Title: Metadata - Temporal Attributes

S-100 Maintenance - Change Proposal Form

Organisation	TWCWG (original) Raphael Malyankar (rev.)	Date	27-Aug-2021/ 25-Oct-2021 (rev. 1)
Contact	Raphael Malyankar	Email	raphaelm@portolansciences.com

Change Proposal Type *(Select only one option)*

1.Clarification	2.Correction	3.Extension
		X

Location (*Identify all change proposal locations*)

No.	S-100 Version No.	Part No.	Section No.	Proposal Summary
1	4.0.0	4a	App. 4a-D	Add optional discovery metadata attribute to indicate temporal validity of the dataset. (5.0.0 draft: X-X.X)
2		4a	App 4a-D	Add temporal validity as an optional attribute to the discovery metadata block for each coverage in a dataset. (5.0.0 draft X-X.X)
3		4a	App. 4a-D	Add discovery metadata attribute to indicate availability of successor dataset using MD_MaintenanceInformation. (5.0.0 draft: X-X.X)
4		4a	App. 4a-D	New clauses describing the ISO 19115-1 class MD_MaintenanceInformation, the S-100-specific restrictions on it in S-100 dataset discovery metadata, relevant ISO types, and examples.
5		1	3	Add ISO 19108 and the XML Schema datatypes specification to the S-100 references: ISO 19108:2002, Temporal Schema XML Schema Part 2: Datatypes Second Edition, W3C Recommendation, 28 October 2004, URL: https://www.w3.org/TR/xmlschema-2/

Change Proposal

Some S-104 (Water Level Information) and S-111 (Surface Currents) data will be published at regular intervals. For example, successive datasets covering a region may be released daily or even more frequently. This proposal requests the addition of discovery metadata indicating the temporal validity of a dataset and the interval before the next dataset can be expected to be available (by whatever means the producer uses for releasing datasets - push, stream, etc.).

Other products are also expected to need temporal validity and availability information, for example S-411 (Sea Ice), S-412 (Weather Hazards), etc., so a common solution at the S-100 level is appropriate.

Temporal validity metadata is needed for forecasts and other dynamic datasets. Water level adjustment needs temporal validity metadata so the adjustment procedures can filter available datasets by the planned dates and times of transits over different portions of the planned route.

*A "**closure**" attribute was considered but the case for using a "**closure**" attribute in temporal extent in discovery metadata is weak and it is not included in this proposal. See the discussion on the S-100 GitHub site at <https://github.com/IHO-S100WG/TSM8/issues>.*

*Attribute names **timeInstantBegin** and **timeInstantEnd** were selected in preference to **begin/end** as being unambiguous in a possible future IHO metadata attribute registry.*

Delivery information is needed for management of dataset transfer operations on end-user systems. In order to avoid the need to update the anticipated delivery information every time for each successive dataset, the interval between datasets is preferred to an indication of a specific time for the availability of successor dataset. A standard format for intervals has already been defined in ISO 8601 and implemented in XML Schema as a built-in datatype, and use of the standard format is preferred to facilitate data validation. The proposal is designed to facilitate automated operations for accessing data updates and/or notifications to human users of systems indicating when the availability of updated data should be checked.

The proposal for delivery information complies with the relevant portions of the ISO 19115-1 model (specifically, MD_MaintenanceInformation), with S-100-specific restrictions on the use of optional attributes and allowed values that are defined in the ISO model.

Item (1) Temporal validity:

[Add a temporalExtent attribute to the class S100_DiscoveryMetadata. Amend the UML diagram in Figure 4a-D-4 to include the attribute and related classes.]

