

Paper for Consideration by TSMAD

S-101 Metadata Comments – or how TSMAD learned to love metadata

Submitted by:	S-101 Work Item Leader
Executive Summary:	This paper specifically addresses the metadata comments from the last comment round.
Related Documents:	Any relevant documents and references to the extent that they are known to the originator.
Related Projects:	Any related projects that may impact upon considerations

Introduction / Background

After TSMAD24, S-101 was distributed for another round of comments. As a result there were additional comments regarding clause 12 – Metadata. As these comments require substantive discussion, the S-101 work item leader felt that they would be better dealt with via a separate paper rather than through the S-101 comments adjudication session.

For clarity, the comment is in **red text** and the proposed resolution is in **blue text**. Editorial notes and proposed edits from the editor are in **green text**

12 Metadata

12.1 Introduction

COMMENT (2J): There is apparently no mention of metadata about the exchange set itself, rather than the individual elements itself such as catalogue, datasets, and support files.

Clause 12.1.1 is titled Exchange Set Metadata but the table has class S100_ExchangeSet which lacks metadata elements like date, identification, etc. The language in 12.1.4 is ambiguous, it refers to the catalogue (e.g., "date" is "the creation date of the exchange catalogue").

PROPOSAL:

(1) Define a new element for exchange set metadata (S100_ExchangeSetMetadata), or,

(2) add exchange set metadata elements to an existing metadata element, or,

(3) clarify the language in section 12.1.1 and 12.1.4.

The exchange set metadata should have the date of issue, date of application, information about the producer (compiler of the exchange set, who might be different from the producer of the datasets and support files)

ED NOTE1: Yes, refinements are definitely possible. We need to keep in mind that S-101 metadata is based on S-100 metadata though. So it would be sensible to consider all such proposals in that context. i.e. if we determine that it would be sensible to include more metadata about the exchange set then we should consider whether it should be a S-100 wide change (something desirable to have in all S-100 products) or a S-101 change (metadata items specific to S-101 only). In the first case this means revising S-100 in which case the change should 'automatically' to S-101. In the second case these mean custom mods, which should be reserved for S-101 specific elements. I realize updating S-101 to include such elements could be viewed as a much simpler thing to do (at this point) but it is really against the spirit of interoperability etc.

Now, it is worth noting that the S100_exchangeSet class in an aggregation and the related metadata about the exchange set is currently intended to be located in the S100_ExchangeCatalogue. For example Cl_ResponsibleParty type there, is intended to be used to encode the compiler of the exchange set. Similarly S100_CatalogueIdentifier type includes Date intended for the exchange set creation date. There is an inherent assumption here that the exchange set catalogue creation is the final step in the exchange set creation process.

With that in mind it would appear that the key proposed elements are already there. If some additional ones are desirable they can naturally be added keeping in mind S-100 vs S-101.

ED NOTE2: In reviewing the S-100 section on metadata, I noticed that we should incorporate some of the explanatory text and graphics. It may make things more clear, please note that the graphics will be updated to reflect S-101 and not S-100:

Proposed Addition:

For information exchange, there are several categories of metadata required: metadata about the overall exchange catalogue, metadata about each of the datasets contained in the catalogue, and metadata about the support files that make up the package.

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

Figures 1 to 3 outline the overall concept of an S-101 exchange set for the interchange of geospatial data and its relevant metadata. Figure 1 depicts the realization of the ISO 19139 classes which form the foundation of the exchange set. The overall structure of S-100 Exchanges Sets is modelled in Figure 2. More detailed information about the various classes is shown in Figure 3 and a textual description in the tables at clause 3.

The discovery metadata classes have numerous attributes which enable important information about the datasets and accompanying support files to be examined without the need to process the data, e.g. decrypt, decompress, load etc. Other catalogues can be included in the exchange set in support of the datasets such as feature, portrayal, coordinate reference systems, code lists etc. The attribute “purpose” of the support file metadata provides a mechanism to update support files more easily.

