF exchange file format will be an RDF/XML file with records and URI pointers to linked content online or offline.

The difference between the Oasis eTMF exchange format and the Oasis eTMF exchange package is that the former is a live exchange format for real-time/cloud, while the latter is for ZIP or offline exchange (see Figure 18)
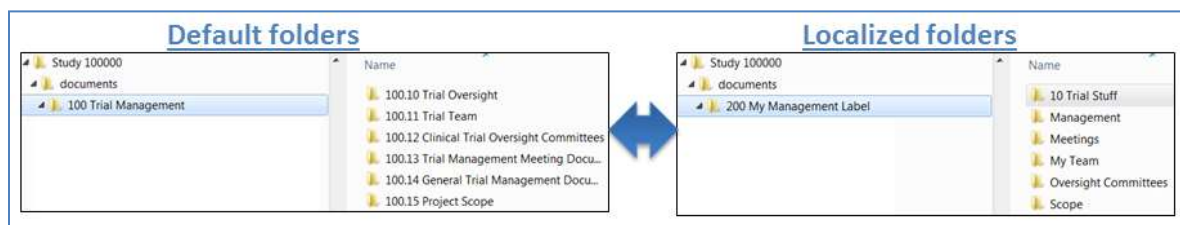

Figure 20: Alternative Taxonomy Example

6.2 Electronic and Digital Signatures

[Non-normative text]

Digital signatures and some electronic signatures can eliminate the need for wet signatures on paper. At publication time, both electronic and digital signatures are acceptable by the EMA and FDA.

One difference between the two signature types is that verification of the signing party is more easily accomplished using digital signatures. The European Medicines Agency (EMA) will start to use digital signatures systematically in outgoing documents that currently require a legally binding signature [4]. This will start with documents related to scientific advice for human medicines, orphan medicines, and pediatric-medicine procedures.

Digital signatures use a digital certificate issued to the signing party and incorporate standards based technology (x.509 PKI). A digital certificate is a cryptographic artifact. It contains a cryptographic key using RSA or Elliptic Curve technologies, where the private key is not discernable by knowing one's public key. The subscriber gets a public key and a private key.

The digital certificate technology is known as the Public Key Infrastructure (PKI). Currently, the FDA requires a digital signature for eSubmissions and a trusted digital certificate on the documents themselves. There are multiple types of digital certificates that are trustworthy. An electronic document, such as a PDF, that is digitally signed using a digital certificate from a recognized certificate authority is more reliable than a wet signature on a paper.

As of September 2013, the EMA moved to require only digital certificates from recognized certificate authorities for three types of submissions: orphan medicines, pediatric submissions, and scientific advice [5]. [Normative text]

Digital certificates are issued to individuals, groups, and devices. In the context of the OASIS eTMF standard content model, digital certificates are used to sign electronic documents. Digital certificates must satisfy the EU qualified certificate policy and the EU advanced electronic signature directive. As of the date of this document, there is a complete overlap between these EU requirements and the FDA PKI policy.

[Normative text]

Digital signatures enable validation of the signing party through digital certificate technologies using a third party certificate authority validation website. Contrary to electronic signatures, digital signatures will be validated in accordance with x.509 digital certificate application standards regarding the requirement for passwords to be used in digital signing. Additionally, digital signing should use the two factor authentication, as per FDA CFR 21 Part 11.

The OASIS eTMF standard supports two types of electronic signatures; electronic signatures (under Part 11 and EMA) and digital signatures. The standard recognizes images with signatures, which can be supported in the OASIS eTMF standard through metadata to indicate documents that contain a scanned signature. Additionally, the Oasis eTMF standard is based on the PKI X.509 V3 digital signing. For digital signatures, person, date/time stamp, and reason for signing are captured and can be extracted from the digital signature within the document. This information is optional in case of electronic signatures.