Role Name	Name	Description	Mult	Type	Remarks
Class	S100_Dataset Discovery Metadata	(see S-100)			
Attribute	temporalExtent	Specification of the temporal extent of the dataset.	0..1	S100_Temporal Extent	The temporal extent is encoded as the date/time of the earliest and latest data records (in coverage datasets) or date/time ranges (in vector datasets). If there is more than one feature in a dataset, the earliest and latest time values of records in all features are used, which means the earliest and latest values may be from different features. If date/time information for a feature is not encoded in the dataset, it is treated for the purposes of this attribute as extending indefinitely in the appropriate direction on the time axis, limited by the issue date/time or the cancellation or supersession of the dataset. This attribute is encoded if and only if at least one of the start and end of the temporal extent is known.

[New documentation table for S100_TemporalExtent]

S100_TemporalExtent

Role Name	Name	Description	Mult	Type	Remarks
Class	S100_Temporal Extent	Temporal extent	--		At least one of the timeInstantBegin and timeInstantEnd attributes must be populated; if both are known, both must be populated. The absence of either begin or end indicates indefinite validity in the corresponding direction, limited by the issue date/time or the cancellation or supersession of the dataset.
Attribute	timeInstantBegin	The instant at which the temporal extent begins.	0..1	DateTime	
Attribute	timeInstantEnd	The instant at which the temporal extent ends.	0..1	DateTime	

NOTES:
(1) In case of overlap in temporal extent between predecessor and successor datasets, the successor dataset prevails. For example, water level or weather forecast datasets may have a temporal extent of N days or hours, but be replaced by new forecast at N - X.
(2) Precedence and succession can be determined from information in dataset discovery metadata (e.g., attributes for dataReplacement, edition and update numbers, issue data and time).

EXAMPLE 1: An S-104 (Water Level Information for Surface Navigation) predictions dataset has the following data for *temporalExtent* encoded in the dataset discovery block in the exchange catalogue:

```
<temporalExtent>
    <timeInstantBegin>2021-07-03T06:00:00Z</timeInstantBegin>
    <timeInstantEnd>2021-07-10T18:00:00Z</timeInstantEnd>
</temporalExtent>
```

indicating that the temporal extent of the predictions in the dataset is the period beginning at exactly 6 a.m. on 3 July 2021 (UTC) and ending at exactly 6 p.m. on 10 July 2021 (UTC).

EXAMPLE 2: The successor dataset to Example 1 has the following data for *temporalExtent*:

```
<temporalExtent>
    <timeInstantBegin>2021-07-03T12:00:00Z</timeInstantBegin>
    <timeInstantEnd>2021-07-10T24:00:00Z</timeInstantEnd>
</temporalExtent>
```

indicating that the temporal extent of the predictions in the dataset is the period beginning at exactly noon on 3 July 2021 (UTC) and ending at exactly midnight at the end of 10 July 2021 (UTC). Since this temporal extent overlaps the temporal extent of Example 1 from noon UTC on 3 July 2021, it supersedes the dataset in Example 1 at and after noon UTC on 3 July 2021.

Item (2) Add temporal validity to metadata for data coverages within a dataset.

S100_DataCoverage

Role Name	Name	Description	Mult	Type	Remarks
Class	S100_DataCoverage	(see S-100)			
Attribute	temporalExtent	Specification of the temporal extent of the coverage.	0..1	S100_TemporalExtent	The remarks for *temporalExtent* in the dataset discovery block (S100_DatasetDiscoveryMetadata) block apply, except that their scope is the individual coverage and not the dataset as a whole.

Item (3) Availability of successor dataset:

[Add a resourceMaintenance block to the class S100_DatasetDiscoveryMetadata, modelled as in ISO 19115-1. (Note that ISO 19115-1 models resourceMaintenance as a role in MD_Identification and MD_Metadata.) Add S-100 restrictions on the ISO datatypes. Add examples in the text. Add the relevant ISO datatypes to Figure 4a-D-4.]

Role Name	Name	Description	Mult	Type	Remarks
Class	S100_Dataset Discovery Metadata	(see S-100)			
Role	resourceMaintenance	Information about the frequency of resource updates, and the scope of those updates	0..1	MD_MaintenanceInformation	S-100 restricts the multiplicity to 0..1 and adds specific restrictions on the ISO 19115 structure and content. See clause MD_MaintenanceInformation later in this Part. Format: PnYnMnDTnHnMnS (XML built-in type for ISO 8601 *duration*). See Notes.

