TMSAD 23-4.5.7

Paper for Consideration by TSMAD

Metadata – down the rabbit hole

Submitted by:	S-101 Work Item Leader (with help from IIC)
Executive Summary:	S-100 greatly expanded the use of metadata. S-101 needs to implement robust metadata for discovery purposes. This paper is to progress S-101 metadata towards completion and recommend extending S-100 to accommodate S-101 metadata needs
Related Documents:	S-101 ENC product Specification IIC S-101 Metadata analysis
Related Projects:	N/A

Introduction / Background

At TSMAD 20 and 21 the group began to review the essential metadata needed for S-101. While general agreement was reached on most of the items, it was pointed out that the current structure needed to be reviewed against S-100 and ISO 19115 to ensure that the proposed S-101 metadata fields were in compliance.

Analysis/Discussion

In August 2011, NOAA contracted with IIC to review the S-101 metadata. As a result, it found that some of the proposed metadata fields in S-101 did not have an S-100 equivalent or they are included in the S-100 Part 4A Annex D-2 – Elements of an Exchange Set. In attempting to harmonize S-101 metadata with S-100 it appears that many of the fields are harmonized that there were inconsistencies between the two. While S-101 can be extended beyond S-100, many of these metadata fields may be useful in other S-100 based products. The purpose of this paper is to review the various S-101 metadata fields needed for the following:

- Exchange Set Metadata
- Dataset Metadata
- Support File Metadata
- Exchange Catalogue Metadata

Once TSMAD has reviewed the S-101 metadata fields to determine if they are valid for inclusion in S-101 there will be a follow on paper to approve the harmonization between S-101 metadata and S-100 metadata in order to bring alignment between the two.

The format of the review will be for each S-101 section there will be series of questions for TSMAD to consider.

The following is the revised S-101 metadata section. The colour coding for the columns is as follows:

GREEN = Mandatory S-100 and S-101 element
BLUE = Elements are located in S-100 Part 4a Appendix D.2(normative) but the S-101 structure does not conform to the S-100 structure
RED = there is no S-100 or ISO 19115 equivalent
ORANGE = called from ISO 19115

Metadata

Introduction

This clause defines the mandatory and optional metadata needed for S-101. In some cases the metadata may also support national language. If this is the case it is noted in the Remarks column.

Exchange Set Metadata

Name	Multiplity	Value	Type	Remarks
S100_ExchangeSet	-		-	Aggregation of the elements comprising an exchange set for the transfer of data.
aggregateFile	0..*		-	Collection of support files in the exchange set
partOf	0..*		-	Collection of datasets which are part of the exchange set
aggregateCatalogue	0..*		-	Collection of catalogues
superset				The master container exchange set which can contain a subSet of exchange sets
subset				Exchange set which is part of the superSet

TSMAD Question: This section has incorporated the elements from Part 4A Appendix D.2 – elements of an exchange set. Does TSMAD agree these are the proper elements for the exchange set?

Dataset Metadata

Name	Multiplity	Value	Type	Remarks
S101_DataSetDiscoveryMetadata	-		-	-
metadataFileIdentifier	1		CharacterString	
metadataPointOfContact	1		CI_ResponsibleParty	
metadataDateStamp	1		Date	
metadataLanguage	1	English	CharacterString	All data sets conforming to S-101 PS must use English language
fileName	1		CharacterString	Dataset file name
filePath	1		CharacterString	Full path from the exchange set root directory
description	1		CharacterString	Short description of the area covered by dataset harbour or port name, between two named locations etc. NATIONAL LANGUAGE enabled
dataProtection	1	{1} to {2}	CharacterString	1. Encrypted 2. Unencrypted
classification	1	{1} to {5}	Class MD_SecurityConstraints>MD_ClassificationCode (codelist)	1. unclassified 2. restricted 3. confidential 4. secret 5. top secret

Comment [JLP1]: ADD to S-100

Name	Multiplity	Value	Type	Remarks
purpose	1	{1} to {5}	CharacterString MD_Identification>purpose (character string)	1. New Dataset 2. New Edition 3. Update 4. Re-issue 5.Cancellation
specificUsage	1	{1} to {3}	CharacterString MD_USAGE>specificUsage (character string) MD_USAGE>userContactInfo (CI_ResponsibleParty)	1. Port Entry – A dataset containing data required: • For navigating the approaches to ports • for navigating within ports, harbours, bays, rivers and canals, for anchorages • as an aid to berthing or any combination of the above. 2.Transit – A dataset containing data required for : • navigating along the coastline either inshore or offshore • navigating oceans, approaching coasts • route planning or any combination of the above. 3.Overview – A dataset containing data required: • for Ocean Crossing • route planning
editionNumber	1		Integer	When a data set is initially created, the edition number 1 is assigned to it. The edition number is increased by 1 at each new edition. Edition number remains the same for re-issue.
updateNumber	1		CharacterString	Update number 0 is assigned to a new data set, and increased by one for each subsequent update.
updateApplicationDate	0..1		Date	this date is only used for the base cell files (i.e. new data sets, re-issue and new edition), not update cell files. All updates dated on or before this date must have been applied by the producer
issueDate	1		Date	Date on which the data was made available by the data producer.

