GUIDE TO THE DIA EDM REFERENCE MODEL, PROCESS ZONE: REGULATORY-SUBMISSIONS

February 2010
Version 1.0

DOCUMENT HISTORY

Date	Modifier	Version	Description
18 May 2009	Kerri DeRosier	1.0	Started
26 May 2009	Kerri DeRosier	2.0	Revised
27 May 2009	Kerri DeRosier	3.0	Revised
28 May 2009	Antoinette Azevedo	4.0	Revised
5 June 2009	Antoinette Azevedo	5.0	Revised – send to Jim Averback
June 8 – June 12	Kerri De Rosier	6.0	Interview info
June 21	Antoinette Azevedo	6.1	Deleted Pharmacovigilence in future direction section. Reviewed interview notes inserted by Kerri and prepared it for email to Jim Averback.
July 1	Kerri De Rosier	6.2	Created introduction based on team interviews and DIA EDM Reference Model Abstract_draft1; wrote "How to" section; added info to Definition of Terms section based on team interviews and these documents: -Clinical domain release 1.0_2008-07-30.pdf - CMC domain release 1.0_2008-07-30.pdf - Cross domain release 1.0_2008-07-30.pdf - Non-clinical domain release 1.0_2008-07-30.pef - Reg admin domain release 1.0_2008-07-30.pdf - Administrative v4 Glossary.xls
September 14, 2009	Jim Averback	6.3	Filled in blanks in metadata tables Added Single and Repeat Value parameter to metadata tables.
September 30, 2009	Jim Averback	6.4	Added explanation of columns for the metadata tables
October 22, 2009	Jim Averback	6.5	- Accept final changes. - Add introduction to benefits section. - Summarize relationship/dependence on the Common Technical Document and remove repeated references throughout the document. - Removed references to company name for those quoted and domain leads.
October 22, 2009	Kerri De Rosier	6.6	- Proofread document; added new image of

Date	Modifier	Version	Description
			Reference model tabs
October 22, 2009	EDM Reference Model Working Group	6.7	- Finalized future direction section
October 29, 2009	Kerri De Rosier	6.8	- Finalized document
January 18, 2010	Antoinette Azevedo	6.9	- Reformatting to facilitate PDF conversion
February 2, 2010	Antoinette Azevedo	6.10	- Added technical support contact information

TABLE OF CONTENTS

HOW THE REFERENCE MODEL CAME TO BE

Implementing an Electronic Document Management (EDM) system typically costs millions of dollars, months if not years of time for customization, and hiring expensive and hard-to-find IT staff and consultants. For companies that were new to implementing an EDM, the process of creating document repositories often started from zero with no prior reference or industry/vendor best practices.

This tedious work was happening over and over again, company to company – and each one was doing it differently.

Kenneth VanLuvanee described the scenario this way: "Take a box, make sure the box is a uniform size, and everyone paints the same color on the outside and puts whatever they want inside." It was clear that technology alone and current implementation practices were not going to address to need to simplify and align electronic document management for the pharmaceutical industry.

It was time for a better way.

In February, 2008, Dimitri Stamatiadis, John Aitken and Nancy Celini of the DIA Documents and Records SIAC decided to form a working group to develop an electronic document management reference model to help companies design systems for authoring, storing, and publishing documents and data intended to be submitted as an application for drug marketing authorization and other types of submissions. The group – a team of subject matter experts on submission content management and publishing from industry, consulting/professional services, and submission publishing services vendors – had this goal: to develop a "flexible, open, free and sustainable model for document management for the industry from the industry and by the industry."

Consider this scenario: You want to buy a drug or even a whole pharmaceutical company. You begin looking into how the company has maintained its documentation, and find that they don't track attributes the same way yours does – much less use the same attributes and metadata. To get the document management systems to work together, you have to create a mapping structure and invent metadata. "The goal of the EDM Reference Model is to provide a starting point to create harmonization across the industry, making it easier for transactions between companies – because the architecture is the same," said Steve Scribner. "Even if data isn't exactly the same, getting people to work together would be so much easier with a model."

Using the ICH Common Technical Document (CTD) structure as a reference point, the working group created a Reference Model prototype. The CTD was the starting point; however, the Reference Model itself will be extended beyond that to address documents for regulatory submissions independent of market or region and other needs including the trial master file.

By design, "the goal was to use the Reference Model as a way to group documents for internal use, not consumption," said VanLuvanee. "We were trying to build a common way to categorize things – to put things into smaller buckets." The other goal was to reduce redundancy, as many times, the same document is used more than once. "That's where the CTD model breaks down," said VanLuvanee. With the Reference Model, "you're not going to store the same document twice."

A consistent structure across the industry also meant that companies could share, reuse and find documents more easily – internally and externally. "The model is set up to allow for early planning and to have a repository for documents so you can get them when you need them," said Michelle Foster-Herrera.

"Stating the goal of a common framework is good in theory, but unless you put it into practice, nothing will move," said Scribner. "We needed a structure for the vendors to see." Hence, the group presented the model to 12 vendors from various industries in the summer and fall of 2008. The DIA EDM Reference Model version 1.0 was released in the fall of 2008 and demonstrated by selected vendors at the DIA US EDM annual conference in February of 2009. At a panel discussion vendors commented that the model, while not yet complete, could easily be extended to address all of the content required for a market authorization application as well as other regulatory submissions required during the product lifecycle.