Item (4) New clauses describing MD_MaintenanceInformation in S-100 metadata and relevant types.
[Since S-100 does not extend the ISO class, only restricts it, an S100_ prefix is not needed.]

[NOTE: This model does not provide for explicitly encoding expected variation, for example "30 days ± 3 days. Comments are invited on whether such explicit encoding is necessary, in the form of an additional "intervalVariation" or similar attribute.]

MD_MaintenanceInformation

Role Name	Name	Description	Mult	Type	Remarks
Class	MD_Maintena nceInformatio n	Information about the scope and frequency of updating.	--	--	S-100 restricts the ISO 19115-class to: • prohibit maintenanceScope, maintenanceNote, and contact attributes; • define restrictions on maintenanceAndUpdateFrequ ency, maintenanceDate, and userDefinedMaintenanceFreq uency attributes.
Attribute	maintenance AndUpdateFr equency	frequency with which changes and additions are made to the resource after the initial resource is completed	0..1	MD_MaintenanceFre quencyCode (codelist)	Must be populated if userDefinedMaintenanceFrequenc y is not present, otherwise optional. See table MD_MaintenanceFrequencyCode in this Part for values allowed in S-100 metadata.
Attribute	maintenance Date	date information associated with maintenance of resource	0..1	CI_Date	Exactly one of maintenanceDate and userDefinedMaintenanceFrequenc y must be populated. Allowed value for dateType: nextUpdate
Attribute	userDefinedM aintenanceFr equency	maintenance period other than those defined	0..1	TM_PeriodDuration	Exactly one of maintenanceDate and userDefinedMaintenanceFrequenc y must be populated. Only positive durations allowed.

NOTE 1: The interval described by *userDefinedMaintenanceFrequency* is with respect to the issue date and time of the dataset described by this dataset discovery metadata block. End-user's and distributor's systems should use this interval for planning any automated operations to obtain the successor dataset, but must allow for delays or variations in the actual availability of successor dataset(s).

NOTE 2: The format for *userDefinedMaintenanceFrequency* is given by the XML built-in datatype *duration*, which can be validated by off-the-shelf XML parsers. See "*XML Schema Part 2: Datatypes (2nd edition) - Clause 3.2.6 duration*" (relevant extracts below):

The lexical representation for **duration** is the ISO 8601 extended format PnYnMnDTnHnMnS, where nY represents the number of years, nM the number of months, nD the number of days, 'T' is the date/time separator, nH the number of hours, nM the number of minutes and nS the number of seconds. The number of seconds can include decimal digits to arbitrary precision.

The values of the Year, Month, Day, Hour and Minutes components are not restricted but allow an arbitrary unsigned integer, i.e., an integer that conforms to the pattern [0-9]+.. Similarly, the value of the Seconds component allows an arbitrary unsigned decimal. Following ISO 8601, at least one digit must follow the decimal point if it appears.

Reduced precision and truncated representations of this format are allowed provided they conform to the following:

- If the number of years, months, days, hours, minutes, or seconds in any expression equals zero, the number and its corresponding designator ·may·

be omitted. However, at least one number and its designator ·must· be present.

- The seconds part ·may· have a decimal fraction.

- The designator 'T' must be absent if and only if all of the time items are absent. The designator 'P' must always be present.

Encoding and interpretation rules in S-100 metadata:

(1) <u>Restriction to non-negative durations</u>: S-100 restricts the *duration* type by prohibiting zero or negative values of duration in *userDefinedMaintenanceFrequency*.

(2) <u>Number of digits</u>: S-100 recommends (but does not require) using 2 digits for the months, days, hours, minutes, components, when they are present. If the seconds component is encoded, two digits are recommended for the number of whole seconds (for example, encode 0.5 seconds as PT00.5S; encode 100 seconds as PT01M40S).