Comment [JLP2]: Does this need a CODE LIST in S-100????

Name	Multiplity	Value	Type	Remarks
productSpecification	1	S-101 version 0.0.1	S-100_ProductSpecification	This must be encoded as S-101
name	1		CharacterString	The name of the product specification used to create the datasets
version	1		CharacterString	The version number of the product specification
date	1		Date	The version date of the product specification
producingAgency	1		CI_ResponsibleParty	Agency responsible for producing the data.
optimumDisplayScale	1	{1} to {13}	S101_DisplayScale	1: <4,000 2: 4,000 3: 8,000 4: 12,000 5: 22,000 6: 45,000 7: 90,000 8: 180,000 9: 350,000 10: 700,000 11: 1,500,000 12: 3,000,000 13: >3,000,000
maximumDisplayScale	1	{1} to {13}	S101_DisplayScale	1: <4,000 2: 4,000 3: 8,000 4: 12,000 5: 22,000 6: 45,000 7: 90,000 8: 180,000 9: 350,000 10: 700,000 11: 1,500,000 12: 3,000,000 13: >3,000,000
minimumDisplayScale	1	{1} to {13}	S101_DisplayScale	1: <4,000 2: 4,000 3: 8,000 4: 12,000 5: 22,000 6: 45,000 7: 90,000 8: 180,000 9: 350,000 10: 700,000 11: 1,500,000 12: 3,000,000

Comment [JLP3]: ADD to S-100

Comment [JLP4]: ADD to S100

Comment [JLP5]: ADD to S100

Name	Multiplity	Value	Type	Remarks
				13: >3,000,000
horizontalDatum	1	WGS84	CharacterString	EPSG:4326
verticalDatum	1	{1} to {30}	Double S100_SoundingAndVerticalDatum	1 : Mean low water springs 2 : Mean lower low water springs 3 : Mean sea level 4 : Lowest low water 5 : Mean low water 6 : Lowest low water springs 7 : Approximate mean low water springs 8 : Indian spring low water 9 : Low water springs 10 : Approximate lowest astronomical tide 11 : Nearly lowest low water 12 : Mean lower low water 13 : Low water 14 : Approximate mean low water 15 : Approximate mean lower low water 16 : Mean high water 17 : Mean high water springs 18 : High water 19 : Approximate mean sea level 20 : High water springs 21 : Mean higher high water 22 : Equinoctial spring low water 23 : Lowest astronomical tide 24 : Local datum 25 : International Great Lakes Datum 1985 26 : Mean water level 27 : Lower low water large tide 28 : Higher high water large tide 29 : Nearly highest high water 30 : Highest astronomical tide (HAT)
soundingDatum	1	{1} to {30}	Double S100_SoundingAndVerticalDatum	1 : Mean low water springs 2 : Mean lower low water springs 3 : Mean sea level 4 : Lowest low water 5 : Mean low water 6 : Lowest low water springs 7 : Approximate mean low water springs 8 : Indian spring low water 9 : Low water springs 10 : Approximate lowest astronomical tide

Comment [JLP6]: ADD Codelist to S-100

Comment [JLP7]: ADD codelist to S-100

Name	Multiplity	Value	Type	Remarks
				11 : Nearly lowest low water 12 : Mean lower low water 13 : Low water 14 : Approximate mean low water 15 : Approximate mean lower low water 16 : Mean high water 17 : Mean high water springs 18 : High water 19 : Approximate mean sea level 20 : High water springs 21 : Mean higher high water 22 : Equinoctial spring low water 23 : Lowest astronomical tide 24 : Local datum 25 : International Great Lakes Datum 1985 26 : Mean water level 27 : Lower low water large tide 28 : Higher high water large tide 29 : Nearly highest high water 30 : Highest astronomical tide (HAT)
dataType	1	ISO 8211 BINARY	S-100_DataFormat	
otherDataTypeDescription	0..1		CharacterString	
boundingBox	1		EX_GeographicBoundingBox	The extent of the cell limits
boundingPolygon	1..*		EX_BoundingPolygon	A polygon which defines the actual data limit
comment	0..1		CharacterString	Any additional Information NATIONAL LANGUAGE enabled
cyclicRedundancyCheck	1		CharacterString NonNegativeInteger	
layerId	1..*	{1} to {3}	CharacterString	Identifies the relationship to other layers that are required to view the complete data set. 1. Scale Independent 2. Scale Dependent 3. Complete

TSMAD Questions:

1. **Does TSMAD have any additions or deletions to the dataset discovery metadata fields?**
2. **Does TSMAD agree that the following fields should be added to S-100: dataProtection, optimumDisplayScale, minimumDisplayScale, maximumDisplayScale and layered?**

3. **Does TSMAD agree that the values for Vertical and Sounding datum should be made into an S-100 codelist that S-101 can call?**
4. **Does the metadata field purpose need a code list in S-100?**
5. **Does TSMAD agree that the values for OPTDSC, MINDSC, and MAXDSC should be captured as an S-101 code list? See example below.**