The next version of the DIA EDM Reference Model is under development by the working group, and will address additional content such as Prescribing Information and Clinical Trial Master File.

Intended Audience

The intended audience for the DIA EDM Reference Model includes biopharmaceutical companies implementing an EDM system or a simple file share to store and prepare documents for regulatory submissions of drugs and biologic products intended for human use. While the model provides the content, a Regulatory Publishing system is still required to assemble and publish a true regulatory submission. The Reference Model is also useful for biopharmaceutical companies needing to organize and transfer documentation assets of a product to another company's repository under an acquisition or partnership scenario. In this case, the Reference Model provides the common document definition and structure that both companies can share.

Benefits

In addition to the time and cost saved by not having to design a new content model when implementing electronic document management, additional benefits accrue as the Reference Model is more broadly adopted. Key performance indicators for these benefits have been outlined by the working group and are presented below.

EDM Reference Model Key Performance indictors (KPI)

- e-CTD Organization Knowledge
  - Training benefits from the model being similar to the CTD
- Content Architecture Definition
  - Reduction of time to define document taxonomy and metadata
- Submission Auto-Assembly
  - Reduction of time to implement the interface to a regulatory publishing tool
  - Reduction of time to transfer documents from functional area repositories to a regulatory publishing tool
- Trial Master File
  - Reduction of time and cost interacting with CROs
  - Optimized business processes with internal clinical operations groups
- Outsourced Business Processes
  - Optimized business processes with
    - Contract Research Organizations
    - Contract Laboratories
    - Contact Manufacturing Organizations
- Technology Transfer
  - Benefits in document exchange and knowledge sharing between CMC “pre-market” and Technical Operations “post-market” manufacturing
  - Reduced time and cost of content migration from merger, acquisition or divestiture

- High Level Benefits
  - Ability to find information
  - Ability to share information
- Mergers and Acquisitions
  - Better understanding and assessment of the value
  - Easier integration with own company

HOW TO USE THE REFERENCE MODEL

Every pharmaceutical company dealing with regulators in Europe, the United States, Canada, and Japan has to collect documents and data over long periods of time, submit them to regulatory authorities during the investigational phase, and then resubmit the data in marketing applications. The documents are a legal record of quality, safety and efficacy.

Electronic Document Management Systems (EDMS) are a critical component for managing thousands of documents required in a product's life cycle – from the laboratory bench to the pharmacy shelf. Electronic document management can be implemented on technology as simple as a file share on a file server or as sophisticated as a commercial, off the shelf (COTS) electronic document management system optimized for submissions and document records management. It is important to analyze the Reference Model with the EDMS and publishing system's capabilities in mind.

Some mechanism is required to describe the documents contained in any regulatory submission; this descriptive information is maintained as "metadata" in an EDMS. In fact, the Electronic Common Technical Document (eCTD) standard ICH M2 also defines a core set of metadata to be supplied with the submission. The Reference Model addresses the need for this metadata so companies can manage their documents and metadata together to provide the information needed by a regulatory publishing system to assemble and create a regulatory submission.

"Instead of starting with a clean piece of paper, we're saying – here's the structure," said Antoinette Azevedo. "Here are the types of granular elements that need to be written for a CTD, and that need to be updated as a compound moves through the R&D process." The Model also helps companies capture metadata about their project.

Before implementing an EDMS, identify the biggest challenges with managing documents. Is it the system, or is it the business process? Also identify who are the key stake holders and decision makers and what EDMS they're currently using. The next step is to ensure that the pharmaceutical company is familiar with CTD requirements. Go through the CTD table of contents to determine the level of granularity and bookmarks you want to use, as well as which documents to place into each section of the CTD.

All the while, remember that the Reference Model is a model – not a standard for managing submission documents. For example, for a CTD Module 3 submission of Quality information, there are many decisions to be made about the level of granularity. Do you cover Stability in one or two sections? How does the granularity affect your ability to update the document? We recommend that companies seek advice from experts before implementing a major document management system.

At a minimum, your EDMS should:

- Manage content used to prepare paper CTD, electronic Common Technical (eCTD), or Non CTD Electronic Submissions (NeES)

- Capture metadata about the content to pass off to a publishing system or outsource vendor for use in compiling submissions in the development, marketing application, and post-approval phases.

Currently, the Reference Model applies to regulatory submissions comprised of documents required by the CTD specification. In the future, there are plans to extend the model further.

How the Reference Model is organized

We'll start with Ken VanLuvanee's description: "It's a way to group documents for internal use, not consumption (meaning, by the reviewer). You're not using it to build a submission – but as a common way to categorize things – to put things into smaller buckets. For example, in a submission, the same document can be used twice. That's where the CTD model breaks down. You're not going to store a document twice. We're trying to say – here's a consistent way to think about and organize documents."

As such, the Reference Model is not organized by the structure of any particular submission standard. Instead, it's organized into context-sensitive domains (or "buckets") that translate to analogous sections in regulatory submissions. In some cases, the same document is required in multiple locations of a submission; you can look at the Reference Model structure to find out how many times a document is used, and where it goes.

