Integrating the Healthcare Enterprise

IHE

5

IHE IT Infrastructure (ITI) Handbook

10

Document Sharing Metadata Handbook

15

Rev. 1.0 – Draft for Public Comment

20 Date: June 13, 2018

Author: ITI Technical Committee

Email: ititech@ihe.net

25

> **Please verify you have the most recent version of this document.** See here for Published Versions and here for Public Comment versions.

Foreword

This is a Handbook of the IHE IT Infrastructure Domain.

30 This handbook is published on June 13, 2018 for Public Comment. Comments are invited and can be submitted at http://ihe.net/ITI_Public_Comments. In order to be considered in development of the subsequent version of the handbook, comments must be received by July 13, 2018.

General information about IHE can be found at www.ihe.net.

35 Information about the IHE IT Infrastructure domain can be found at ihe.net/IHE_Domains.

Information about the organization of IHE Technical Frameworks and Supplements and the process used to create them can be found at http://ihe.net/IHE_Process and http://ihe.net/Profiles.

The current version of the IHE IT Infrastructure Technical Framework can be found at http://ihe.net/Technical_Frameworks.

40

CONTENTS

1 Introduction

90 This handbook serves as a guide to designing, deploying, and maintaining a "Community Strategy" for a well-defined and managed specification for how metadata will be used with the intended purpose of optimizing Document discovery within a Document Sharing system. This process includes defining consensus metadata vocabularies and value sets for items such as typeCode, classCode and practiceSettingCode as well as establishing routine strategies for

95 various common metadata handling, enforcement and life-cycle use cases.

1.1 Purpose of the Metadata Use Handbook

This handbook is intended to assist the reader on the steps necessary to define how document metadata would be used, how to enforce and propagate that use expectation, and how to evolve the document metadata use constraints over time. A well-managed document Community

100 Strategy is an essential component of an efficient and coordinated, registry-focused Document Sharing system and becomes more important as the size of the participating Community and the number of patients and documents under management grow.

This handbook uses the term "Community" as defined in the HIE white paper. This definition is inclusive of an XDS Affinity Domain, an XCA Community, or a set of XCA Communities. This

105 definition is also inclusive of the use of MHD (XDS-on-FHIR®[1]).

This handbook covers the Document Sharing Metadata to enable discoverability. There are other responsibilities of Community management: Patient Identity and Demographics management, Organizational identification and credentialing, Privacy Policy, Security Policy, Partner Certification processes, etc. IHE has a paper covering these broader topics, see Section 1.3

110 "Template for XDS Affinity Domain Deployment Planning".

1.1.1 When to use this Handbook

This handbook is intended to be used in various situations.

- When initially deploying a Community – In this case there are no existing documents published, so the opportunity is to define the expected use of the metadata.

115 - When adding a new document use case to a Community – Any new use cases may drive changes to the Metadata Use Pattern

- When connecting new partners to a Community – This is similar to adding new document use cases, but may bring in many new factors

- When connecting multiple Communities together – This will focus on harmonizing the
120 various Community Strategy

- On a regular basis (roughly between 2-5 years), to re-examine past use case analysis and decisions – this focuses on how the environment or user base has changed.

---

[1] FHIR is the registered trademark of Health Level Seven International.

- As a starting point for national and international XDS metadata harmonization initiatives.

1.1.2 How to use this Handbook

125 This handbook is designed to establish an understanding of the intended use cases and technology limitations of metadata within various scales of Document Sharing Communities. Some important scale considerations include:

- Single Purpose XDS Affinity Domain such as a research project – the Community Strategy may be simpler in this case, but use of this handbook may still bring some value;

130 - A small isolated XDS Affinity Domain used for specific disease or treatment – the use of this handbook focuses on the primary set of use cases, but should also consider future expansion;

- An XDS Affinity Domain that is connected to other Communities using XCA – the scale of the use cases is much larger, and thus there is more focus on alignment of metadata
135 elements;

- An XCA Community that is not natively XDS based – this is similar to an XDS Affinity Domain that is connected to other Communities;

- A set of Communities interacting using XCA – When multiple Communities are interacting, the Community Strategy will focus more on documenting how metadata are
140 used, as it is harder to impose constraints upon them.

1.2 Intended Audience

The intended audience of the IHE ITI Metadata Use Handbook is:

- Those involved in deployment of an XDS Affinity Domain, or XCA Community;

- Those involved in joining multiple Communities using XCA;

145 - Clinicians involved in the use of documents from the Communities, and publishing into their own Community;

- Developers involved in implementing actors supporting the Community.

The intended audience is expected to be well versed in the Document Sharing metadata concept, Vocabulary use, and understand the intended use cases for their Community.

150 ## 1.3 Background References

The following references are considered critical foundational text that should be understood by those participating in the use of this handbook (all links can be found at https://wiki.ihe.net/index.php/Metadata_Handbook).

