S-124 metadata and S-100 metadata requirements

Submitted by Canadian Coast Guard

SUMMARY
Executive Summary: The S-100 defined exchange set structure impose a discovery metadata file on each datasets of approximate 10KB. When S-124 datasets are estimated to average at 3-5KB, this is a significant data volume increase. Action to be taken: Determine path forward for S-124 data exchange. Related documents: S-124 Product Specification Draft 2.0.0.

1. The S-100 defined exchange set structure impose a discovery metadata file on each S-124 datasets of approximate 10KB, while preliminary tests indicate an average NW dataset to be about 3-5KB. This means that for the total exchange set, metadata will account for 66-75% of total data amount. This issue can to some extent be mitigated by compression. S-100 Ed 4.0.0 permit ZIP compression (see S-100 part 15). Another mitigation can be to use a different encoding (e.g. 8211 or HDF5), but since the metadata is external, the improvements offered is limited and likely offset by increased complexity in the production systems, such as when reusing the same data for different purposes, e.g. publishing on websites. Another option to reduce data volume is to use the Online Communication Exchange (see S-100 Part 14), which would not use the exchange set methodology, and rather send metadata at the beginning of a session and then send only the NW datasets. A drawback is that the specification is still new and largely untested. IALA ENAV committee is working on testbeds and improved documentation.

2. Figure A shows the complexity of dataset discovery metadata and includes a proposed S-124 dataset discovery metadata structure that significantly restrain the S-100 default dataset discovery metadata structure down to mostly the optional items. A similar, but less complex metadata structure is expected for support files and for catalogue files.
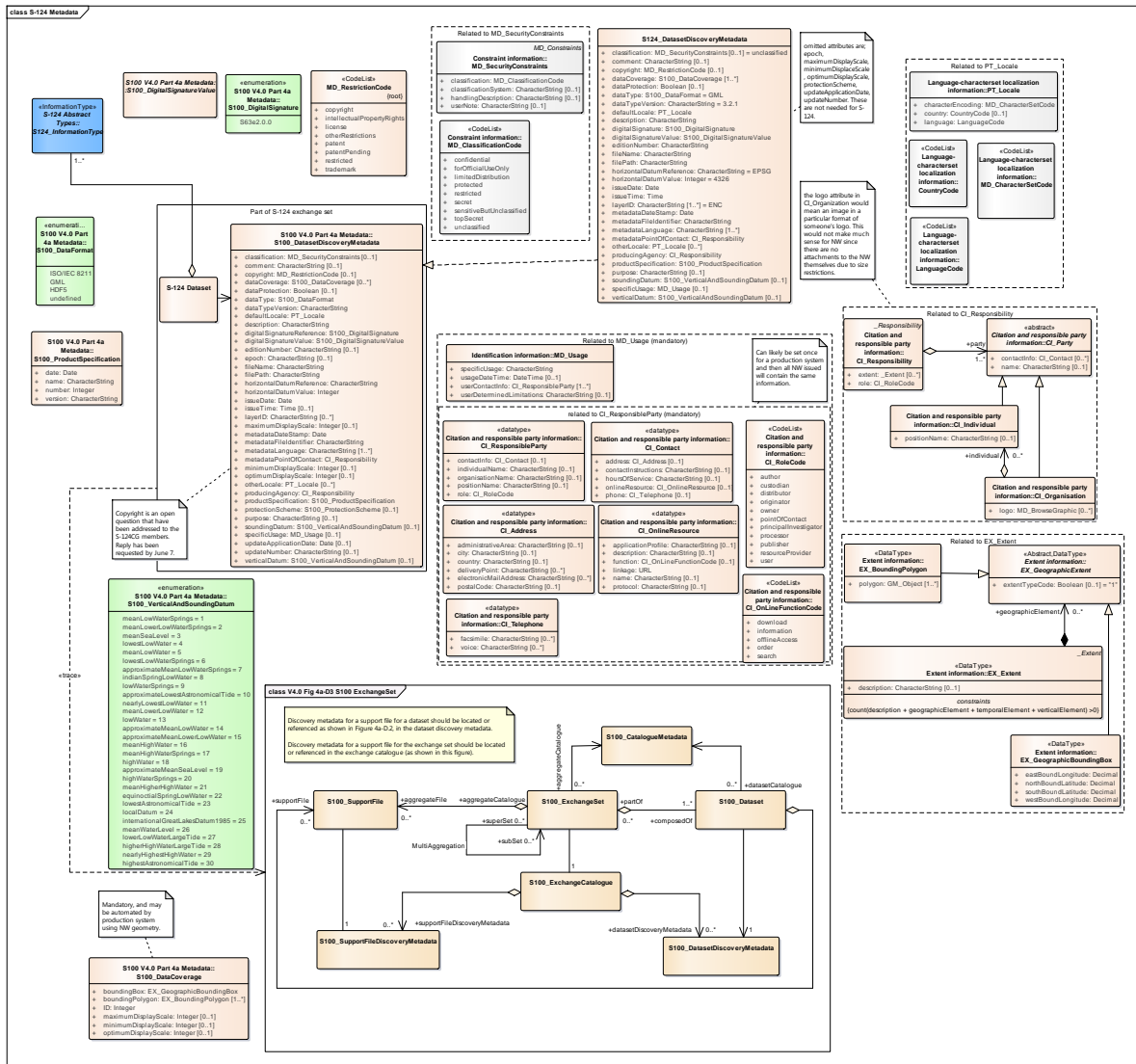
Figure A