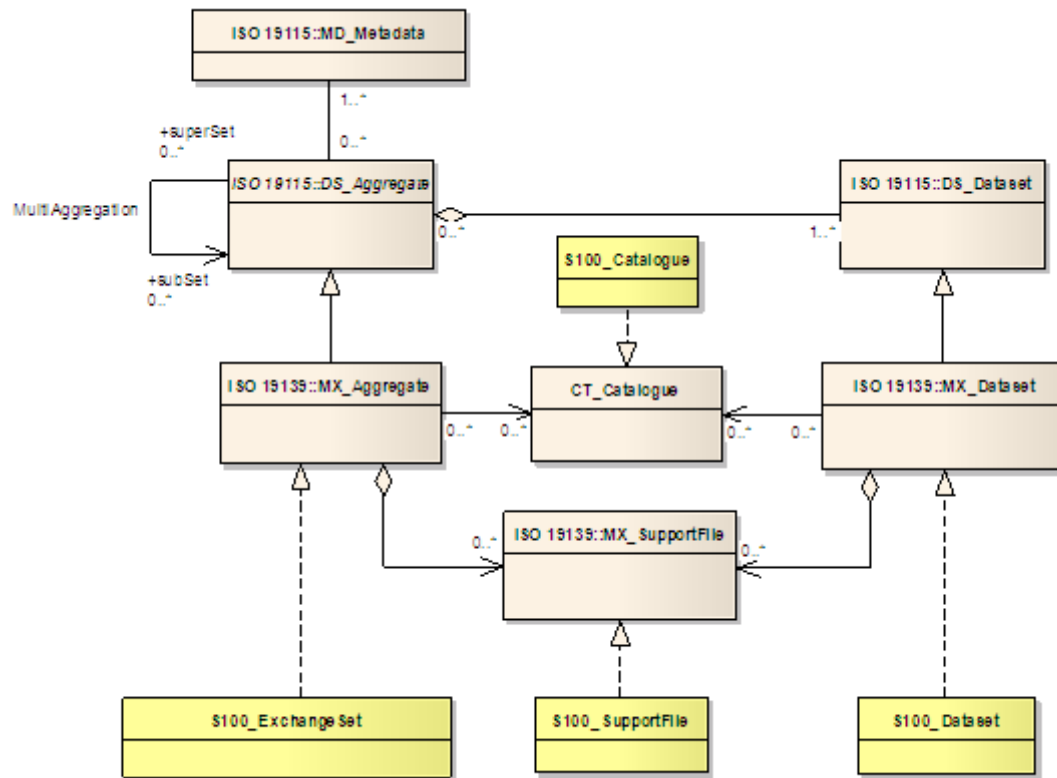


Figure 1 Realization of the Exchange Set Classes

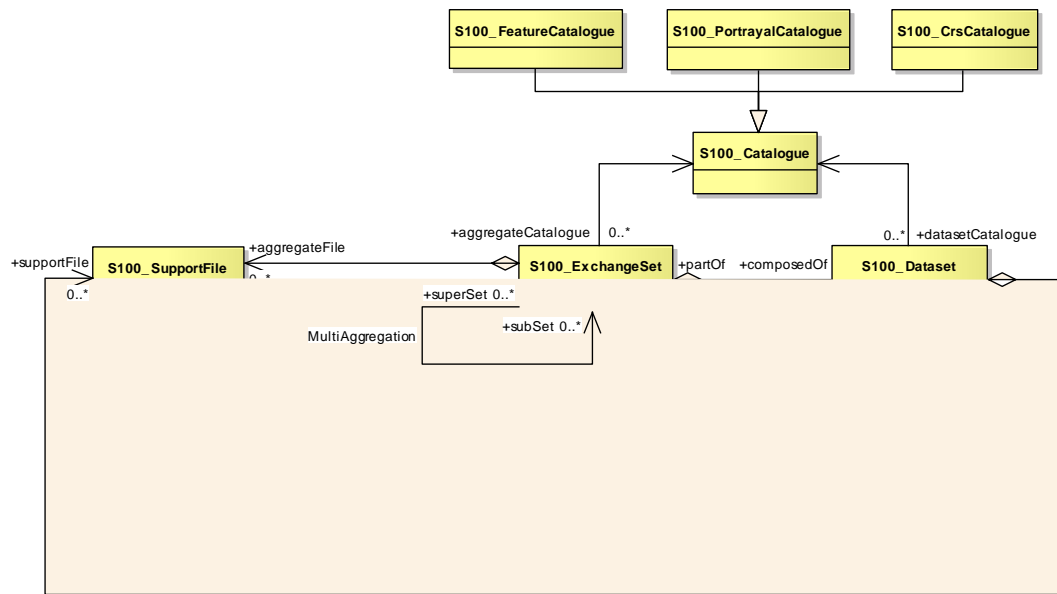


Figure 2 - S100 ExchangeSet

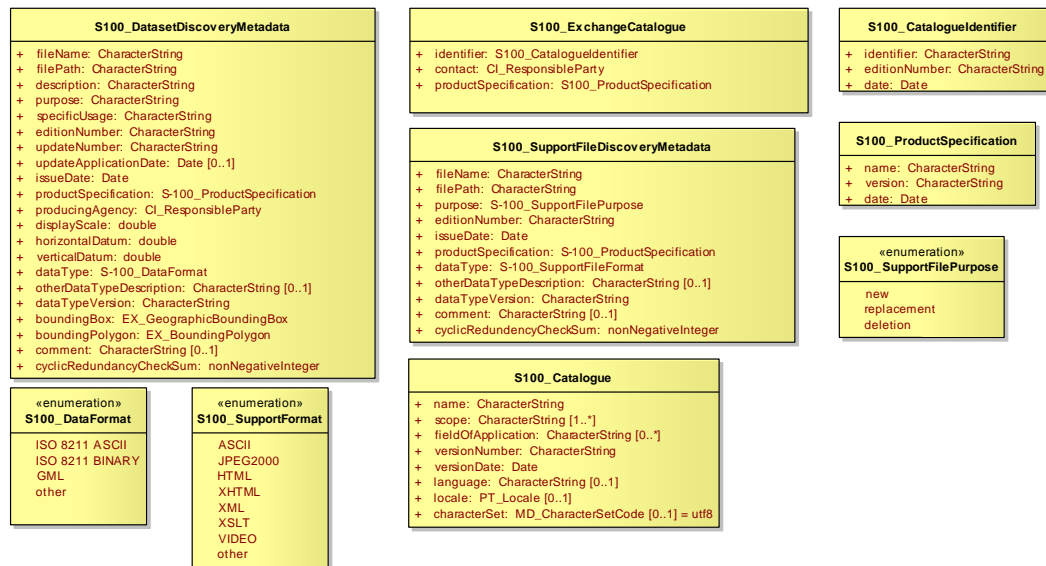


Figure 3 S100 Exchange Set - class details

12.1.1 Exchange Set Structure

COMMENT (UKHO): UKHO: These items define the relationships not actual metadata values this should be made clear or the section removed. It could be argued Figure 13 conveys this more clearly.

PROPOSAL:

Clarify or delete clause; lack of type and values makes it difficult to understand how it is to be implemented

COMMENT (2J): Not clear what this clause “Exchange set Metadata” does or how it is supposed to be implemented. Its name suggests metadata about the exchange set, but in its present form it appears to be more of a catalogue giving the structure of an exchange set rather than metadata. Taking the current definition literally, implementation would be as a collection of pointers to datasets, support files, and exchange sets of which this exchange set is a part, or which are part of it.