- HIE using IHE Profiles – defines the basics of Document Sharing and the use of the IHE
155 profiles, http://www.ihe.net/Technical_Framework/upload/IHE_ITI_White-Paper_Enabling-doc-sharing-through-IHE-Profiles_Rev1-0_2012-01-24.pdf

- Section 4 of Volume 3 of IHE ITI Technical Framework – defines the XDS and XCA metadata, http://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_TF_Vol3.pdf#nameddest=4_0_Metadata_used_in_Document_S

- Template for XDS Affinity Domain Deployment Planning – White Paper covering the broader process and policy recommendation for deploying an XDS Affinity Domain, http://www.ihe.net/Technical_Framework/upload/IHE_ITI_White_Paper_XDS_Affinity_Domain_Template_TI_2008-12-02.pdf

- Basics of doing Document Sharing Query right – blog article that explains use of critical few query parameters and using further processing of the resulting document entries to avoid false-positive and false-negative results, https://healthcaresecprivacy.blogspot.com/2018/04/basics-of-doing-document-sharing-query.html

- ALL IHE defined profiles in the Document Sharing family (e.g., XDS, XCA, MHD, etc.) http://wiki.ihe.net/index.php/Category:DocShare

- All IHE defined profiles of CDA®2, http://wiki.ihe.net/index.php/Category:CDA

- European metadata project - initiative to collect and analyze XDS metadata initiatives in 10 European countries and general introduction to categorization topics, 2017, https://ihe-europe.net/sites/default/files/2017-11/IHE_ITI_XDS_Metadata_Guidelines_v1.0.pdf

1.3.1 Samples and ongoing Discussion

IHE as gathered valuesets from various projects. IHE has started a discussion forum to encourage sharing by the community of individuals working on Metadata use and constraints. These and other examples can be found at https://wiki.ihe.net/index.php/Metadata_Handbook

1.4 Comment Process

IHE International welcomes comments on this document and the IHE initiative. They can be submitted by sending an email to the co-chairs and secretary of the ITI domain committees at ITI@ihe.net.   A Community forum of peers also using this handbook is available at ihe-metadata@googlegroups.com

1.5 Open and Closed Issues

Open issues:

- MDH-0: This text is known to be incomplete at Public Comment. We expect that it includes the critical messages for the Handbook. We expect after Public Comment the text will be made more consistent and readable. Your comments and additional

---

2 CDA is the registered trademark of Health Level Seven International.

190   suggestion are welcome, although we invite the reader at this stage to focus on the essence and not on the linguistics.

- MDH-1: to what extent should we recommend use of Metadata Update to correct historic metadata that no longer fits new Community Strategy?

195
- MDH-2: We have covered the few critical metadata elements but have not covered all possible. The first public comment phase is kept minimal to enable this handbook to be approachable and effective. Please comment on other topics that we should cover, and why that topic is critical.

- MDH-3: We did not cover creating a strategy for use of Folders, ReferenceIdList, or Associations. We recognize these as important aspects to include, but don't yet have a
200   good consensus of the "Best Practice". Please comment on the criticality.

- MDH-4: We recognize authorInstitution and sourcePatientInfo are constrained by various projects for specific uses. Such as use of authorInstitution use being consistent with the XUA SAML assertion content, and sourcePatientInfo use being consistent with national patient ID assigned. These are useful constraints within those projects, but it is not clear
205   to what extent they are globally defined Best Practices that his handbook could recommend.

Closed issues:

- MDH-5: typeCode is a value selected by the publisher of a document, expressing its intention for the document. The value set within a Community for typeCode will have to
210   be comprehensive and is likely to contain overlapping codes. There can be preferred use codes within a Community and guidance to reduce overlap. But, in the end, it will be the publisher of a document that choose the best code from their perspective. This is a reality of clinical practice that the publisher must be allowed to pick the typeCode that best expresses their intentions for the document.

215 - MDH-6: classCode is a value intended to assure discoverability success. The classCode should come from a small value set of non-overlapping codes. Ensuring non-overlaps delivers the certainty that no clinically relevant document may be missed.

- MDH-7: Historic metadata is unlikely to be changeable, given the challenged in semantic mapping, so the old vocabulary is most likely to remain visible and valid. There are
220   methods, such as Metadata Update, to revise metadata; however, these methods are discouraged as their use for this purpose becomes very expensive as the longitudinal data grows. Therefore, one must always accept that historic data may not be following current metadata expectation. This means that when looking for document within a time span that covers different metadata definition regimes one may have to query two different ways or
225   be more liberal in the query expectations. This also means that having to redesign/adjust a community strategy for metadata should be avoided, by adopting a forward-looking approach from the beginning.

2   Principles of good document metadata control

The overall use case for Document Sharing makes metadata critical to data discovery. The *Document Source* must carefully use the right metadata to describe the document, the quality and accuracy of the metadata produced by the source is critical to any reliable and efficient access to shared documents. The *Document Consumer* must leverage the Query model, and the resulting list of available files must answer the expectations of the end users. If a Source "misfiles" a record (incorrect creation of metadata), the record may not be found by the Consumer.

To avoid these potential hazards, the following measures must be considered:

- Any Consumer within the Document Sharing Community should be able to accurately interpret the correct meaning of the metadata;

- Harmonization of metadata sets is needed when multiple Affinity Domains want to exchange documents, either through common metadata sets or through mapping;

- Metadata attributes and value set elements should be clearly defined to ensure that sources can publish content accurately;