Generally, electronic signatures should comply with FDA and EU regulations.. Digital signatures should use x.509 PKI certificates, should comply with FDA and EU regulations, and should support any file format approved by FDA and EU for e-signing. While not the purview of version 1.0 of this draft standard, V2.0 of the OASIS eTMF standard should support EU compliant Digital Signatures per emerging EU regulations.

6.3 Business Process Model

[Non-normative text]

Many organizations use automated business processes or workflows for specific operations, such as electronic document approvals. The Business Process Model component provides a mechanism to capture basic business process model task completion information found in metadata linked to documents. The Business Process Model is based on the Business Process Model and Notation (BPMN) V2.0 specification, which is maintained by the Object Management Group [7]. BPMN is a standard for business process modeling, i.e., the representation of processes in an enterprise, with an objective to support business process management. The BPMN V2.0 includes two elements: Business Process and Task. A Business Process is a collection of related tasks that produce a specific output, e.g., a service or a product. A Task is a unit of work that cannot be broken down into a further level of business process details. An example is provided in Figure 21 to illustrate Business Processes and Tasks. In this figure, the Business Process "Sign and Review" is split into three Tasks: Send email request for Signature, Sign Document, and Review Complete.

[Normative text]

The aforementioned BPMN elements are supported in the OASIS eTMF Business Process Model component. They are mapped to their equivalent OASIS eTMF model Business Process Metadata (BPM) terms. For example, whenever a process, or a specific task in the process, is completed for a particular content item, a date/time stamped entry is captured in the OASIS eTMF model BPM, forming an auditable document history log (see Figure 22 for an example). The OASIS eTMF model Business Process component utilizes XML to capture the details of BPMN processes and tasks. No specific software, system, or language is required to implement it.

The OASIS eTMF model BPM enables capturing details about business processes and tasks associated with a specific document. It also allows organization and person entities to be associated with processes and tasks. The OASIS eTMF model BPM includes Process, Task, Content Type, Organization, Person Name, Source, Digital Signature, and Date. The OASIS eTMF model BPM can be captured for any Content Type.

When used in conjunction with digital signatures, the OASIS eTMF model Business Process Component offers automation of paper-based document approval and signing processes. Figure 21 shows an example of a BPMN V2.0 process for an automated signature approval on a confidentiality agreement. Figure 22 illustrates how the BPMN signature process example in Figure 21 is mapped into BPM, with a date/time stamp captured to indicate completion of each task. By capturing the date/time stamp in the BPM for each completed process and task, each document resource has its own auditable workflow metadata history; this enables detailed auditing and reporting in applications. For each Content Type that uses BPM, a new entry is made in the BPM history log for each task in a process. For example, one of the tasks is Send email request for signature, which is associated with the Content Type 'Confidentiality Agreement'. This information is captured in the OASIS eTMF model BPM history log as shown in the figure.
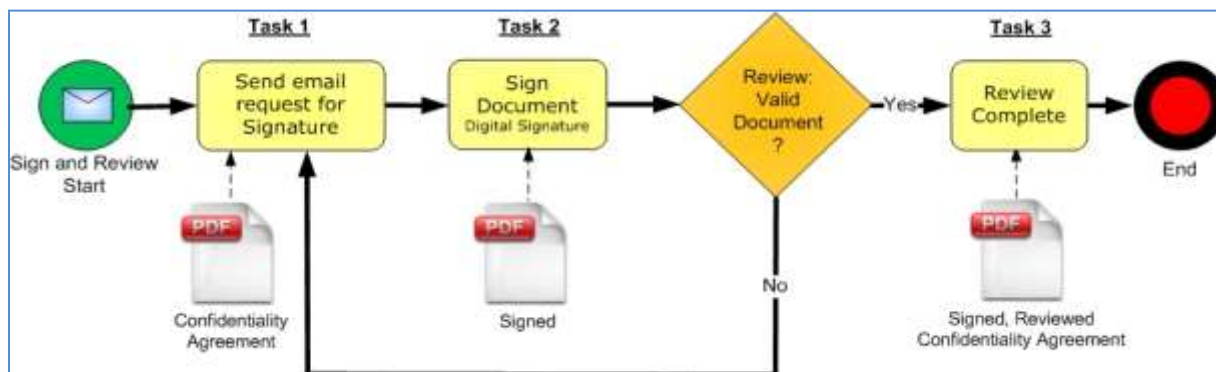