The domains are:

- Regulatory/Administrative

- Prescribing Information (to be included in the next release)

- Pharmacovigilence (to be included in a future release)

- Quality

- Non-Clinical

- Clinical

In the Reference Model, there's a tab for each major category:

The Non-clinical and Clinical Report Components tabs contain supporting documents for the Clinical and Non-clinical domains respectively. These tabs will be covered in a later release of this document.

Let's look at a line from Reference Model to get an idea of how to use it.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Pharmaceutical Content Architecture		Date:												
2	Context of Use:	Regulatory Submission	10/30/2009												
3								Domain Specific -Columns apply to 'Domain' column							
4	Process Zone	Domain	Group	Sub-Group	Sub-Group	Artifact Name		NON-CLIN	Study#	Study Type	Study Director	Dosage Strength	Dosage Form	Test Facility Location	Route Admi
5	Regulatory-Submission	Non-Clinical	Summary documents	Non-clinical Summaries	General										
6	Regulatory-Submission	Non-Clinical	Summary documents	Non-clinical Summaries	General	Nonclinical Overview			M	O	O	O	O	O	O
7	Regulatory-Submission	Non-Clinical	Summary documents	Non-clinical Summaries	General	Nonclinical Summary			M	O	O	O	O	O	O
8	Regulatory-Submission	Non-Clinical	Summary documents	Non-clinical Summaries	General	Introduction			M	O	O	O	O	O	O
9	Regulatory-Submission	Non-Clinical	Summary documents	Non-clinical Summaries	General	Pharmacology Written Summary			M	O	O	O	O	O	O
10	Regulatory-Submission	Non-Clinical	Summary documents	Non-clinical Summaries	General	Pharmacology Tabulated Summaries			M	O	O	O	O	O	O
11	Regulatory-Submission	Non-Clinical	Summary documents	Non-clinical Summaries	General	Pharmacokinetics Written Summary			M	O	O	O	O	O	O
12	Regulatory-Submission	Non-Clinical	Summary documents	Non-clinical Summaries	General	Pharmacokinetics Tabulated Summaries			M	O	O	O	O	O	O
13	Regulatory-Submission	Non-Clinical	Summary documents	Non-clinical Summaries	General	Toxicology Written Summary			M	O	O	O	O	O	O
	Regulatory-		Summary	Non-clinical		Toxicology Tabulated									

RefMod-combined / Regulatory Administrative / Quality / **Non-Clinical** / Clinical / Nonclinical Report Components / Clinical Report Components

There's a color assigned to each domain. In the above example, you're looking at the Non-Clinical domain, which is light blue. The information on the left in the grey columns identifies the type of the document, or taxonomy. The Artifact Name column contains names of documents created for the submission process. The information in the next set of columns (green) contains metadata specific to each domain. The metadata can be used as search criteria in a EDMS.

	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI
2	Entity	Attribute Label													
3	Submission										Product Name	NOTE: At least one required			
4	Region	Submission Country	Application Type	Submission Type	Procedure Type	Transaction	CTD/Section	Project	Submit Date			Chemical Name	Compound ID	Generic Name	Trad Nam
5															
6							2.4					M	M	O	O
7							2.6					M	M	O	O
8							2.6.1					M	M	O	O
9							2.6.2					M	M	O	O
10							2.6.3					M	M	O	O
11							2.6.4					M	M	O	O
12							2.6.5					M	M	O	O
13							2.6.6					M	M	O	O
14							2.6.7					M	M	O	O

RefMod-combined / Regulatory Administrative / Quality / **Non-Clinical** / Clinical / Nonclinical Report Components / Clinical Report Components

The remaining columns (Submission, Product Name, Content Management (not shown) and Document Management (not shown) are process-related, containing information like product name and author, for example. These columns are described in full detail in the following sections.

DEFINITION OF TERMS

A glossary of terms used repeatedly throughout this document is provided below. While some of the terms may be familiar, others such as "Context of Use" are relatively new. This glossary is provided as a baseline to assist all levels of readers in understanding the EDM Reference Model documentation.

Reference Model Terms

Term	Definition
Document Taxonomy	Taxonomy is the practice and science of classification. The EDM Reference model classifies the content of one or more CTD submissions to be managed in a document management system or network file share (“repository”). The taxonomy provided here includes the hierarchical relationship of CTD content as well as attributes or metadata required to find content in the repository. The taxonomy also allows that content to be organized for review in paper or electronic format using metadata defined in the ICH CTD specification. http://en.wikipedia.org/wiki/Taxonomy
Context of Use	Version 1 of the EDM Reference Model is focused on Regulatory Submissions, with the potential for other contexts of use to be added in the future.
Process Zone	Regulatory Submission is the only process zone at this time, with the potential for other process zones to be added in the future.
Domain	The high level name for the functional area or discipline in which a document is created. Currently, the four domains are based on CTD modules: • Administrative, which mostly corresponds to CTD Module 1 • Quality, which corresponds to CTD Module 3 and Module 2.3, Quality Overall Summary • Non-Clinical, which corresponds to CTD Module 4, Safety, and Modules 2.4 and 2.6, Non-clinical Overview and Summaries • Clinical, which corresponds to CTD Module 5, Efficacy, and Modules 2.5 and 2.7, Clinical Overview and

Term	Definition
	Summaries
Group	Categorizes the content of the domain based on logical groupings or CTD modules
Sub-group	Categorizes the content of the group, usually based on CTD module subsets
Region	Geographic regions of ICH; currently EU, USA, and General
Artifact Name	The document created by a user based on the document type selected. Often referred to as the “instantiation” of the document”.
Report Component	A supporting document for Clinical and Non-Clinical domains that isn’t directly tied to a CTD section number.