- A guideline and 'decision tree' are very useful to ensure the correct use of the metadata for entry and retrieval of documents. Disambiguation between semi-overlapping values should be explained and linked to a decision tree.
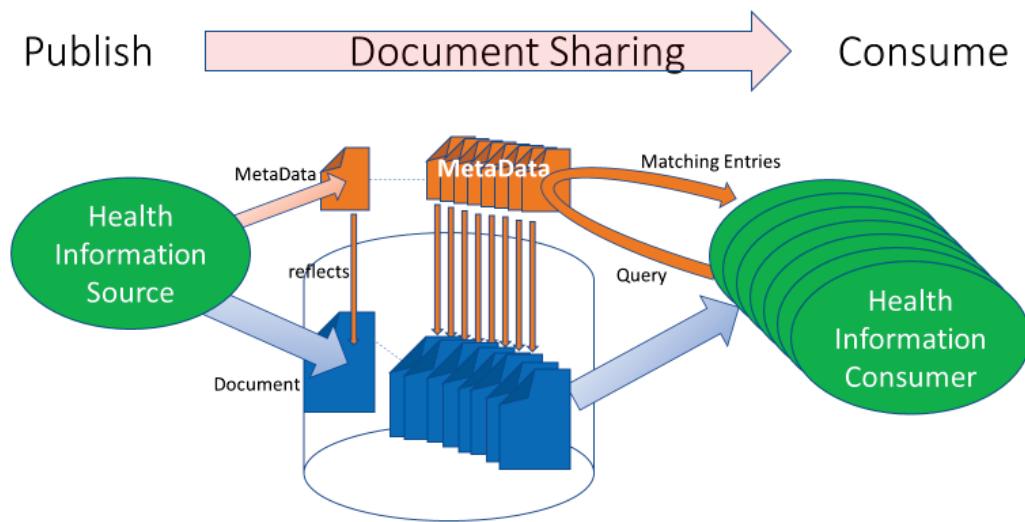
Figure 2-1: Document Sharing Lifecycle

The XDS/XCA Query transaction has a number of stored queries. The FindDocuments is the one that is most powerful.

The other stored queries are not useless but are far more special purposes. Some are more focused on SubmissionSets, Folders, and Associations. These may be useful for specific use cases, but not for a general-purpose Document Consumer.

2.1 XDS/XCA FindDocuments Query Parameters

The FindDocuments query has 18 query parameters, but 5 of them play a critical role. They are the "critical few", especially for the initial query that performs the *primary filtering* (Further discussed in Section 3.4) among all available documents for a patient that may be in the thousands with a mature deployment. The other 13 parameters can also be used, but the use of these additional parameters in any primary filtering is fragile. Fragile in that the consuming system must be in really good synchronization with the publication system, which is unlikely over the timespan of decades. Later we will explain how to deal with this fragility. Here are the 5 query parameters and metadata attributes that are relevant for primary filtering:

- **patientId** - this is required in XDS/XCA. You must have a Patient ID you are interested in. Use of PIX, PDQ, XCPD, or some other Patient Identity Management system is a required prerequisite, that will not be further discussed in this handbook.

- **classCode** - this is the most poorly understood metadata element, yet it was intended to be the most powerful. The idea is a major focus of this handbook, where we explain that a small number of vocabulary terms should be allowed, that group documents into logical 'classifications', where these classifications are useful to the consuming end user. That is, they should be designed (vocabulary design -- value set) so that for any use case where someone is looking for documents they can pick one term from this value set, possibly combine that with another one of the 5 parameters and then get the results where no relevant documents are missing.

- **serviceStartTimeTo** - **serviceStopTimeFrom** - these work together to give a period of time that overlaps so as to give more positive results within which the documents were relevant. The service times are specific to the time range of the treatment or episode. This is different than the creation time, which is when the document was created. So the query result will return any document whose "service time" falls within that range. Important to note that these two parameters work together to give a period of time, and that period of time can only have a start, or only have an end.

- **practiceSettingCode** - this is the clinical specialty where the act that resulted in the document was performed. Like the classCode, this should have been filled with a controlled value set of pre-negotiated vocabulary that represents broad classifications of clinical specialties. By restricting the value set to the high-level clinical specialties, one should avoid the misfiling associated with documents produced by sub-specialties.

2.2 Classification – value sets are critical

285 What the above shows is that two of the critical FindDocuments query parameters (classCode and practiceSettingCode) should come from well controlled value sets, each having a limited number of value set elements that represent broad classifications.

These codes need to be useful and recognizable for the end users. They expect to find documents under a certain 'heading' or category. The more values a value set has, the harder it is to store 290 and retrieve the document under the right category.

So, how does a Community determine what these value sets should contain? That is the main topic of this document.

2.3 Query is not enough – local processing

The purpose of the XDS/XCA Metadata is to enable document/search/filter/group the documents 295 so that the right information can be found without missing any relevant documents. The above four query parameters are necessary, but not always sufficient.

Critical in the "Best Practice" is that a Document Consumer must be ready to do some form of local processing. This local processing would leverage all of the metadata. It might further eliminate unnecessary entries, group and sort the results, put emphasis on some entries, show 300 relationships between documents et cetera. This local processing might be an automated algorithm and/or the possibility to interactively change the order, look for search words, by the end users.

3 Process to define document metadata

305

In this section, the process steps of the definition of an XDS metadata set are described and tips are provided. The following steps will be explained:

1. Identify shared business needs;

2. Gather relevant use cases;

310

3. Gather existing patterns of how metadata are used;

4. Consider the various uses for metadata in the relevant use cases:

5. Determine the needs for constraints;

6. Assemble the metadata elements;

7. Use the guidance in Sections 4 and 5 to address overall and specific metadata

315

    characteristics;

8. Record all decisions and actions into a Community Strategy.

3.1 Identify shared business needs

Collaborating parties and stakeholders in a Community must agree on the need for a shared definition and implementation of a metadata set. Here are the steps to organize this:

320

- Set up a project and working group to organize the processes;

- Define what the business cases are that bring the Community together;

- Define specific shared principles of that Community;

- Achieve agreements: organization, legal, financial, governance;

- Set groundwork for deployment and maintenance;

325

- Engage with regulators, and professional societies.