PROPOSAL

Clarify or delete clause; lack of type and values makes it difficult to understand how it is to be implemented

EDITORS NOTE: I'm not sure what the value is of this clause. I do know that it is part of S-100 metadata and we had this discussion back at TSMAD23 regarding harmonization and we agreed to leave it in. However what is missing is the introductory language:

An S-100 Exchange Set is an aggregation of all the various elements required to support the interchange of geospatial data and metadata. The MultiAggregation association introduces the concept of using subsets which could be domain oriented e.g. packaged by scale, producer, region etc.

The S100_ExchangeSet class is a conceptual entity intended to illustrate the relationship between the various components of the exchange set. It is not intended to be implemented or contain metadata about the exchange set, which will be encoded in S100_ExchangeCatalogue.

Name	Multiplicity	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

12.1.2 Dataset Metadata

COMMENT (SHOM): The only way to encode a Copyright is using the Comment field. Create a new metadata to encode a copyright. The number of copyright references for one ENC may be multiple.

PROPOSAL: Create a metadata field for copyright. Inherits from MD_LegalConstraints ->MD_RestrictionCode <copyright>, multiplicity 0..*

COMMENT (2J): : Meaning of “full path” should be clarified. Should it start with a “/”? Does it include the file name too? If it includes the file name too, that is redundant because it is already given in the fileName metadata attribute

PROPOSAL: Change remarks to: Path to the dataset file, relative to the root directory of the exchange set. The location of the dataset file after the exchange set is unpacked into directory <EXCH_ROOT> will be: <EXCH_ROOT>/<filePath>/<fileName>

COMMENT (US – SPAWAR): CRC should be expressed in hex notation

PROPOSAL: Add to remarks should be expressed in hex notation

PROPOSAL (NO): Should the layer complete still be listed? In clause 4.5.4 and 4.5.5 only the SI and SD layers are mentioned. If it is to be kept it should it also be described in the document? Also clause C4.5.3 only mentions the SI and SD layers.

PROPOSAL: Amend listing based on comments to 4.5.4 and 4.5.4.

Name	Multiplicity	Value	Type	Remarks
S101_DataSetDiscoveryMetadata	-		-	-
metadataFileIdentifier	1		CharacterString	The file name must be unique. Each file name must have a MD suffix added to the S-101 file name. Dataset: GB45678.000 Metadata: MD_GB45678_000.xml Update 1: GB45678.001 Metadata: MD_GB45678_001.xml

Name	Multiplicity	Value	Type	Remarks
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		Boolean	e.g. Encrypted or Unencrypted
protectionScheme	0..1		CharacterString	e.g. S-63
classification	1	{1} to {5}	Class MD_SecurityConstraints>MD_ClassificationCode (codelist)	1. unclassified 2. restricted 3. confidential 4. secret 5. top secret
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	1. Port Entry – A dataset containing data required: <ul style="list-style-type: none"> For navigating the approaches to ports

Name	Multiplicity	Value	Type	Remarks
			MD_USAGE>specificUsage (character string) MD_USAGE>userContactInfo (CI_ResponsibleParty)	<ul style="list-style-type: none"> • for navigating within ports, harbours, bays, rivers and canals, for anchorages • as an aid to berthing <p>or any combination of the above.</p> <p>2.Transit – A dataset containing data required for :</p> <ul style="list-style-type: none"> • navigating along the coastline either inshore or offshore • navigating oceans, approaching coasts • route planning <p>or any combination of the above.</p> <p>3.Overview – A dataset containing data required:</p> <ul style="list-style-type: none"> • 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.
updateApplicationDate	0..1		Date	this date is only used for the base dataset files (i.e. new data sets, re-issue and newedition), not update dataset 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

Name	Multiplicity	Value	Type	Remarks
				data producer.
productSpecification	1	S-101 version 0.0.1	S100_ ProductSpecification	This must be encoded as S-101
producingAgency	1		CI_ResponsibleParty	Agency responsible for producing the data.
maximumDisplayScale	1		Integer	Example: A maximum display scale of 1:22,000 would be encoded as 22000
horizontalDatumReference	1	EPSG	CharacterString	
horizontalDatumValue	1	4326	Integer	WGS84
verticalDatum	1	{1} to {30}	S100_VerticalAndSoundingDatum	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}	S100_VerticalAndSoundingDatum	1 : Mean low water springs 2 : Mean lower low water springs 3 : Mean sea level 4 : Lowest low water