3.   S-124 Metadata table below shows the draft structure of the dataset discovery metadata for S-124 datasets.

Metadata for datasets

Dataset discovery metadata for new datasets and new editions of published datasets is described in Table Error**! No text of specified style in document.**.1.

Name	Multiplicity	Value	Type	Remarks
S100_DataSetDiscovery Metadata			Class	The following S-100 attributes are not used: epoch, maximumDisplayScale, minimumDisplaceScale, optimumDisplayScale, protectionScheme, updateApplicationDate, updateNumber.
fileName	1		CharacterString	Dataset file name (see **Error! Reference source not found.**)
filePath	1		CharacterString	Full path from the exchange set root directory

description	1		CharacterString	Short description of the area covered by dataset, e.g., area, harbour, or port name, between two named locations etc.
dataProtection	0..1		Boolean	TRUE: Encrypted FALSE: Unencrypted
protectionScheme	0..1		S100_ProtectionScheme	See S-100 Appendix 4a-D.
digitalSignatureReference	1		S100_DigitalSignature	Specifies the algorithm used to compute digitalSignatureValue. S-100 Appendix 4a-D.
digitalSignatureValue	1		S100_DigitalSignatureValue	The value resulting from application of digitalSignatureReference. Implemented as the digital signature format specified in S-100 Part 15.
copyright	0..1		MD_LegalConstraints>MD_RestrictionCode <copyright> (ISO 19115-1)	"copyright" for copyrighted datasets, omitted otherwise
classification	0..1	(one of the literals from the ISO codelist)	Class MD_SecurityConstraints>MD_ClassificationCode (codelist) ISO 19115-1	"unclassified" for S-124 datasets
purpose	0..1	{1}	MD_Identification>purpose (character string)	"New dataset" for S-124 datasets
specificUsage	0..1		MD_USAGE>specificUsage (character string) MD_USAGE>userContactInfo (CI_Responsibility)	brief description of the resource and/or resource series usage
editionNumber	1		CharacterString	Edition number is always "1" for S-124 datasets.
issueDate	1		Date	Date on which the dataset was generated.
issueTime	1		Time	Encoded only if time of issue is significant.
productSpecification	1		S100_ProductSpecification	See Notes below this table for constraints on values.
producingAgency	1		CI_Responsibility>CI_Organisation or CI_Responsibility>CI_Individual	Party responsible for generating the dataset. See S-100 Part 4a Tables 4a-2 and 4a-3.
horizontalDatumReference	1	EPSG	CharacterString	
horizontalDatumValue	1	4326	Integer	WGS84
dataType	1	GML	S100_DataFormat	The only value allowed is "GML".
dataTypeVersion	1	3.2.1	CharacterString	
dataCoverage	1..*		S100_DataCoverage	See S-100 Appendix 4a-D.
comment	0..1		CharacterString	Any additional Information
layerID	1..*	S-101	CharacterString	Dataset must be used with ENC in an ECDIS.

defaultLocale	1		PT_Locale	S-100 Appendix 4a-D.
otherLocale	0..*		PT_Locale	S-100 Appendix 4a-D.
metadataFileIdentifier	1		CharacterString	For example, identifier for ISO 19115-3 metadata file
metadataPointOfContact	1		CI_Responsibility>CI_Individual or CI_Responsibility>CI_Organisation	See S-100 Part 4a Tables 4a-2 and 4a-3.
metadataDateStamp	1		Date	Metadata creation date, which may or may not be the dataset creation date
metadataLanguage	1..*		CharacterString	
--	0..*		Aggregation S100_SupportFileDiscovery Metadata	One for each support file linked to this dataset and present in the exchange set.

Table Error! No text of specified style in document.**.1 Dataset discovery metadata**
NOTES:

1) **Attribute** *productSpecification***: The values of sub-attributes** *name* **and** *version* **must correspond to this version of the S-124 product specification. (Clause** Error! Reference source not found.**). The value of sub-attribute** *number* **must be the number assigned to this version of the S-124 product specification in the GI registry.**

4. Options for dealing with the data volume is compression (zip), or a different data distribution paradigm. Online Data Exchange (Part 14) is the only other option defined in S-100.