3.2 Gather relevant Use cases

Make an inventory of the current and planned use cases in the Community and look at use cases in other Communities. Look at what types of documents are used in these use cases, such as:

- Medical

330

- o Referral letters*

- o Intake reports

- o Medication information*

- o History & physical reports

- o Consultation documentation

335
- o Laboratory results*
- o Genomics, biomics, proteomics information
- o Diagnostic test reports*
- o Multidisciplinary board reports
- o Discharge letters*

340
- o Encounter Summaries*
- o Patient Summaries*
- o Surveys/Assessments
- o Advance directives
- o Notes

345
- Workflow related
  - o Referrals*
  - o Care plans*
  - o Appointments
  - o Diagnostic study requests*

350
  - o Timelines
- Registrations
  - o Quality registrations
  - o Outcome measurements
  - o Research studies

355
- Administrative
  - o Demographics
  - o Legal
  - o Insurance / financial
  - o Consents*

360
- Multimedia and Images
  - o Imaging studies*
  - o Medical photographs and movies
  - o Microscopy / pathology studies
  - o Audio and video

365 Some of the above use cases have been flagged with asterisk (*) to indicate that the IHE Community has noticed that these are more widely deployed than others world-wide.

From these use cases, end user- and technical requirements can be set up for the metadata set. Consult with all stakeholders to get a clear view of the expectations of the end users towards the categories choices.

370 3.3 Gather existing patterns of how metadata are used

- Within existing Community there will be patterns of use. These may have grown organically, or follow some existing governance;

- Gather patterns of use from other similar Communities.

375 Note that historic and current use of metadata might have deviated from the current Community Strategy. This is an indication that there is an unmet need, or a misunderstanding of the Community Strategy. These deviations should be investigated carefully to guide the updated Community Strategy.

3.4 Various uses for metadata

Consider the various uses for metadata in the relevant use cases:

380 - Define the use of the metadata for the most relevant "critical few" – see below;

- Define the use of the metadata for the additional elements extending the most relevant "critical few";

- Define the use of the metadata for the remaining elements that are "less critical" (e.g., used for filtering by processing the returned query response matching entries, either by
385 human review of application analysis) - see *secondary filtering* (below).
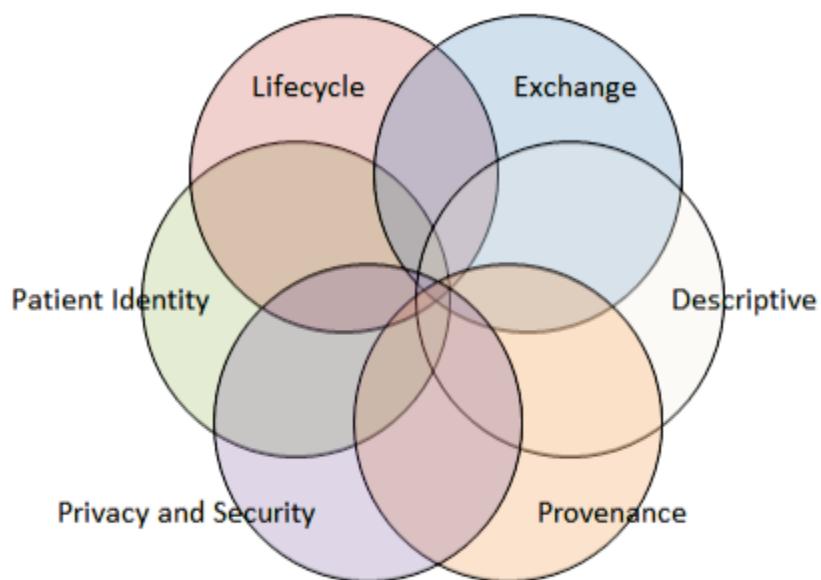
Figure 4.1.3.1-1: Pictorial of Overlapping Document Sharing Metadata Purpose

Note: The above figure is from IHE ITI Technical Framework Volume 3.

390   The IHE ITI Technical Framework Volume 3 contains details on each metadata element, and the various use cases that each assist with. The following snippet from "Table 4.1.3.2-1" is from tables found in the ITI Technical Framework Volume 3. The reader needs to be very familiar with the definitions of the metadata found in ITI Technical Framework Volume 3.

Table 4.1.3.2-1: DocumentEntry Metadata Attribute Definition

DocumentEntry Metadata Attribute	Description	Patient identity	Provenance	Security &Privacy	Descriptive	Object Lifecycle	Exchange
author	The humans and/or machines that authored the document. This attribute contains the sub-attributes: authorInstitution, authorPerson, authorRole, authorSpecialty and authorTelecommunication.		X	X	X		X
availabilityStatus	The lifecycle status of the DocumentEntry			X		X	
classCode	The code specifying the high-level use classification of the document type (e.g., Report, Summary, Images, Treatment Plan, Patient Preferences, Workflow).			X	X		
comments	Comments associated with the document.				X		
confidentialityCode	The code specifying the level of confidentiality of the document.			X			
creationTime	The time the author created the document.		X	X	X	X	
entryUUID	A globally unique identifier used to manage the entry.			X	X	X	X
eventCodeList	This list of codes represents the main such as a colonoscopy or an documented.						
formatCode	The cod						

395 The XDS- (and XDS-I, XD*Lab, etc.) metadata attributes cover a number of domains that focus on a number of axes: patient identity, privacy and security, provenance (authors, senders, location, date and time), processes, purpose (descriptive), document lifecycle information and technical information about the document itself. The why, where, what, when, how and other metadata aspects document the context within which the document was created. This information
400 is vital for the correct interpretation of the information contained in the document. All these metadata attributes can be used, often in combination, for the filtering and selecting of specific information types.