Name	Multiplicity	Value	Type	Remarks
				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)
dataType	1	ISO 8211 BINARY	S100_DataFormat	
otherDataTypeDescription	0..1		CharacterString	
dataCoverage	0..*		S101_DataCoverage	Provides information about data coverages within the dataset
checksum	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

12.1.3 S101_DataCoverage

COMMENT (2J): Important enough to be defined in the body of the document should receive more explanation in main body of document

PROPOSAL: Does TSMAD agree to add more content to the main document.

IF SO: What should be added, and who will write it.

COMMENT (UKHO): UKHO: Is this consistent with S-100/ ISO 19115? Which uses geographic extent?

PROPOSAL: Amend to be consistent with S-100

ED NOTE: Both EX_GeographicBoundingBox and EX_BoundingPolygon are from ISO 19115 and are consistent with in S-100 4A-A. Now, the way the metadata is currently handled in S-100 consists of two sections: 4A-A (dataset level) S100_Metadata class and 4A-D (exchange set level) S100_DataSetDiscoveryMetadata. While the first one is consistent with ISO 19115, the second one is a custom implementation only partially conforming to ISO 19115. S-100 could benefit from a clear indication how they should be implemented. Presumably the full S100_Metadata is get implemented as XML and exists along with the dataset itself while the S100_DataSetDiscoveryMetadata is created as a consolidated listing at the exchange set level, one XML file for the entire series. I believe a more detailed description of the above would be desirable.

Name	Multiplicity	Value	Type	Remarks
S101_DataCoverage	-	-	-	-
ID	1		Integer	Uniquely identifies the coverage
boundingBox	1		EX_GeographicBoundingBox	
boundingPolygon	1..*		EX_BoundingPolygon	
maximumDisplayScale	1		Integer	

12.1.4 Support File Metadata

COMMENT (2J): 2J: dataType values are different from S-100 4a.D.2-11 and TSMAD23-4.3.6. The labels in S-100 are slightly different e.g., HTML instead of HTM. TSMAD23-4.3.6 conforms to S-100 (and TIFF is missing from those lists). Other product specifications will take S-100 as the norm for labels. It would be nice to avoid confusion like one product specification using Text and another using TXT (this row specified the metadata and not the file extensions). Make consistent with S-100.

PROPOSAL: edit either S-101 or S-100 or both to use the same labels for the same type).

ED NOTE: S100 uses the following:

Value	Text	
Value	JPEG2000	

Value	HTML	
Value	XML	
Value	XSLT	
Value	VIDEO	
Value	Other	

Name	Multiplicity	Value	Type	Remarks
S101_SupportFileDiscoveryMetadata	-		-	-
fileName	1		CharacterString	
filePath	1		CharacterString	Full location from the exchange set root directory
purpose	1	{1} to {3}	class S100_SupportFilePurpose	<ol style="list-style-type: none"> 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		S100_ProductSpecification	Version of S-101
dataType	1	{1} to {4}	class S100_SupportFileFormat	<ol style="list-style-type: none"> 1. TXT =Text files 2. XML = Text files 3. HTM = Text files

Name	Multiplicity	Value	Type	Remarks
				4. TIFF = Picture files
dataTypeVersion	1		CharacterString	The version number of the dataType
Comment	0..1		CharacterString	Any additional Information NATIONAL LANGUAGE enabled
checksum	1		CharacterString	
digitalSignatureReference	0..1		CharacterString	Reference to the appropriate digital signature algorithm
digitalSignatureValue	0..1		CharacterString	

Comment [N1]: 2J: What is the purpose of digitalSignatureReference and digitalSignatureValue?