Domain-specific metadata

For each domain, a table is provided with metadata attributes that are specific to content created in the domain. The columns of the table are defined as follows:

Column Name	Definition
Term	The business name of the metadata attribute as it would be presented to a user. For some technologies, this may be considered the label for the attribute, as in how it is labeled for a user.
Definition	A brief business definition of the attribute's contents. This does not describe any characteristics of the attribute as it is implemented in a system, such as length, data type or validation criteria.
Example	A brief example of the attribute's contents. This should not be taken to define a data standard for the contents, even though in some cases, references are made to standard data sources.
Single / Repeat Valued	An indication of whether in common practice one would expect there to be 1 or multiple values populated to the attribute as it is applied to a single document. There are always exceptions to any rule, but the "S" or "R" value in this column is meant to reflect whether one should expect a single value or multiple values for the attribute in the context of a single document.

Regulatory/Administrative Metadata

The Regulatory/Administrative metadata encompass documents created to support the administration of a Regulatory Application during the full product lifecycle; from molecule/biologic to market through loss of exclusivity.

The documents include forms, cover letters, correspondences, commitments, authorizations, financial disclosures, regulatory submissions and responses to queries. This domain does not include Investigational Brochures, study documents, summary documents, labeling documents, and product update reports.

The attributes provide context to content, such as product information, submission information, document management, and records management.

In the table below, only attributes for Regulatory / Administrative are presented. There are additional attributes common to all documents presented in sections for: Submission Management, Product Name, Content Management and Document Management.

Term	Definition	Example	Single / Repeat Valued
Applicant Address	Business address of the Applicant	Wilson Pharmaceutical Research & Development, LLC, 110 Wilson Way, Newport RI	S
Applicant Name	Name of the sponsoring organization submitting application to conduct clinical trials or authorization to market a regulated pharmaceutical product.	Wilson Pharmaceutical Research & Development, LLC	S
Application Number	Identifier supplied by a Regulatory Authority for reference to an Application made by a sponsoring entity. For the EU, this is the number assigned to the application by the receiving agency. If known to the applicant prior to submission, it can be added.	BB-IND-10940 (IND/IDE Number), 020844 (NDA Number), Health Canada Control Number	R
Application Status	Status with respect to actions of a Regulatory Authority on an application to conduct clinical trials or gain authorization to market a regulated pharmaceutical product.	Approved Not Submitted Pending Approval Rejected Submitted	R

Term	Definition	Example	Single / Repeat Valued
Application Type	Type of application submitted to a Regulatory Authority as required to conduct clinical trials or gain authorization to market a regulated pharmaceutical product. Values dictated by Regulatory authorities.	ANDA – Abbreviated New Drug Application BLA – Biologics License Application DMF – Drug Master File IND - Investigational New Drug Application MAA - Marketing Authorization Application NDA - New Drug Application CTA - Clinical Trial Authorization	R
ATC Code	The Anatomical Therapeutic Chemical classification of the medicinal product. This can be the assigned or proposed code. The top-level code should be used if this code is included in the envelope. ATC Code is a required element in electronic application form for the MAA in Europe: Electronic Application Form: New Application, Specification Version 2.1 March, 2007.	L01CA01	S
Authorization Number	The marketing authorization number of the medicinal product as granted by the regulator in the respective territory/country.	PL00014/0634, O426/0074	R
Concentration	The concentration of active drug substance contained in a single dose of drug product.	1000 IU/ml	R
Dispatch Date	Date on which the submission was sent to an external regulatory authority.	4/Jul/2005	R
Dosage Form	The pharmaceutical form of a drug substance.	capsule, tablet, suppository	R
Dosage Strength	The amount of drug contained in one dose as specified in the label.	200 mg	R

Term	Definition	Example	Single / Repeat Valued
Drug Indication	As described for internal use, the defined disease, syndrome, or clinical condition for which a patient may be treated with a pharmaceutical product. For example, diabetes is an indication for insulin. Once a drug is marketed for one approved "primary" indication, the manufacturer can add "secondary" indications for the drug by meeting requirements for the additional indications.	Epilepsy, Migraine, HIV	R
Fill Volume	The volume of a liquid containing the active drug substance in a single dose drug product.	.3 ml, .5 ml	R
Generic Name	Generic name of the product as approved by USAN and/or INN.	sodium chloride, paracetamol, acetaminophen	R
Health Authority	Name of the country-specific health authority responsible for ensuring the safety and effectiveness of drugs, biologics, vaccines, and/or medical devices.	FDA, EMEA (European Union), Health Canada,	R
Label Format	Format of the label.	SPL, PIM, …	R
Manufacturer Name	The business entity engaged in making, assembling, processing or modifying devices; mixing, producing or preparing drugs in dosage forms by encapsulating, entableting or other process; or packaging, repackaging or otherwise changing the container, wrapper or label of any package containing a drug or device in furtherance of distributing the drug or device from the original place of manufacture to the person who makes final delivery or sale to the ultimate consumer.	Wilson Pharmaceutical Research & Development, LLC...	R