The main purposes for metadata are:

- Documenting

405
- o All metadata attributes define the contextual information about the document (why, what, when, who, …). They are vital for the correct interpretation of the information contained in the document;

- Filtering

Four types of filtering can be distinguished: primary, secondary, tertiary and additional

410
- o Primary Filtering: attributes primarily used for selecting documents and sets of documents submitted together for sharing. This filtering may support a narrowly targeted query (looking for a specific or small set of documents) or a broad query intended to select a manageable set of likely relevant documents;

415
- o Secondary filtering: returned metadata attributes intended to be associated with the documents matching a primary query to enable a human (or application) to filter out among the returned candidate entries, the ones that are relevant and need be retrieved;

- o Third-level filtering: Once the relevant documents have been retrieved, their content may be processed, and relevant information extracted. This third level is important for the user but is not included in the metadata set as metadata are set at the document
420
  level, not at parts of their content;

- o Additional mechanisms

  - ▪ Technical filtering: Metadata attributes critical for the operation of the queries, but generally not visible to the clinical user. They are used for integrity verification, performance management, configuration, etc.

425
- • Grouping

  - o On folders and groups of documents (submission sets) Folders can be used to make a persistent record of a group certain documents that belong together or are in the same category.

  - o Categorization attributes can also be used for grouping documents of a certain type, for instance letters, notes, summaries et cetera. Grouping is a specific form of filtering
430
    based on metadata with the purpose of creating groups of certain types of documents. In many EHR user interfaces, tabs are used to quickly find the right type of document or image, and grouping is used to place the right documents under the right tabs.

- • Sorting

435
  - o Category attributes have predefined value sets. The captions of these value sets can be shown in alphabetic order, or in another predefined order (for instance, showing the most frequently used options first);

  - o Linking of coded value set elements to terminologies allows for additional descriptions for sorting and finding documents and images, even in other languages.
440
    Attributes that have coded lists as options to choose from can also be sorted in different ways by linking the codes to different captions, so that end users can choose the caption list that they are accustomed to;

  - o End-users should be able to interactively sort on different kinds of parameters (date, specialty, functional type, author, department, et cetera).

445
Filtering, grouping and sorting the available documents greatly enhances the discoverability of specific documents. Setting up a recognizable and intuitive set of metadata elements helps the end users in quickly accessing the right information at the right time for the right purpose.

3.5 Determine the needs for constraints

450

Each Community will determine the desired level of constraints. Some Communities may only define a specific value set for classCode and leave all other metadata with the default constraints from XCA or XDS; other communities will need to define constraints for every metadata element across all object types. This is especially the case with document exchange between Affinity Domains.

455

- Where there are rules already in place in a Community, determine how well these are working and what the gaps are. If more Affinity Domains are involved in the Community, what are the differences between the metadata sets of these Affinity Domains? Can the metadata sets be harmonized?

460

- The technology used within the Community may have technical constraints. For example, a publishing system that has fixed codes that can't be changed. These technical constraints are unfortunate but are a reality. It is better to understand them, and work around them.

- Take into account that IHE has predefined the use of metadata elements for some document types:

Table 3.5-1: Sample of IHE Document Content Profiles – with metadata constraints

IHE Domain	Profile Acronym	IHE Profile Links
ITI	XDW	Cross-Enterprise Document Workflow
ITI	XDS-SD	Cross-enterprise Sharing of Scanned Documents
ITI	BPPC	Basic Patient Privacy Consents
ITI	APPC	Advanced Patient Privacy Consents
ITI	DSG	Document Digital Signature
RAD	XDS-I	Cross-enterprise Document Sharing for Imaging
RAD	XCA-I	Cross-Community Access for Imaging
CARDIO	XCHT-WD	Cross Enterprise Cardiovascular Heart Team
CARDIO	CRC	Cath Report Content
CARDIO	CIRC	Cardiac Imaging Report Content
PCC	XPHR	Exchange of Personal Health Record Content Profile
PCC	EDR	Emergency Department Referral Profile
PCC	IC	Immunization Content
PCC	XDS-MS	Medical Summaries Profile
PaLM	XD-LAB	Sharing Laboratory Reports

465

Notes:

1: XDW workflow definitions can be found at http://wiki.ihe.net/index.php/Category:XDW and at http://wiki.ihe.net/index.php/Cross-enterprise_Tumor_Board_Workflow_Definition

2: Certain images can be found at http://wiki.ihe.net/index.php/Category:DICOM

3: All IHE profiles of CDA can be found at http://wiki.ihe.net/index.php/Category:CDA

3.6 Assemble the metadata elements

In setting up a Community metadata set, it is advised to start with the 'easy', non-ambiguous metadata attributes:

- Define codes that are one-for-one descriptive of the document
  - mimeType, FormatCode, typeCode
- Define codes that are descriptive of regions and author roles
  - Author role, practiceSettingCode
- Define expected use and the minimal accuracy for each of the date-time attributes

3.7 Use the guidance in the Handbook to assemble the metadata set

Section 5 of this document provides principles, guidelines and other considerations for the assembly of a metadata set. In the definition process, keep the following action points in mind:

- Define a recommended value set, or a required value set for coded elements
  - It is better to start with small value sets for any coded element. It is far easier to expand a value set than it is to correct historic data that was liberally published;
  - Smaller value sets are less error-prone in the publishing systems, especially when human assistance is needed (human make many errors when having to peruse value sets larger than 20 or 30 values (50 or more values require a two level pull-down which is ergonomically ill-advised);
  - Use of a value 'unknown' or 'other' should be avoided as much as possible;
  - Well defined codes are critical to successful use;
- Define rules for how to behave when populating elements where the value is unknown;
- Define what to do with attributes that the core XDS specification allows to be left empty.
- Often a Community, especially as time passes must have two perspectives for value sets, (1) The value set that is recommended to be use when publishing "new" documents, and (2) The value set that is understood to represent all historic and new documents. This recognizes the need to be very specific in new publications, while recognizing that historic entries are not going to conform to the new Community Strategy rules.

