

## Annex A

### Validation Checks - Concepts, Terms, Syntax, Operations

#### 1 Introduction

This Annex defines validation checks for developers of <PS>. The checks can be administered at any time during the production phase. They can also be applied downstream in the distribution and end user systems to test the conformance of a dataset to (S-100 Part 10a/10b/10c and <PS>).

The checks are based on the data quality measures and elements described in S-97 Part C and <PS>.

#### 2 References

S-58 IHO Publication S-58, ENC Validation Checks, Edition 6.1.0, September 2018.

S97C IHO S-97 Part C IHO data quality checklist Edition 1.0.0, [Draft, March 2019].

S-100 IHO S-100 Universal hydrographic data model. Edition 5.0.0, [TBD].

#### 3 Check classification

##### 3.1 General classification

Checks are classified as critical, error, or warnings as described in Table B-1.

**Table B-1 - Classification of checks**

C	Critical Error	An error which would make a dataset unusable in ECDIS through not loading or causing an ECDIS to crash or presenting data which is unsafe for navigation.
E	Error	An error which may degrade the quality of the dataset through appearance or usability but which will not pose a significant danger when used to support navigation.
W	Warning	An error which may be duplication or an inconsistency which will not noticeably degrade the usability of a dataset in ECDIS.

##### 3.2 Specialized classifications

[Insert language about specialized classifications being allowed to elaborate on the general classification.]

(Example from S-98.)

Checks are also categorized according to their scopes (see Clause 6 in “S-98 – Main”), as follows:

- Checks which apply to the Interoperability Catalogue itself are categorized as “IC” checks.
- Checks which apply to the output of interoperability operations in the interoperability catalogue are categorized as interoperability output (“OP”) checks.

#### 4 Check application

Checks do not apply to dataset terminations or cancellations, except where the check description explicitly states it applies in case of a termination or cancellation.

##### 4.1 HDF5 check application

[S-111 language example]

The checks apply to each HDF5 file which constitutes a dataset, in the S-100 sense of “dataset”.

There being no update dataset format defined in S-111 Edition 1.0.x, checks are not designated as applying to “base” or “update” datasets.

## 4.2 GML check application

Checks are designated as applying to base datasets, update datasets, post-update datasets, or some combination of the three.

**Table 2. Check application annotations**

B	Base	Apply check to new dataset, new edition, and post-update dataset (after updates have been applied to the base).
U	Update	Apply check to update datasets in isolation.
S	Post-update	Apply check only to a post-update dataset, i.e., subsequent to application of all available updates.

Checks do not apply to dataset terminations or cancellations, except where the check description explicitly states it applies in case of a termination or cancellation.

## 5 Check operations

### 5.1 Comparison and Logical Operators

The following comparison and logical operators are used:

- Equal
- Not equal
- Less than
- Less than or equal to
- Greater than
- Greater than or equal to
- AND
- OR (inclusive OR)

### 5.2 Spatial Operators

Within this document the spatial operators (EQUALS, DISJOINT, TOUCHES, WITHIN, OVERLAPS, CROSSES, INTERSECTS, CONTAINS, and COINCIDENT), based on those laid out in the ISO standard 19125-1, are used to describe spatial relationships tested within the checks.

For all spatial operators a default tolerance should be applied in validation software which is equivalent to 1 pixel on an ECDIS of the lowest acceptable resolution according to the performance standards at the minimum display scale of the S-101 dataset involved in the test.

### 5.3 Values

The following terms are used for types of values:

- Present – An attribute is present and has been populated with:
  - a value (for attributes not derived from XML string type), or
  - the empty string (for attributes derived from the XML string type), or

- empty content, and is nilled as specified in the W3C XML specification.
- Null – An attribute is present and has no content or is nilled as specified in the XML specification.
- notNull – The attribute is present and has been populated with a value.

## 5.4 Statements

The checks must be structured using the following statements:

- If – A conditional statement which determines whether a further statement should be executed.
- For – Repeat a statement until a statement is met (evaluates to “true”). For the purposes of the checks the statement being met generates the error or warning specified.

## 6 Geometry and spatial operators: Terms and definitions

The terms and definitons of geometry and spatial operators are as described in S-58 Clause 2.

[To do: Integrate that material from S-58 into this document.]

## 7 Application of validation checks

### 7.1 Validation processing for GML formats

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas porttitor congue massa.

Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna. Nunc viverra imperdiet enim.

### 7.2 Validation processing for HDF5 formats

Checks are divided into five phases:

**Table 3. Phases in validation processing**

Phase	Name	Description
1	Validate Dataset Root and Feature Information	Validation of root group of HDF5 file and feature type information.
2	Validate Feature Container Groups	Validation of metadata and structure for each feature type (“Feature Container”). In S-111 1.0.X there is only one feature container, so this set of checks is executed only once. If future editions introduce multiple feature container HDF5 groups, this set must be executed for each feature container HDF5 group.
3	Validate Feature Instance Groups	Validation of feature instances. This set of checks, along with Phase 4 and 5 checks, must be executed once for each feature instance group contained within a feature container.
4	Validate Position Information	Validation of positioning data. This set of checks is executed if and only if the data coding format requires the presence of explicit position arrays.
5	Validate Values Datasets	Validation of values data. This set of checks is executed for each values group in a feature instance group.

Figure 1 below depicts the sequence of processing. Certain check failures in Phases 1-3 prevent progress to later phases (because information needed to test conditions is not available). If one of these checks fails, processing of other checks in the current phase is allowed to proceed, but subsequent phases cannot be executed due to a lack of necessary information.