Figure 21: Business Process Management "Sign and Review Process" Example Using the BPMN Notation

Date	Process	Task	Content Type	Source	Person Name	Organization	Digital Signature
10/25/2011 3:07:01 PM	Sign and Review	Send email request for signature	Confidentiality Agreement	Email	Admin	PharmaCorp	
10/26/2011 12:15:52 PM	Sign and Review	Sign Document	Confidentiality Agreement	Web	Lori Albright, MD	Albright Medical	http://sign.pharmacorp.com/lalbright
10/26/2011 12:45:10 PM	Sign and Review	Review Complete	Confidentiality Agreement	Web	Admin	PharmaCorp	

Figure 22: Example for The OASIS eTMF BPM Audit Trail Log Captured for Tasks in Figure 21

7 Conformance

An implementation is a conforming eTMF Content Classification System if the implementation meets the conditions:

a) Conforms to specifications detailed in the OASIS eTMF *Content Classification System (CCS) Layer (Section 5),* with support for the Content Classification System, the RDF/XML [6] based Content Model and includes support for the Metadata.
  I.   Conforms to CCS specifications for naming, numbering and organizing content classifications (Sections 5.1.1, 5.1.2, and 5.1.3).
  II.   Conforms to CCS specifications and policy rules for modifying and editing content classification entities (Section 5.1.4).
  III. Includes metadata tags for all content classified in the system (Sections 5.2).
  IV.   Produces validated RDF/XML code per V1.1 of W3C RDF/XML [6].

b) OASIS eTMF *Metadata Vocabulary* spreadsheet sources all eTMF classification terms from the National Cancer Institute's NCI Thesaurus term repository, which are replicated for reference in the Oasis eTMF Metadata Vocabulary spreadsheet [7], [Oasis: reference for Oasis eTMF Metadata Vocabulary spreadsheet].
  I.   Supports the *Metadata Vocabulary Interoperability Layer* requirement that classification entity terms must have a user-modifiable display label to enable localized terms to end-users (Section 4).

c) Conforms to specifications detailed in the OASIS eTMF Web Standard Core Layer (Section 6).
  I.   Supports content exchange through the RDF/XML based data model component. (Section 6.1).
  II.   Content exchange via RDF/XML must be a valid RDF/XML code per V1.1 of W3C RDF/XML [6].

Appendix A. Acknowledgments

The following individuals have participated in the creation of this specification and are gratefully acknowledged:

Participant:	Organization:
Sharon Ames	NextDocs
Michael Agard	Paragon Solutions
Jennifer Alpert Palchak	CareLex
Peter Alterman, PhD	SAFE-BioPharma Association
Aliaa Badr	CareLex
Lou Chappuie	SureClinical
Sharon Elcombe	Mayo Clinic
Chet Ensign	OASIS
Robert Gehrke	Mayo Clinic
Troy Jacobson	Forte Research Systems, Inc.
Rich Lustig	Oracle
Christopher McSpiritt	Paragon Solutions
Jamie O'Keefe	Paragon Solutions
Oleksiy (Alex) Palinkash	CareLex
Fran Ross	Paragon Solutions
Catherine Schmidt	SterlingBio
Zack Schmidt	SureClinical
Mead Walker	Health Level Seven, Inc.
Trish Whetzel, PhD	SureClinical

Appendix B. OASIS eTMF Terms

B.1 OASIS eTMF Classification Terms

Classification Category terms, used in the OASIS eTMF Content Model, are sourced from NCI [4] using the CareLex Preferred Term, as approved by the Oasis eTMF Technical Committee. Classification terms are published online and curated by NCI.