3.8 Community Strategy

The Community Strategy is the plan produced by the use of this Handbook. It includes value sets, policies, and Metadata use expectations. The Community Strategy contains the preferred rules for publishing new documents and the realistic rules under which historic documents will be found. A document sharing environment is longitudinal and will over time contain medical information that spans decades. During this time, changes will be made to the preferred metadata set. The Community Strategy Changes records these changes.

➤ Experience shows that a well-designed Community Strategy is effective for 5 to 10 years.

505

The exact format of the Community Strategy is up to you. Value sets are critical component of a Community Strategy that could be maintained as simplistically as a set of spreadsheets or use purpose specific tools. Often for any metadata element there might be two independently managed value sets: the preferred value set for new publication, and the comprehensive value set holding all possible values for that element representing longitudinal historic record. The important part of a Community Strategy is the information it holds. The Community Strategy captures, at a point in time, the current preferred metadata usage, and the historic metadata usage under which older documents will be found among all participants in the community, to the maximum degree of accuracy.

510

515 Suggested information to capture:

- Value set used, including references to sources. Specify version used. If a value set management system is used (see Section 4), it may be considered part of the Community Strategy;

- Time frame of the value set;

520 - Participants or groups using the value set;

- Common values;

- Most used values;

- Guidelines for the proper use of the metadata.

The Community Strategy must include historic Metadata use that may be seen in queries. This allows participants to deal with this reality via user interfaces, education and other means.

525

4   Deployment and Maintenance

After the definition of a new Community Strategy, deployment of the Community Strategy is a delicate task, and must be managed carefully. We recommend the following general project considerations:

530
- Reach out to others who have gone through this; look at examples of Community Strategies as well as messaging (a community of others can be found at the IHE wiki discussed in Section 1.3.1);

- Manage the deployment and maintenance as a project, including all stakeholders from the beginning (participating organizations, vendors, end users, implementers, payers);

535
- Engage and motivate all affected parties, highlighting the purpose, its advantages and required efforts;

- Maintain lines of communication and mechanisms to ask questions and report problems;

- Manage all transition states (for example, bringing a specific partner into conformance with the metadata usage);

540
- Add some ceremony when milestones are reached – "Have a nice party of it".

For deploying value sets, we offer the following specific suggestions:

- Use a Value set management system, to allow community participants to easily reference and/or consume values;

- Value sets should be published in electronically consumable form (examples can be
545   found on the IHE wiki referenced in Section 1.3.1):

  o Using a spreadsheet;

  o using IHE Sharing Value Sets (SVS) Profile;

  o using tools like ART-DECOR;

  o using FHIR ValueSet resource.

550
- It is recommended to publish the value set including clear definition of each value set element, and guidance on the proper use to avoid ambiguity of choice. This might be done with a document expressing a decision tree or other, or with more sophisticated (integrated) tools;

- Registry actors will have a mechanism to update value sets and ensure that newly
555   published documents conform to the latest deployed Community Strategy;

- The publication transaction should be strictly validated and favor rejecting invalid publication requests (for correction and re-submission) rather than allowing poor quality metadata.

4.1 Managing metadata changes

560 Healthcare is changing rapidly, with a continuous introduction of new technologies, new use cases, new workflows, new data and new documents. This will lead to change proposals and to the need to update your current metadata set and Community Strategy. Each new version will require implementation management.

During the transition from an old Community Strategy to the new Community Strategy the
565 Community participants need to be managed and kept informed of the progress. Let's look at how you manage this transition:

- Define transition dates and governance for when newly created documents will be expected to use the new Community Strategy (e.g., value sets).

- Decide whether you will manage transitions per-participant or across the community. It is
570 easier for participants clients to deal with all participants transitioning to a new metadata usage on the same date, but this is often not possible. Be prepared to manage separate dates per participant…

- Decide on the length of the transition period, in which either the new or the old metadata values may be used. Transition periods can make it easier to manage transitions at the
575 Community level. The length of transition period depends on the ambition, coordination, resources, number of participants etc.

- Decide how you will handle existing documents (i.e., created prior to the change in allowable metadata values). If you will be changing anything, define those transition periods as well.

580 Let's also look at how you might handle historic existing documents and their metadata. There are some strategies to choose from:

- Modify metadata for the existing documents. This is discouraged as the process of changing all historic entries is very expensive as the longitudinal data grows. If you do update the metadata on historic entries, track the dates these are modified for each
585 participant. You may then be able to deprecate values or value sets entirely, removing them from the Community Strategy. Only do this if these values are truly not in use and will not show up in queries.

- Develop "smart" implementations that can map between new and old values. Note that this is not part of the XDS specification but is a rather a potential value added by the
590 implementer. This automatic mapping may be done at either the client or server side. For example,

  o A GUI that implements an XCA Initiating Gateway could present only new values to the user, and convert a request for new value X, to a query for X OR old value Y.

  o A server that implements an XDS Registry could respond to a query for X with
595 matches on X OR Y.

- Communicate with all stakeholders, so they can adjust their querying and processing accordingly. If necessary, inform end users of the changes and additions to the metadata set and the UI.

600
- Document Consumers will need to be robust with their ability to process historical documents[3].

Note that these strategies rely on the Community Strategy containing accurate and current information on the state of metadata usage.

4.2 Maintenance and Governance

Once the Community Strategy is deployed there are some considerations to maintain control.

605
- Set up a governance for making changes that engages the Community and assures everyone is empowered to consider the impact, provide feedback, and is informed of the change;

- Monitor for partners that are not following the Strategy. This might be an indication that:

  o the current Strategy is not sufficient for the use cases;

610
  o the participant is unaware of the Community Strategy;

  o end users have not been fully informed of the Community Strategy;

  o the participant is tardy with implanting the Community Strategy.