Term	Definition	Example	Single / Repeat Valued
Manufacturing Site	The site of record at which the drug, medical device or component of such is manufactured.	San Juan, Puerto Rico	R
Market Authorization Holder	Legal entity responsible for releasing regulated pharmaceutical products into the market.	WonderPil Pharmaceuticals	S
Packaging Component	The name of a component used in the packaging of a clinical supply or marketed drug.	vial, container label, carton	R
Procedure Type	Procedure used for submission to the EU.	Centralized, Decentralized, Mutual Recognition, National	R
Region	Region within which the document is intended to be used.	European Union, North America, Africa European Union, AsiaPAC	R
Related Sequence Number	The sequence number of any related submissions in the application.	0000, 0001	R
Renewal Date	Date when MAA must be renewed (for EU submissions only).	1-Apr-09	R
Route of Administration	The path by which a drug is brought into contact with the body.	oral, intravenous, subcutaneous	R
Sequence Number	The sequence number of the submission (for eCTD submissions).	0000, 0001, 0002	R
Submission Date	Date on which the submission was acknowledged as being received by the recipient.	4/Jul/2005	R

Term	Definition	Example	Single / Repeat Valued
Submission Description	Brief description of the purpose of a submission.	Protocol Amendment, Follow up measure to provide a toxicology report, ... Advisory Committee stage of review (while under HA review) Review Complete (while under HA review) Questions Received (while under HA review) Response Submitted (while under HA review)	R
Submission Type	The type of a regulatory submission describing its purpose. Submissions are usually related to a specific regulatory application (e.g., NDA, MAA) but may also be independent of a specific application for a marketed product.	original-application, amendment, resubmission, presubmission, annual-report, establishment-description-supplement, efficacy-supplement, labeling-supplement, chemistry-manufacturing-controls-supplement	R
Submission Country	Country representative of the Regulatory Authority a submission was sent to and the country in which a product is approved for marketing.	US, Germany, Australia, Japan	R
Trade Name	The proprietary name of a drug product designated for regulatory approval/marketing. Same as brand name.	WonderPill	R

Quality Metadata

The Quality domain contains CTD Module 3 CMC (Chemistry, Manufacturing, and Controls) documents, the Module 2 Quality Overall Summary produced during all phases of clinical development (IND, NDA, BLA, etc.) and post-marketing documents. The types of documents include those produced during the course of product development that are submitted to regulatory agencies in clinical or marketing applications (initial, amendments, and supplements/variations), and documents and other files created for use in specific regulatory submissions.

Those expected to use the Quality information include:

- Regulatory Reviewers (agencies and internal reviewers, Reg Affairs)
- R&D Contributors and Authors; Functional Areas
- Manufacturing and QA (GMP)
- Regulatory Operations/Publishing
- Project Team and Project Managers

For a CTD, we recommend that you employ the finest level of granularity adopted in the model. You can use further granularity such as sub-sections or attached reports and submit them as separate documents.

The Quality documents are divided into three major categories:

1. Module 2 Summary documents, which are created for a specific regulatory submission
2. Module 3 Quality documents, which are CMC documents filed in original applications and amendments and supplements, according to CTD/CTD granularity. Representative supportive documents associated with each CTD are listed for further expansion of the model (outside the scope of submission documents).
3. Module 3 Literature References

In the table below, only attributes for the Quality domain are presented. There are additional attributes common to all documents that are presented in other sections for: Submission Management, Product Name, Content Management and Document Management.

Term	Definition	Example	Single / Repeat Valued
DS Name	The name of the active ingredient in a drug product. Typically the generic name is used once it has been registered.	sodium chloride, acetaminophen	R
Manufacturer	The business entity engaged in making, assembling, processing or modifying devices; mixing, producing or preparing drugs in dosage forms by encapsulating, entableting or other process; or packaging, repackaging or otherwise changing the container, wrapper or label of any package containing a drug or device in furtherance of distributing the drug or device from the original place of manufacture to the person who makes final delivery or sale to the ultimate consumer.	WonderPill Pharmaceutica	R
Formulation	The specification of dosage form, dosage strength and administrative route for an investigational or marketed product. Often associated by the manufacturer with a unique formulation identifier.	[dependent on the manufacturer]	R
MFG Process	The process for manufacturing a finished drug product or any of the intermediates or raw materials required as a component of the drug product. All drugs manufactured for human consumption must comply with relevant manufacturing regulation and practice generally referred to GMP or Good Manufacturing Practice in the United States.	[dependent on the manufacturer]	R
Impurity	An identifiable and measurable constituent of a manufactured drug substance or drug product that is not intended or required. Acceptable levels of impurities are identified in specification documents for the material, and analytical	reagents, ligands, catalysts, heavy metals, inorganic salts, etc...	R