(3) <u>Start and end instants</u>: The start and end instants of the interval calculated by combining *userDefinedMaintenanceFrequency* with the issue date/time must be interpreted according to Part 3 Clause 3-8. The value must be encoded appropriately; this means that smaller date/time components must not be encoded unless the availability of the successor dataset is known to the corresponding level of precision. Smaller units should be used when the availability is known to the corresponding precision, such as "48 hours" instead of "2 days" when the successor dataset availability is planned to the hour.

(4) <u>Encoding of zero components</u>: Zero components must be encoded if and only if they are significant for indicating the granularity of the start/end instants of the interval.

(5) <u>Variability</u>: A variation of $\pm X$ should be allowed for, where X is the component of smallest granularity; if the value of the smallest component is 1, variability is unspecified.

Commented [rmm1]: See request for input on encoding (or not encoding) variations in the note at the beginning of Item 4

(6) <u>Stability for successive datasets, and exceptions</u>: The value of this attribute will normally be stable over a sequence of predecessor/successor datasets. The alternate encoding using *maintenanceDate* should be used for known exceptional circumstances affecting the release of a successor, such as an office closure at the end of the intervening period, reverting to normal encoding with *userDefinedMaintenanceFrequency* when the normal update schedule is restored.

(7) <u>Off-schedule updates</u>: Communication of exceptional, unforeseeable off-schedule issues of data such as emergency hurricane forecasts should be provided for by other means than *userDefinedMaintenanceFrequency* or *maintenanceDate* attributes, since they are by definition unforeseeable.

(8) <u>Supersession</u>: If both *userDefinedMaintenanceFrequency* and *maintenanceDate* are encoded in the same discovery metadata block, the *maintenanceDate* supersedes the *userDefinedMaintenanceFrequency*.

EXAMPLES

No.	maintenanceAnd UpdateFrequency	maintenanceDate	userDefined Maintenance Frequency	Remarks
1	--	--	P3DT10H30M	An interval of 3 days, 10 hours, and 30 minutes. Variability +/-1 minute.
2	--	--	PT6H	An interval of exactly 6 hours, with a variability of +/1 hour.
3	--	--	P30M	an interval of 30 months.

4	--	--	PT30M	an interval of 30 minutes.
5	--	--	P6H P30S P30M10S	invalid (they contain time components but lack the 'T' designator)
6	--	--	PT30m	Invalid ('m' should be upper-case).
7	--	--	PT12:30 P3DT10H 30M	Invalid (the ':' or space separators are not allowed, only the separators specified by the XML Schema datatypes specification for *duration* are allowed)
8	--	--	P1M	One month, variability unknown. According to the "Start and end instants" rule, will be interpreted as the same day in the following month, or the nearest preceding day if there is no such date in the following month. If the issue date of the current dataset is 30 August, the successor dataset can be expected to be issued between midnight at the beginning of 30 September and midnight at the end of 30 September.
9	--	--	P1M00D	One month, with a variability of +/- 1 day. With a dataset issued on January 31 2021, the next dataset is expected on February 28, 2021; with a dataset issued on January 31, 2024 means the next dataset is expected February 29, 2024. A 1-day variation before after those dates should be anticipated.
10	--	--	P30D	30 days, variability +/- 1 day. With a dataset issued on January 31, 2021 it means the next dataset is expected on March 2, 2021; with a dataset issued on January 31, 2024 it means the next dataset is expected on March 1, 2024. A 1-day variation should be allowed for in both cases.
11	irregular	cit:CI_Date > cit:dateType=nextUpdate cit:date=2021-10-25	--	On 25 October 2021, at an unspecified time on that date.