- Ensure that the current version of the Community Strategy reflects the current state of metadata usage and any transitions. As explained earlier, historic value sets and

615
configurations should be maintained to aid with transition.

---

[3] Postel's law on robustness: "Be liberal in what you accept, and conservative in what you send"

5   Metadata principles and guidelines

5.1 Good Metadata Principles

The goal of Good Metadata is to enable effective discovery of the 'right' document. This goal is focused on the Document Consumer success, as enabled by well controlled Document Source filling out the metadata when the content is Published or otherwise made available for discovery.

Constraints on the use of Document Metadata are not limited to coded elements. Also critical to good discovery is the use of date/time stamps, structured elements, identifiers, and narrative elements. The handbook will guide the team to evaluate the critical few metadata elements and their search and refinement. Further refinement beyond these critical few is encouraged but not the focus of this handbook.

This section outlines some principles of good metadata use. Following these principles is not required but is highly recommended. Past experience has learned these lessons.

5.1.1   Principles for Codes

Use of coded and structured metadata values will result in better discovery and better preservation of the meaning of those metadata values. This principle discourages use of free text, or minimally filed out structured values (e.g., author element) as a means to perform an initial query. Use of textual content in metadata values is effective in a second step filtering (e.g., when a human is involved in browsing and selecting relevant document entries based on reading free text). The following metadata attributes use codes:

Table 5.1.1-1: Excerpt of Metadata Definitions from ITI Technical Framework Volume 3

Attribute	Description		Source	Section
classCode	The code specifying the high-level use classification of the document type (e.g., Report, Summary, Images, Treatment Plan, Patient Preferences, Workflow).	Code	ebRIM Classification	4.2.3.2.3
Confidentiality-Code	The code specifying the level of confidentiality of the document.	Code	ebRIM Classification	4.2.3.2.5
eventCodeList	This list of codes represents the main clinical acts, such as a colonoscopy or an appendectomy, being documented.	Code	ebRIM Classification	4.2.3.2.8
folderCodeList	A list of codes used to semantically identify a set of documents for a specific purpose.	Code	ebRIM Classification	
formatCode	The code specifying the detailed technical format of the document.	Code	ebRIM Classification	4.2.3.2.9
healthcareFacility TypeCode	This code represents the type of organizational setting of the clinical encounter during which the documented act occurred.	Code	ebRIM Classification	4.2.3.2.11

Attribute	Description		Source	Section
languageCode	Specifies the human language of character data in the document.	String	ebRIM Slot	4.2.3.2.13
practice-SettingCode	The code specifying the clinical specialty where the act that resulted in the document was performed (e.g., Family Practice, Laboratory, Radiology).	Code	ebRIM Classification	4.2.3.2.17
typeCode	The code specifying the precise type of document from the user perspective (e.g., LOINC code).	Code	ebRIM Classification	4.2.3.2.25

Here are some general principles for good metadata:

- Some metadata are simply value sets representing facts (mime-type, FormatCode, languageCode), where there are other value sets that need to be designed using this handbook (classCode, practiceSettingCode) to assure discoverability. Some metadata attributes cluster into related pairs of metadata that work together, such as:

  o classCode – typeCode;

  o mimeType – formatCode;

  o formatCode – eventCodeList;

  o practiceSettingCode – authorRole.

- A value set should be made up of similar kinds of objects. For example, one should not define values in a typeCode value set made up of Document codes and Procedure codes, as documents are larger kind of an object than a procedure. Also, do not use perspectives that are already defined in other metadata attributes, such as 'Dermatology discharge summary' (specialty and document type).

- Values in a value set should be non-overlapping as much as possible. Overlapping values lead to ambiguity in the selection process. Possible exceptions to this principle should be clearly substantiated and should be well understood by the end users;

- Where codes are selected for a metadata attribute, they need to be very clearly defined and should express the cases where the code should be used. Clear and specific definitions of vocabulary are critical to success. Look at (nationals extensions of) international terminologies first (LOINC, SNOMED), and then at national code systems, depending on the national requirements and regulations;

- The displayName of that code is not critical in the metadata, as the text displayed to a user to represent the definition and meaning of that code can be customized to the application, user, workflow, and background. Often these displayName alternatives can be provided through a vocabulary profile like Sharing of Value Sets (SVS);

640

645

650

655

660

- Definitions of codes in the metadata system must not conflict with the definition given for that code by the standard that defines that code;

- Select codes from internationally defined code systems (e.g., SNOMED, LOINC), then nationally managed value sets, only when no code can be found should a locally defined code be used;

- A value set may contain codes from different code systems.

- Within and across Community boundaries the use of shared and consensus vocabulary will lead to better search results. Exceptions may be unavoidable but should be discouraged as much as possible. Exceptions should be interpreted as an indication of the need to re-examine the current Community consensus for Document Metadata vocabulary;

- When a Community publishes a value set, it should indicate the conformance expectation (example, preferred, extensible, required)

5.1.1.1 Good classCode and typeCode

- The typeCode should be specific to the type of document published. The value set for typeCode will then contain potentially overlapping codes to be most expressive of the exact type of document published. Different clinical perspectives at the document source may result in preferring to use of one of these different overlapping codes.

- The classCode should be optimized for query discoverability. The value set for classCode should be small and representing non-overlapping concepts. This will ensure that any document consumer query expression results in getting deterministic query results, across multiple document sources.

- For every value in the typeCode value set, there should be one agreed value from the classCode value set. In this way all typeCodes are classified in an agreed way for that Community.