Term	Definition	Example	Single / Repeat Valued
	methods or assays are defined to evaluate the concentration of the impurity in the manufactured product.		
Batch Number	A unique identifier associated with a manufactured batch of any constituent of a product.	[dependent on the manufacturer]	R
Container Closure	The container of the drug substance or drug product used for the investigational or marketed product. Specific vocabularies for container closure values are defined; see example.	FDA SPL Package Type list http://www.fda.gov/ForIndustry/DataStandards/StructuredProductLabeling/ucm163380.htm BOTTLE, GLASS, BOTTLE, PLASTIC	R
Controlled Document #	Sponsor- or manufacturer-defined number associated uniquely with a document for which change control is implemented. Controlled Document# is usually printed on the cover page of a document, available as metadata and is intended to provide an unequivocal way of associating hardcopy with electronic versions of a document.	[dependent on applied document management nomenclature]	S
DP Name	A name to uniquely identify the product ready for packaging and use. Depending on the product, packaging and/or formulation may be indicated in the drug product name.	WONDERDRUG 50/50MIX 100U/MLX 1 10ML VIAL AM, WONDERDRUG	R
Dosage Form	The form in which a drug product is provided for a defined administration route. There are many controlled vocabularies for dosage form.	FDA SPL Dosage Forms list http://www.fda.gov/ForIndustry/DataStandards/StructuredProductLabeling/ucm162038.htm	R
Dosage Strength	The amount of active ingredient provided in a dosage form specified in terms of a defined unit of measure.	FDA SPL Units of Measure and Units of Presentation list http://www.fda.gov/ForIndustry/DataStandards/StructuredProductLabeling/ucm162049.htm	R

Non-clinical Metadata

The Non-clinical domain contains CTD Module 4 submission documents covering non-clinical study reports, data and literature references, as well as the nonclinical overviews and summaries located in Module 2.

This is information that would be important for a nonclinical or regulatory affairs person to know about the document without having to read it. For example, the reviewer might want to view the descriptive information to make the read/don't read decision, or look at only the studies associated with specific information.

The types of documents include:

- Documents submitted to agencies in initial applications, amendments, or supplements/variations produced during the course of nonclinical research
- Documents and other files created specifically to support a nonclinical report and submitted in a regulatory application
- Documents and other files created for use in specific regulatory submissions

The documents can be organized into the following categories:

1. Module 2 Summary documents created for a specific regulatory submission or submissions
2. Module 4 Nonclinical study documents, reports created as a result of nonclinical studies that are not specific to a submission; data generated during the course of nonclinical studies that is also not specific to a submission
3. Module 4 Literature References

The table below only presents attributes for the Non-clinical domain. There are additional attributes common to all documents which are presented in sections for: Submission Management, Product Name, Content Management and Document Management.

Term	Definition	Example	Single / Repeat Valued
Study #	A unique identifier associated with a nonclinical study as provided to Module 4 artifacts. Usually single valued, but may be associated with one or more synonyms, for example,	[dependent on the manufacturer]	R

Term	Definition	Example	Single / Repeat Valued
	when an outsourced study has both a contract laboratory study number and an internal number maintained by the drug sponsor.		
Study Type	A categorization of nonclinical studies identifying the kind of dosing or kind of assessment being made.	Absorption, Cytotoxicity, Phototoxicity, Single Dose Toxicity	S
Study Director	The person responsible for conducting a nonclinical study and findings. For a study that is in compliance with GLP regulation, the study director is identified within the formal study report as having approved the document and any translations of such.	[name of the individual]	S
Dosage Strength	The amount of active ingredient provided in a dosage form specified in terms of a defined unit of measure. For example, the amount of active ingredient contained in one dose as specified in the label.	http://www.fda.gov/ForIndustry/DataStandards/StructuredProductLabeling/ucm162049.htm	R
Dosage Form	The form in which a drug product is provided for a defined administration route. There are many controlled vocabularies for dosage form.	For example: film-coated tablets, oral solution, tablet, capsule, etc. FDA SPL Dosage Forms list http://www.fda.gov/ForIndustry/DataStandards/StructuredProductLabeling/ucm162038.htm	R
Test Facility Location	The address of a facility that conducted testing nonclinical testing.	[address]	R
Route of Administration	The path by which a drug is brought into contact with the body.	Oral, Intravenous, Intramuscular, Intraperitoneal, Subcutaneous, Inhalation, Topical	R
Duration	Duration of exposure to the drug for in vivo testing. The International Committee on Harmonization (ICH) M2, "The ICH eCTD Valid Values Description" document, specifies the values of Duration to be: Short, Medium and Long. Companies typically manage an internal vocabulary in terms of days, weeks or	Short, Medium, Long, 2 weeks, 12 weeks	R

Term	Definition	Example	Single / Repeat Valued
	months.		
GLP	Good Laboratory Practice (GLP) is a term describing regulation and practice standards in nonclinical testing. Good Laboratory Practice embodies a set of principles that provides a framework within which laboratory studies are planned, performed, monitored, recorded, reported and archived.	Good Laboratory Practice (GLP) for Non-clinical Laboratory Studies, OMB No. 0910-0119, Supporting Statement Docket Number 2004N-0296	S
Species	The International Committee on Harmonization (ICH) M2, "The ICH eCTD Valid Values Description" document, specifies the values of Species to be: mouse, rat, hamster, other-rodent, rabbit, dog, non-human-primate, other-non-rodent-mammal, non-mammal. Companies typically manage additional internal vocabularies for species that are more specific.	mouse, rat, rabbit, dog, non-human-primate, non-mammal	R

Clinical Metadata

The Clinical domain contains CTD Module 5 clinical study reports and supporting documents, plus the clinical overview and summaries from modules 2 and 5. This includes clinical documents and other file types listed in ICH CTD guidance, ICH E3 CSR guidance, and FDA STF specification.