12	irregular	cit:CI_Date > cit:dateType=nextUpdate cit:date=2021-10-25T14:00:00Z	--	On 25 October 2021, at 2 p.m. UTC.
13	asNeeded	cit:CI_Date > cit:dateType=nextUpdate cit:date=2021-10-25T14:00:00Z	--	To encode an exception to a dataset sequence normally on a regular schedule. Next dataset will be available on 25 October 2021, at 2 p.m. UTC.

NOTES:

1) The codeList attributes must be populated with the URL of the appropriate codelist, which will be made known in the S-100 schema distribution documentation.

2) The ISO schemas permit a codelist element to have its content empty, or any text in any single language, but not multilingual text, ISO schemas do not allow free text here. Product specification authors may prescribe content, but note that applications are not required to do any processing of the element content, and may or may not display it. (To obtain information about the value, the key encoded in the XML attribute *codelistValue* must be used to look up the entry in the codelist identified by attribute *codeList.*)

XML encoding examples:

EXAMPLE 1: Dataset is updated at an interval of 6 hours:

```
<mri:resourceMaintenance>
    <mmi:MD_MaintenanceInformation>
        <mmi:userDefinedMaintenanceFrequency>
            <gco:TM_PeriodDuration>PT06H</gco:TM_PeriodDuration>
        </mmi:userDefinedMaintenanceFrequency>
    </mmi:MD_MaintenanceInformation>
</mri:resourceMaintenance>
```

EXAMPLE 2: Dataset is normally updated on a regular schedule, but the next update will be on 1 January 2022 at 5 a.m. local time in the time zone with UTC offset -5 hours (e.g., 5 a.m. US Eastern Standard Time). The codeList attributes must be populated with the URL of the appropriate codelist, which will be in the ISO or S-100 schema distribution package.

```
<mri:resourceMaintenance>
    <mmi:MD_MaintenanceInformation>
        <mmi:maintenanceAndUpdateFrequency>
            <mmi:MD_MaintenanceFrequencyCode codeList="http://...."
codeListValue="asNeeded">
                empty, or any text in any single language
            </mmi:MD_MaintenanceFrequencyCode>
        </mmi:maintenanceAndUpdateFrequency>
        <mmi:maintenanceDate>
            <cit:CI_Date>
                <cit:date>
                    <gco:DateTime>2022-01-01T05:00:00-05:00</gco:DateTime>
                </cit:date>
                <cit:dateType>
                    <cit:CI_DateTypeCode codeList="http://..."
codeListValue="nextUpdate">
                        empty, or any text in any single language
                    </cit:CI_DateTypeCode>
                </cit:dateType>
            </cit:CI_Date>
```

```
            </mmi:maintenanceDate>
        </mmi:MD_MaintenanceInformation>
</mri:resourceMaintenance>
```

EXAMPLE 3: Dataset has no consistent update schedule. The next update will be on 1 January 2022 at an unspecified time.

```
<mri:resourceMaintenance>
    <mmi:MD_MaintenanceInformation>
        <mmi:maintenanceAndUpdateFrequency>
            <mmi:MD_MaintenanceFrequencyCode        codeList="http://...."
codeListValue="irregular"/>
        </mmi:maintenanceAndUpdateFrequency>
        <mmi:maintenanceDate>
            <cit:CI_Date>
                <cit:date>
                    <gco:Date>2022-01-01</gco:Date>
                </cit:date>
                <cit:dateType>
                    <cit:CI_DateTypeCode        codeList="http://...."
codeListValue="nextUpdate"/>
                </cit:dateType>
            </cit:CI_Date>
        </mmi:maintenanceDate>
    </mmi:MD_MaintenanceInformation>
</mri:resourceMaintenance>
```

MD_MaintenanceFrequencyCode

S-100 uses a subset of the values allowed in ISO 19115-1.