B.2 Content Item Numbering Policies

In the OASIS eTMF Standard, the document version text values follow the same formatting that is familiar and commonly implemented in software and in other health science standards: **Major Version.Minor Version**. Version numbering text is an integer value separated by a period, without leading zeros. There can be a new **Major version** every time the document/content item changes and   a new **Minor version** every time the metadata changes. .

Within the OASIS eTMF archives, content item version management shall be application specific to provide application flexibility. However, for consistent content item exchange, version number text formatting should be implemented using the OASIS eTMF content item numbering policies (based on NCI/CDISC/FDA/BRIDG definition: **C93816**) as follows:

1. Each document **Major version** number is an integer starting at '1' and is incremented by 1. The first instance, or the original content item, should always be valued as '1'. The version number value must be incremented by one when a document is replaced or its content is modified, but can also be incremented more often to meet application specific requirements.

2. Different versions of the same document belong to the same Content Type group.

3. The document **Minor version** number would be an integer starting at '0' and incrementing by 1. The first instance of an original document with no minor version should always be valued as '1.0', where '0'indicates that no minor version exists.

4. Documents with a change to the **metadata values** would require a minor version. The first minor version for a 1.0 document would be indicated as 1.1. Successive changes to any of the document's metadata would increment the Minor version by 1. For example 1.2 indicates major version 1 and minor version 2.   The Minor version number value must be incremented by one when a document's metadata is changed, but can also be incremented more often to meet the application specific requirements.

Finally, it is worth mentioning that the Content Item version can be different from that of the actual document version. Additionally, the same document can be sent with different metadata and/or content and this condition is supported (not an error condition) by the Oasis eTMF standard.

Appendix C. Glossary

Term	Definition	Reference
EDMS	Computer-based applications dealing with the management of documents throughout the document life cycle	https://www.iso.org/obp/ui/#iso:std:iso:tr:14105:ed-2:v1:en
eTMF	A formalized system of organizing and storing digital content such as documents, images, and other digital content items in an electronic record management system for clinical trials. The Term TMF encompasses documents related to strategies, methods, tools and personnel engaged throughout the lifecycle of the clinical trial regulated content. Items in a TMF may be required for compliance with government regulatory agencies.	http://ncit.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&version=14.10d&code=C115785&ns=NCI_Thesaurus&key=1804103479&b=1&n=null
eCTD	The eCTD is a global specification produced by the International Conference on Harmonisation (ICH) for electronic submissions of clinical trial data. It is used by regulatory agencies such as the US FDA and European Medicines Agency	http://www.ich.org
NCI	National Cancer Institute: is part of the National Institute of Health (NIH) and coordinates the U.S national Cancer program and conducts and supports research, training, health information dissemination, and other activities related to the causes, prevention, diagnosis, and treatment of Cancer.	http://www.cancer.gov/
NCI EVS	NCI Enterprise Vocabulary Service: provides terminology content, tools, and services to code, analyze, and share cancer and biomedical research, clinical, and public health information.	http://evs.nci.nih.gov/
Presentation Layer	The presentation layer (or tier) in n-tier architecture is the tier in which users interact with an application.	http://msdn.microsoft.com/en-us/library/bb384398.aspx

Appendix D. Oasis eTMF Audit Log

[Non-normative text]

Audit logs are time stamped records generated as a result of numerous system and user events affecting content items. Metadata properties Created, Modified, Created By, and Modified by are basic audit information associated with every document or content item to provide general information about the user who created and modified a content item. These metadata properties do not provide sufficient auditing information about system or user modifications to content items. For instance, when user modifies metadata properties of a document or its content, these metadata properties are updated to reflect the date/time for the modification event. However, it does not provide detailed information about the applied modification such as person name, system ID, description of the modification, etc. Additionally, metadata properties generally reflect the most recent change applied to a content item but not the list of all applied changes. The lack of detailed audit log makes it difficult to track system and user modifications to content items. This tracking information is critical for clinical trials' external auditing. The Oasis eTMF standard provides a comprehensive list of attributes that describe an audit log record.