The Clinical domain excludes documents with clinical content in the regulatory administrative domain, such as the Investigator's Brochure, General Investigational Plan, Investigator sign Form FDA 1572.

The types of documents in the Clinical domain include:

- Those submitted to agencies that are produced during the course of a clinical study (not the full clinical TMF but documents from the TMF that are submitted to agencies)
- Documents and other files collected or created specifically for a clinical study report
- Documents and other files collected or created for specific regulatory submissions
- Documents supporting the clinical development process even though they may not be submitted to authorities, such as Clinical Development Plan, Statistical analysis Plan

The documents can be organized into the following categories:

1. Module 2 summary documents created for a specific regulatory submission
2. Module 5 ISS and ISE documents created for a specific NDA or BLA
3. Module 5 clinical study reports and related information generated at different times during a clinical study. Multiple versions of some documents may be used in a study before the study is over or the study report is final, e.g. protocol.

The table below presents only attributes for the Clinical domain. There are additional attributes common to all documents which are presented in sections for: Submission Management, Product Name, Content Management and Document Management.

Term	Definition	Example	Single / Repeat Valued
Indication	The defined disease, syndrome, or	type 2 diabetes melitus	R

Term	Definition	Example	Single / Repeat Valued
	clinical condition for which a patient may be treated with an investigational or marketed pharmaceutical product. For example, diabetes is an indication for insulin. Once a drug is marketed for one approved "primary" indication, the manufacturer can add "secondary" indications for the drug by meeting requirements for the additional indications.		
Clinical Study #	A unique identifier associated with a clinical study as provided to Module 5 artifacts. Usually single valued, but may be associated with one or more synonyms; for example, when an outsourced study has both a contract research organization study number and an internal number maintained by the drug sponsor.	[dependent on the manufacturer]	S
Clinical Study Title	The actual title of the study as written on the cover page of the protocol. • Applies to all documents/components in the study	A double-blind, placebo-controlled study of the effect of LSD-25 on patients with mild to moderate depression	S
Route of Administration	The path by which a drug is brought into contact with the body.	Oral, Intravenous, Intramuscular, Intraperitoneal, Subcutaneous, Inhalation, Topical	R
Study Phase	Categorization of clinical studies based on the objectives of the study. Generally, Phase 1 studies are designed to determine dosing and safety in healthy subjects; Phase 2 studies are designed to assess safety at recommended doses and to begin to establish primary efficacy; Phase 3 studies are designed to evaluate safety and efficacy at recommended doses in patient populations for the indication being evaluated.	1, 2a, 2b, 3, 4	S
Clinical Study Identifier	A unique identifier provided by a Regulatory Authority when registering a trial. This is distinct from the Clinical Study#; however, they both refer to the same clinical trial. The European Medicines Agency requires the EudraCT number. In the US, the FDA requires	NCT00961480 (US)	S

Term	Definition	Example	Single / Repeat Valued
	trials to be registered on www.clinicaltrials.gov and provides an NCT number.		
Type of Control	Control groups in clinical trials can be classified on the basis of two critical attributes: (1) the type of treatment used and (2) the method of determining who will be in the control group.	placebo, no-treatment-control, dose-response-without-placebo, active-control-without-placebo, and external-control.	S
Site Identifier or (Clinical Site Identifier)	Sponsor or CRO code for a specific clinical trial site where subjects were enrolled and treated. Applies only to CRFs, subject profiles and investigator CVs in m5.	[dependent on the drug sponsor or entity conducting the clinical trial]	R
Subject Identifier (or Patient Identifier)	A sponsor or CRO code identifying the specific subject or patient enrolled in a clinical study.	[dependent on the drug sponsor or entity conducting the clinical trial]	R

Submission Metadata

Submission metadata is required in addition to Regulatory / Administrative metadata required to identify, facilitate assembly and track each submission.

Term	Definition	Example	Single / Repeat Valued
Region	Region within which the document is intended to be used.	European Union, North America, Africa European Union, AsiaPAC	R
Submission Country	Country representative of the Regulatory Authority where a submission was sent, and the country in which a product is approved for marketing.	US, Germany, Australia, Japan	R
Application Type	Type of application submitted to a Regulatory Authority as required to conduct clinical trials or gain authorization to market a regulated pharmaceutical product. Values dictated by Regulatory authorities.	ANDA - Abbreviated New Drug Application BLA - Biologics License Application DMF - Drug Master File IND - Investigational New Drug Application MAA - Marketing Authorization Application NDA - New Drug Application CTA - Clinical Trial Authorization	R
Submission Type	The type of a regulatory submission describing its purpose. Submissions are usually related to a specific regulatory application (e.g., NDA, MAA) but may also be independent of a specific application for a marketed product.	original-application, amendment, resubmission, presubmission, annual-report, establishment-description-supplement, efficacy-supplement, labeling-supplement, chemistry-manufacturing-controls-supplement	R
Procedure Type	Procedure used for submission to the EU.	Centralized, Decentralized, Mutual Recognition, National	R
Transaction	A unique identifier for dispatching information to a Regulatory Authority. A transaction may be a complete submission, a letter, a query or response or any other transfer of documentation to a Regulatory Authority related to an investigational or marketed product.	[dependent on the drug sponsor or entity dispatching the submission]	R