Role Name	Name	Description	Code	Remarks
Enumeration	MD_MaintenanceFrequencyCode	frequency with which modifications and deletions are made to the data after it is first produced	-	S-100 is restricted to only the following values from the ISO 19115-1 codelist. The conditions for the use of a particular value are described in its Remarks.
Value	asNeeded	resource is updated as deemed necessary	-	Use only for datasets which normally use a regular interval for update or supersession, but will have the next update issued at an interval different from the usual. Allowed if and only if userDefinedMaintenanceFrequency is not populated.
Value	irregular	resource is updated in intervals that are uneven in duration	-	Use only for datasets which do not use a regular schedule for update or supersession. Allowed if and only if userDefinedMaintenanceFrequency is not populated.

CI_DateTypeCode

This codelist is documented in the ISO schemas documentation, available in the S-100 schemas distribution. It is used in several places in S-100 metadata.

Changes to Figure 4a-D-4 (common to both temporal extent and maintenance information):
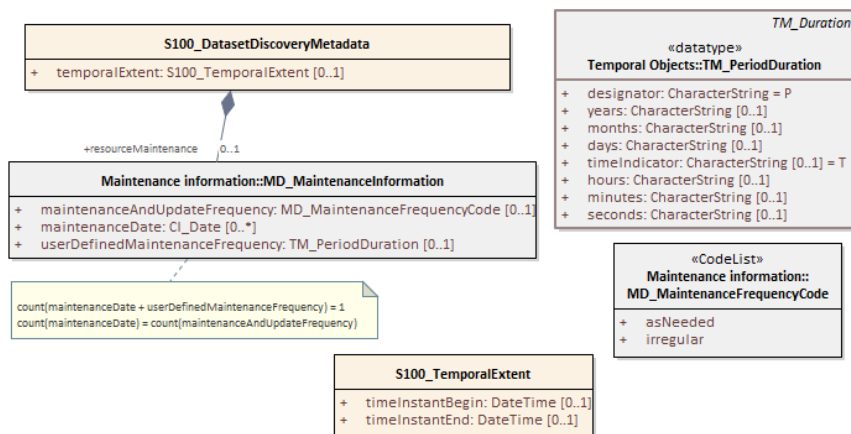
Change Proposal Justification

(1) S-104 (Water Level Information) and S-111 (Surface Currents) envisage the issue of successive water level information and surface currents datasets at regular intervals. Selection of the proper dataset for water level adjustment, route planning, and also portrayal in general requires information about the start and end times delimiting the period when the data is valid. Certain other data products are also valid for specific periods. A common method of indicating validity periods is necessary so that custom programming and/or opening/loading HDF5 files (or other formats) is not needed to determine the validity period of the data.

(2) The intervals between successive datasets are expected to vary at present from a few hours to monthly, depending on the type of water level or current data and the producer. (Longer and shorter intervals are possible in the future.) Certain other data products are also expected to issue successive datasets at intervals, for example sea ice and WMO meteorological data products. To facilitate management of data products and planning of operations, as well as planning by end-users, information about the expected availability of successor datasets is needed in a machine-readable form that can be used by distributors and end-user systems.

(3) An alternative to delivery interval is to encode the actual anticipated date/time of the successor dataset. This would require updating the corresponding attribute(s) in the discovery metadata block for each successive dataset, which is more complex and more susceptible to human or computer error.

(4) The end instant of temporal extent is not an acceptable proxy for delivery interval because it is not necessarily true that the successor dataset is available at the very instant the validity ends and not before or after, so it is better to separate availability information from dataset validity information. For example, an S-111 or S-104 hydrodynamic model forecast could include predictions/forecasts out to 180 hours (or beyond), whereas the model produces a forecast (and *datasetDeliveryInterval* would be) every 6 hours ("6 hourly cycle").

What parts of the S-100 Infrastructure will this proposal affect?

☐   S-100 Feature Concept Dictionary Interface or Database

☐   S-100 Portrayal Register
☐   S-100 Feature Catalogue Builder
☐   S-100 Portrayal Catalogue Builder
☒   S-100 UML Models
☒   S-100 GitHub Schemas

Please send completed forms and supporting documentation to the secretary S-100WG.