5.1.1.2 Good practiceSettingCode

- The value set for practiceSettingCode should be optimized for discovery, and thus define the clinical specialties that would be interested in the document being published. Given that clinical specialties are often defined in a very fine-grained level to account for the evolutionary way to span the complexity of healthcare, it is recommended to limit the definition of the Practice-Setting value set to be only a coarse-grained value set, containing non-overlapping broad specialties (e.g., include Neurology and Pediatrics, but exclude Neuro-Pediatrics and considered part of Pediatrics).

5.1.2 Use of Specific Attributes for a good discovery

- When discovering documents of interest, one should search on a critical few search criteria, and further refine locally leveraging the other metadata. In this way one avoids missing documents due to overly specific query requests. There are exceptions to this, such as when a very specific type of document is the target, these exceptions should be used sparingly and with care.

  The most relevant "critical few criteria" typically relate to:

  - Patient ID → Who ?

  - classCode, → What class of document ?

  - practiceSettingCode, → Which clinical specialty originated the document ?

  - healthcarefacilityTypeCode, →Where in the health system ?

  - StartTime-End/Time, → When treated ?

- These generic "critical few" can be extended with class/specialty specific "critical few". The presence of these class/specialty specific "critical few" may form a second layer of search criteria. For their use to be further refining the above generic critical few, their presence need to be conditional to the presence of one or more of these generic critical few (e.g., imaging modality metadata use in queries is conditional to classCode being related to images and PracticeSettingCode to be cardiology or radiology).

- As you look for older data one should be more liberal in what you look for, as the older the data the less likely it adheres to current rules.

  - This might mean you manage two different value sets for any element. One that is the recommended use, another that contains all known current and historic value sets.

5.1.3 Coordination of publication and discovery

- The granularity of both publishing and discovery must be understood in the same way across all implementation of the metadata usage among the sharing systems. For example, with the author location element, bad search results will happen when the searching is expecting the metadata to be broad granularity (address of the facility within which many departments exist), yet the publishing is done with fine grain (specific location within a department)

- Patient Identity quality is critical to Document Sharing, use of IHE PIX, PDQ, and XCPD enable this, where data used for identity will be influenced by national, regional, and local policies

5.1.4 Miscellaneous

- When publishing a document where a specific metadata element value is unknown, there should be well defined rules on how unknown is handled:

- o Allowing an empty or null value (is it allowed?) or creating a specific code value "unknown" in metadata value sets weakens the quality of the query responses as it would generate either a systematic match or no match ever

- o Not allowing publishing when the most relevant "critical few criteria" elements (See Section 2.1.2) are not known may be an acceptable policy.

These cases should be kept to a minimum but are realistic. E.g., don't publish or send later, allow placeholder 'unknown' values and require to replace later when information is known.

- A Community may mandate that specific elements must contain a value. Thus, within that Community an element is mandatory where it might otherwise be considered optional.

- There should be a well-defined set of known error and warning codes as well as the capability of handling unexpected error and warning conditions. By defining expected codes, new error conditions will automatically trigger the need for and discussion of new error codes. (e.g., new error codes may be needed to accommodate conditions associated to new technology stacks)

5.2 Principles per metadata attribute

5.2.1 classCode

(See also Section 4.2.1 for the organization of classCode and typeCode).

- Critical for searching for a document of interest;

- Coarse level granularity, non-overlapping, preferably from one perspective;

- Value set managed at Community level should be closed, only values from the value set are used for publication;

- Nonoverlapping concepts to enable most effective searching success;

- Clear definition and guidance for proper creation and lookup of documents.

5.2.2 typeCode

- should not be used for searching unless looking for a very exact type of document. The classCode combined with practiceSetting should be used for searching;

- would be inclusive of all document types known to be published within the Community;

- value set managed at Community level should be extensible to enable exceptional use cases.

5.2.3 practiceSettingCode

- critical for search for a document of interest produced by a specific clinical specialty;

- should be a small well controlled value set to enable searching;

- Nonoverlapping concepts to enable most effective searching;

- Coarse level of granularity but explain how to deal with 'border' cases (for example, pediatric neurosurgery).

5.2.4 healthCareFacilityTypeCode

- Critical for search for a document of interest produced by a certain type of healthcare facility well understood by patients (hospital, doctor office, nursing home, imaging center, etc.);

- Coarse level of granularity;

- small, well controlled value set to enable searching;

- Non-overlapping concepts to enable most effective searching.

5.2.5 author

5.2.5.1 authorRole

- Critical for search for a document of interest produced by a certain type of healthcare facility well understood by patients (hospital, doctor office, nursing home, imaging center, etc.

- should be a small well controlled value set to enable searching.

5.2.5.2 authorSpeciality

- Useful in a secondary filtering (See Section 3.4), when looking for documents produced by a few very specific specialties;

- Critical for secondary filtering for a document of interest produced by a certain type of specialty or sub-specialty;

- Fine-grained level of granularity;

- should not be used for searching unless looking for a very exact specialty. The practiceSetting should be used for searching specialty for primary filtering;

- would be inclusive of all specialties and sub-specialties known to be published within the Community;

- value set managed at Community level should be extensible to enable exceptional use cases.

Glossary

Term	Definition
Community	A combination of legal, organizational, financial and operational rules regarding healthcare. A Community can be organized on multi-national, national, and/or regional level.
Community Strategy	The output of the use of the Metadata Handbook, made up of Policy, Business Needs, Use cases, Procedures, Value Sets, and other use constraints.
Document	Patient specific document of any type or format.
Participant	Usually an organization within an Affinity Domain or Community.

800