Term	Definition	Example	Single / Repeat Valued
CTD/Section	The section as defined by the Common Technical Document specification to which the document is applied. For an eCTD submission, this is equivalent to the XML element of the eCTD DTD.	Primary Pharmacodynamics, Adventitious Agents Safety Evaluation, Reports of Human Pharmacokinetic (PK) Studies	R
Project	A unique identifier for the regulatory project that is responsible to assemble and dispatch the submission.	[dependent on the drug sponsor or entity dispatching the submission]	R

Product Name Metadata

Product name metadata is required to unequivocally identify the product in both investigational and marketed stages. Product metadata is associated with every document that is product specific and is also transferred to the submission.

Term	Definition	Example	Single / Repeat Valued
Chemical Name	Full chemical name of the product including salt form.	2,3:4,5-Bis-O-(1-methylethylidene)-β-D-fructopyranose sulfamate	S
Compound ID	Unique identifier naming a product, small molecule or biologic. It can be company specific and is used to identify data associated with nonclinical research and testing.	MKA12345, JNZ12345	R
Generic Name	The unique nonproprietary or common name for a drug defined by establishing logical nomenclature classifications based on pharmacological and/or chemical relationships and approved by USAN and/or INN.	sodium chloride, acetaminophen	R
Trade Name	The proprietary name of a drug product designated for regulatory approval/marketing. Same as brand name.	WONDERPIL	R

Content Management Metadata

Content management metadata is required to define the content contained in the document, the parties responsible for and knowledgeable of it and when it was first ready for use.

Term	Definition	Example	Single / Repeat Valued
Author	Person(s) who authored or are responsible for the content of the document.	[author name(s)]	S
Business Owner	Person(s) or role responsible for maintenance of the document at each stage of its lifecycle.	[business owner or role name(s)]	S
Component Name	The name of a part that will be assembled with other parts to create a compound document	body, tables, figures, appendix	S
Contributor	Person(s) who contributed content to the document.	[contributor name(s)]	R
Document Title	Full title as it appears on the cover of the document.	Efficacy and Safety of Aliskiren Administered in Combination With Amlodipine versus Amlodipine alone in African American Patients With Stage 2 Hypertension	S
Issued Date	Date on which the document was released for use.	[a date]	S

Document Management Metadata

Document management metadata is required to manage the document throughout its lifecycle, including information about who created it, and its current status is as a document.

Term	Definition	Example	Single / Repeat Valued
Creation Date	Date on which the document was originally created.	[a date]	S
Creator	Name of person who created the original document. Not necessarily the author.	[a name]	S
Language	Primary language in which the document is written.	English, French	R
Litigation Hold	Indication of whether any changes to the document should be delayed due to litigation.	true or false	S
Modify Date	Date on which the document version was last modified.	[a date]	S
Status	State of the document as defined by the applicable lifecycle process. Each lifecycle has a defined set of states and a document is always in one of them.	Draft, In Review, Approved	S
Version Label	The label presented on a document or content management system's user interface indicating the document's version.	0.1, 1.0, Current	S
Version Number	The current number in a sequence indicating the document's version.	0.1, 1.0	S

FUTURE DIRECTION:

Where do we go from here? This is the very first version of the Reference Model and the authors sincerely hope that it will prove useful to industry users as well as vendors and consultants. We are fully aware that it may not be perfect or totally comprehensive, and that it may even conflict with some of the existing approaches and philosophies. But is a starting point that has the enormous advantage of being consensual among many different parties. Our ambition is to continue both refining and enriching the model in order to give more robust and more consensual guidance to all the players in the business landscape.

For this, we call for the help and participation of all those who believe that progress is simplifying and not complicating, that competitive advantage lies with scientific innovation and not within administrative tasks and that functions like EDM are part of the common ground on which we all try to build better medicines for tomorrow.

Many companies today have gone through the process of renewing EDM tools; this is a unique opportunity for considering the Reference Model, not as an obligatory framework, but at least as an option for comparison and to which we all, users, vendors and consultants, can contribute comments and suggestions.

Beyond the current scope of the Reference Model, several areas have been identified as candidates that could benefit from harmonization in the Pharma industry. These are:

Prescribing Information

Clinical Trial Master File (CTMF)

Pharmacovigilance and Risk Management

Metadata data model (ER diagrams) and relationships

Metadata terminology standards

EDM Reference Model Certification process

RESOURCES

Following are resources for:

- Technical support
- Implementation discussion with members of the DIA EDM Reference Model Working Group
- Request enhancements to the Regulatory-Submissions Reference Model or User Guide

- Report errors or omissions with the Regulatory-Submissions Reference Model or with the User Guide.

Access to Regulatory-Submissions Reference Model

http://www.diahome.org/en/HomePage/EDM+Corner.htm

LinkedIn group (**http://www.linkedin.com**)

DIA EDM Reference Model, Process Zone: Regulatory-Submissions

Email Address

edm.ref.model.reg.submissions@gmail.com