S101_DisplayScale

Role Name	Name	Description	Mult	Type	Remarks
Class	S-101_DisplayScale	The allowed display scales for S-101 ENCs	-	-	-
Value	<4,000		-	-	-
Value	4,000		-	-	-
Value	8,000		-	-	-
Value	12,000		-	-	-
Value	22,000		-	-	-
Value	45,000		-	-	-
Value	90,000		-	-	-
Value	180,000		-	-	-
Value	350,000		-	-	-
Value	700,000		-	-	-
Value	1,500,000		-	-	-
Value	3,000,000		-	-	-
Value	>3,000,000		-	-	-

Support File Metadata

Name	Multiplicity	Value	Type	Remarks
S101_SupportFileDiscoveryMetadata	-		-	-
aggregateFile	0..*		-	Collection of Support Files

Name	Multiplicity	Value	Type	Remarks
supportFile	0..*		-	File which has information about a dataset
fileName	1		CharacterString	
filePath	1		CharacterString	
Purpose	1	{1} to {3}	class S-100_SupportFilePurpose	1. New – A file which is new 2. Replacement – A file which replaces an existing file 3. Deletion – deletes an existing file
editionNumber	1		CharacterString	When a data set is initially created, the edition number 1 is assigned to it. The edition number is increased by 1 at each new edition. Edition number remains the same for a re-issue.
issueDate	1		Date	Date on which the data was made available by the data producer.
productSpecification	1		S-100_ProductSpecification	Version of S-101
dataType	1	{1} to {4}	class S-100_SupportFileFormat	1. TXT =Text files 2. XML = Text files 3. HTM = Text files 4. TIFF = Picture files
dataTypeVersion	1		CharacterString	The version number of the dataType
Comment	0..1		CharacterString	Any additional Information NATIONAL LANGUAGE enabled
CyclicRedundancyCheck	1		CharacterString	

TSMAD Questions: Is TSMAD satisfied with the fields needed for support file metadata?
Is there a need for a field such as fileType to identify if the file contains a T and P notice, Chart Note or Other?
Is there a need to add a support file expiration date?

Exchange Catalogue File Metadata

The catalogue file is defined in XML schema language and the data set files are encoded as ISO/IEC 8211 data records, fields, and subfields. The Exchange catalogue inherits the dataset discovery metadata and support file discovery metadata.

Name	Multiplicity	Value	Type	Remarks
S101_ExchangeCatalogue	-			An exchange catalogue contains the discovery metadata about the exchange datasets and support files
identifier	1		CharacterString S-100_CatalogueIdentifier	Uniquely identifies this exchange catalogue
editionNumber	1		CharacterString	The edition number of this exchange catalogue
date	1		Date	Creation date of the exchange catalogue
contact	1		S-100_CataloguePointofContact CI_ResponsibleParty	
organization	1		CharacterString	This could be an individual producer, value added reseller, etc. The organization distributing this exchange catalogue
phone	0..1		CI_Telephone	The edition number of this exchange catalogue
address	0..1		CI_Address	The Address of the organization
productSpecification	0..1			Details about the product specifications used for the datasets contained in the exchange catalogue. This value is conditional on all datasets using the same product specification
name	1		CharacterString	The name of the product specification used to create the datasets
version	1		CharacterString	The version number of the product specification
date	1		Date	The version date of the product specification
MetadataLanguage	1	English	CharacterString	All data sets conforming to S-101 PS must use English language
exchangeCatalogueName	1	CATALOG.101	CharacterString	Catalogue filename
exchangeCatalougeDescription	1		CharacterString	Description of what the exchange catalogue contains NATIONAL LANGUAGE enabled
ExchangeCatalogueComment	0..1		CharacterString	Any additional Information

Comment [JLP8]: Not sure if it belongs at this level – but rather at the dataset level as the exchange set can contain multiple datasets based on different product specifications.

Comment [JLP9]: Should this be at the dataset level?

Name	Multiplicity	Value	Type	Remarks
				NATIONAL LANGUAGE enabled
compressionFlag	1	{1} to {2}	CharacterString	1. Yes 2. No
algorithmMethod	1	{1} to {2}	CharacterString	1. ZIP 2. RAR
sourceMedia	1			
replacedData	1			If a data file is cancelled is it replaced by another data file
dataReplacement	0..1			Cell name

Comment [JLP10]: Everything below needs review by DPSWG.

ADD to S100

Questions for TSMAD:

1. Should MetadataLanguage be at the dataset level?
2. Does TSMAD agree to the addition of the inherited fields for identifier, contact and product specification?
3. Does TMAD know what identifier is for? There is already a field for the name of the catalogue.

Recommendations

Based on the TSMAD discussion these revised and harmonized sections will be added into the S-101 metadata section. The extensions to S-100 will be discussed under a separate paper.

Action Required of [HSSC] [Relevant HSSC WG]

The TSMAD is invited to:

a. discuss the questions proposed in this paper
b. recommend inclusion of the resultant discussion into S-101 metadata
c. note that this discussion may affect the S-100 metadata section.