D.1 Oasis eTMF Audit Log Attributes

Each audit log record is composed of several attributes (listed in Table 8) to provide comprehensive information about the content item modification event. Some of these attributes are optional, while others are assigned values if the changes are applied to the content item metadata properties. Figure 23 illustrates an example for the Oasis eTMF Audit Log.

Table 8: Oasis eTMF Standard Audit Log Attributes

Attribute	Description and Format	Requirement
Timestamp	GMT +/- local offset – 00:00:00 +[offset] (standard format xml definition)	Required
Creator Name	First, Last Name of person making the change	Required
Username	System username (e.g., joesmith, joesmith@instance.issuer.com, or other text)	Required
System ID	ID of the system on which the changes occurred. The ID format is [app-instance-name.issuer.com]. The **app-instance-name** section of system ID uniquely identifies a specific instance of the same application within an organization.	Required
Content Item UUID	A universal unique identifier (UUID) using RFC 4122 and 128 bit ID of content item	Required
Content Item Name	Name of the content item from the referring system	Required
Content Item Version	The automated version number created by the system. It is a free text attribute that support any content item version format (e.g., 10-01, 10, 10.1, or 10.0001A)	Required
Type of Change	Free text representing the action on a content such as Create, Modify, Delete, etc.	Required

Change Detail	Free text representing the change applied to a content item, such as Signed, Reviewed, Annotated, Edited, etc.	Optional
MD Attribute Changed	If a metadata property of a content item is modified, this attribute lists the metadata property name.	Required if change is related to metadata
Previous MD Value	If a content item metadata property is modified, this attribute lists the previously assigned metadata property value. This attribute is related to the attribute **MD Attribute Changed**.	Required if change is related to metadata
New Value	If a content item metadata property is modified, this attribute lists the new metadata property value assignment. This attribute is related to the attribute **MD Attribute Changed.**	Required if change is related to metadata
Description	Free text describing the log entry or reason for change.	Optional

Figure 23: Oasis eTMF Audit Trail Record Example

D.2 Oasis eTMF Audit Log Location

- Audit log must be included for every content item within the content item's metadata.
- Every change to a content item must be time stamped and recorded in the audit log, creating a history of changes.
- Audit logs are formatted in RDF/XML V1.0.
- Reference the sample audit log file in the OASIS package

D.3 Oasis eTMF Audit Log Exchange Policy

- Audit logs can be exchanged in real time or via batch methods.
- The protocol or communications method for exchange is up to the implementing system (e.g., https, sftp, or other popular methods).
- Communications error handling is not part of this spec.
- Audit log is exchanged in a separate file.

D.4 Oasis eTMF Audit Log Validation

- An audit log is valid if it passes RDF/XML validation and contains data values for the required audit log metadata attributes.
- Every audit log time stamped entry must have these required audit log metadata attributes.

Appendix E. Revision History

Revision	Date	Editor	Changes Made
R01	April 18 2014	Aliaa Badr	Working draft – first version
R02	May 24 2014	Aliaa Badr	Working draft – apply edits by Jennifer, Rich, and Airat.
R03	June 7 2014	Aliaa Badr	Updated graphics (Figure 6 and 7) to reflect correct classification terms. Added sub-section in the Appendix for OASIS Classification Terms
201406-R01	June 13, 2014	Zack Schmidt	Renamed document and version for public review as committee specification draft
201406-R02	September 14 2015	Aliaa Badr	Committee specification draft – second revision
201607-R03	July 28 2016	Zack Schmidt	Updated cover page, added reference to National Cancer Institute's published eTMF controlled vocabulary terms