

Title: Metadata - Temporal Attributes**S-100 Maintenance - Change Proposal Form**

Organisation	TWCWG	Date	27-Aug-2021
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 discovery metadata attribute to indicate temporal validity of dataset.
2		4a	App. 4a-D	Add discovery metadata attribute to indicate availability of successor dataset.
3		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. 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.

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 records in the dataset. 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. 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]

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.
Attribute	timeInstantBegin	The instant at which the temporal extent begins.	0..1	dateTime	Must be a dateTime value in XML Schema dateTime format. E.g., 2021-08-03T06:00:00Z for 6 a.m. UTC on August 3, 2021.
Attribute	timeInstantEnd	The instant at which the temporal extent ends.	0..1	dateTime	Must be a dateTime value in XML Schema dateTime format. E.g., 2021-08-03T06:00:00Z for 6 a.m. UTC on August 3, 2021.

Item (2) Availability of successor dataset:

[Add a datasetDeliveryInterval attribute to the class S100_DatasetDiscoveryMetadata. Add the relevant ISO datatype to Figure 4a-D-4. Add a Note describing the implementation format below the documentation table for S100_DatasetDiscoveryMetadata.]

Role Name	Name	Description	Mult	Type	Remarks
Class	S100_Dataset Discovery Metadata	(see S-100)			
Attribute	dataset Delivery Interval	The expected time interval between availability of successive datasets for time-varying data.	0..1	CharacterString <TM_PeriodDuration>	Format: PnYnMnDTnHnMnS (XML built-in type for ISO 8601 duration). See Notes.

NOTE 1: The interval described by *datasetDeliveryInterval* 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 *datasetDeliveryInterval* 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.

S-100 restricts the ISO 8601/XML format by disallowing negative values.

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).

The start and end instants of the interval 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.

Zero components must be encoded if and only if they are significant for indicating the granularity of the start/end instants of the interval.

A variation of $\pm X$ should be allowed for, where X is the component of smallest granularity (see Examples 7 *et seq.* below).

EXAMPLE 1: P3DT10H30M and P0Y0M3DT10H30M both indicate an interval of 3 days, 10 hours, and 30 minutes.

EXAMPLE 2: P0Y0M0DT06H00M00S, PT6H, PT06H00M00S, PT6H0.00S all indicate an interval of exactly 6 hours.

EXAMPLE 3: P30M indicates an interval of 30 months, not 30 minutes. PT30M indicates an interval of 30 minutes.

EXAMPLE 4: P6H, P30S, and P30M10S are invalid (they contain time components but lack the 'T' designator).

EXAMPLE 5: PT30m is invalid ('m' should be upper-case).

EXAMPLE 6: PT12:30 and P3DT10H 30M are invalid (the ':' or space separators are not allowed, only the separators specified by the XML Schema datatypes specification for *duration* are allowed).

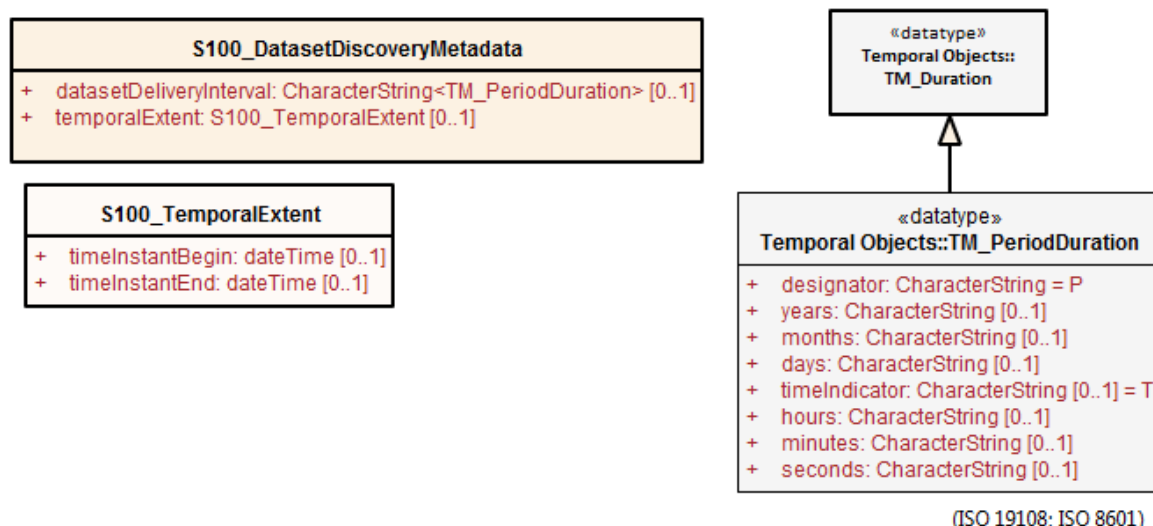
EXAMPLE 7: P1M00D means 1 month and zero days, and according to the interpretation rule given earlier will be interpreted as the same day in the following month, or the nearest preceding day if there is no such date in the following month. If the issue date of the current dataset is 30 August, the successor dataset can be expected to be issued between midnight at the beginning of 30 September and midnight at the end of 30 September. P1M on the other hand will be interpreted as between 30 September and 29 October.

EXAMPLE 8: P1M00D with a dataset issued on January 31 2021 means the next dataset is expected February 28, 2021; with a dataset issued on January 31, 2024 means the next dataset is expected February 29, 2024.

EXAMPLE 9: P30D means 30 days; 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.

[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 "deliveryIntervalVariation" or similar attribute.]

Changes to Figure 4a-D-4 (common to 1 & 2):



(ISO 19108; ISO 8601)

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.