Ed NOTE: I think this comes from S metadata

TSMAD24: UK to seek clarification from the DPSWG

STILL WAITING FROM THE UKHO

Comment [EK2]: Yes, I believe this was first proposed to be added in order to handle encryption requirements. I believe the concept was discussed a number of times thereafter and was considered to be a sensible idea and at one point a hash was proposed to be used.

12.1.5 Exchange Catalogue File Metadata

The catalogue file is defined in XML schema language. The Exchange catalogue inherits the dataset discovery metadata and support file discovery metadata.

COMMENT (2J): Unclear language “The Exchange catalogue inherits the dataset discovery metadata and support file discovery metadata.” taken literally this is multiple inheritance. Should it be “The Exchange catalogue metadata is inherited by the dataset discovery metadata and support file metadata?”

Also, given the current attribute multiplicities, it is hard to see what inheritance achieves. For example, if the productSpecification is inherited it need not be mandatory (i.e., multiplicity=1) in the child element

PROPOSAL: Clarify language, is it inheritance, aggregation, or composition that is really desired?

If “inheritance”, review multiplicities of attributes supposed to be inherited and make them conditional in the child. (Conditional on them not being present in the parent?)

ED NOTE: I agree, we are dealing with an aggregation between the catalog and metadata classes and should describe it accordingly. Similar to one of my comments above I believe the implementation aspect should also be provided. The UML diagram shows the relationships but it does not indicate how it all should be implemented i.e. should it be one XML for for a catalogue plus one XML for each dataset, plus one XML for each support file. Or should the individual dataset metadata files be aggregated into one dataset series (which is often done in ISO 19115 world) or should it be all consolidated into one bit XML catalogue consisting of the above subsections. ...

COMMENT (2J): editionNumber and Contact: why 2 types?

PROPOSAL: clarify type specification – also for contact.

ED NOTE: Both seem correct. Perhaps strangely, edition is defined as character string in 19115 and therefore in S-100 and therefore in S-101. Similarly contact is defined as a complex type CI_ResponsibleParty in order to cover its possible multiple characteristics (name, address, tel etc.). The S100_CataloguePointofContact should probably be removed from type and maybe a short note under remarks should be added that this POC is for catalogue/exchange set.

COMMENT (2J): attribute "identifier" already has edition number and date sub-fields from type S100_CatalogueIdentifier

PROPOSAL: delete these attributes; use attributes of "identifier" instead

COMMENT (SHOM?): Algorithm method. This is mandatory, but what if the exchange set is not compressed?

PROPOSAL multiplicity = 0..1, make conditional on compressionField=Yes

Name	Multiplicity	Value	Type	Remarks
S101_ExchangeCatalogue	-			An exchange catalogue contains the discovery metadata about the exchange datasets and support files
identifier	1		CharacterString S100_CatalogueIdentifier	Uniquely identifies this exchange catalogue
editionNumber	1		CharacterString	The edition number of this exchange catalogue
contact	1		S100_CataloguePointofContact CI_ResponsibleParty	
date	1		Date	Creation date of the exchange catalogue
MetadataLanguage	1	English	CharacterString	All data sets conforming to S-101 PS must use English language

Name	Multiplicity	Value	Type	Remarks
exchangeCatalogueName	1	CATALOG.101	CharacterString	Catalogue filename
exchangeCatalogueDescription	1		CharacterString	Description of what the exchange catalogue contains NATIONAL LANGUAGE enabled
productSpecification	1			S-101 Version Number
exchangeCatalogueComment	0..1		CharacterString	Any additional Information 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			Dataset name

12.2 Language (S-57 PS 3.11)

The exchange language must be English. Other languages may be used as a supplementary option. National geographic names can be left in their original national language in the international attributes, or transliterated or transcribed and used in the international attributes.

Character strings must be encoded using the character set defined in ISO 10646-1, in Unicode Transformation Format-8 (UTF-8). A BOM (byte order mark) must not be used.