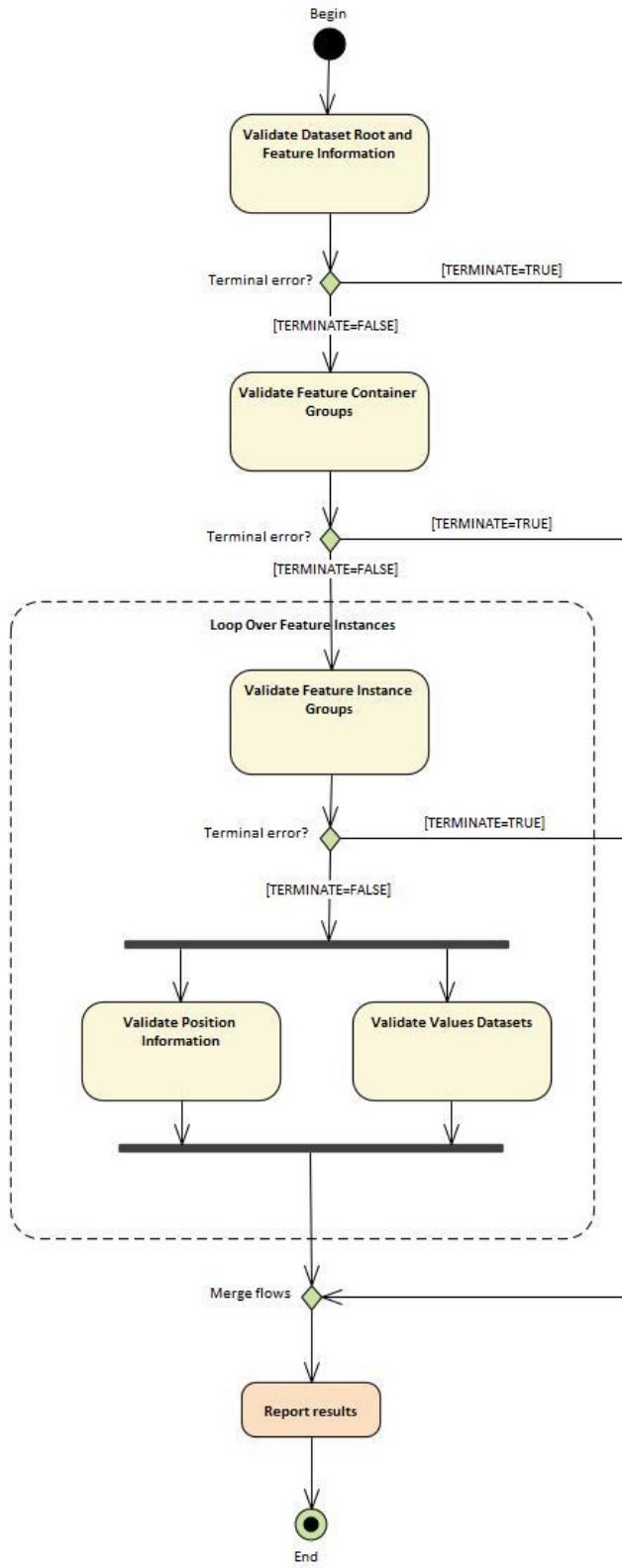


Figure 1. Check processing

## 8 Check syntax

In order to ensure that checks can be interpreted clearly and consistently a defined syntax has been used for the reworded checks wherever possible. Each check is a statement which generates a Critical Error, Error or Warning if the expression returns 'true'.

The syntax for check descriptions is the same for all formats. The differences are in the additional information as to the applicability (base or update datasets) and in the addition of prerequisites and contexts needed for the HDF5 checks.

### 8.1 Check syntax for GML formats

**Table 4. Example of syntax for GML formats**

No	Check description	Check message	Check solution	Conformity to	Category	Apply to
131	For each feature instance where periodicDateRange subattribute is notNull AND dateStart is Null OR not Present.	Object has dateEnd without a value of dateStart.	Populate dateStart or remove dateEnd.	Logical consistency	E	B, U

The check described above would return true and give an error if there is no starting date for a periodic date range complex attribute.

### 8.2 Check syntax for HDF5 formats

Individual checks are defined in the following format:

**Table 5. Check specification format**

Column	Description
Quality Measure (see S-97 Part C)	Quality measure from S-97 Part C. If two measures are included in this column, the Comments column explains how the error should be classified.  Note: Pending advice from DQWG about classification of checks for embedded metadata, errors in the values of <i>metadata</i> attributes are classified as "Conceptual Consistency" errors (breaches of rules of the conceptual schema, which controls the values assigned to metadata attributes) rather than "Domain Consistency" errors (breaches of rules about ranges of allowed values). "Domain Consistency" is more appropriate for <i>data</i> attribute values rather than <i>metadata</i> . This distinction should produce quality measure results which better reflect the actual data quality, by separating data errors from metadata errors.
Check No.	Identifying number for check
Prerequisite check(s)	Checks which must succeed (check condition evaluates to FALSE) before this check can be executed.

Column	Description
Context test (IF ...) or initialization (SET ...)	<p>Combination of test conditions and initialization statements.</p> <p>Test conditions check for the existence of an HDF5 attribute, group, or other element (e.g., an HDF5 array), or test the value of a metadata attribute.</p> <p>Initialization statements set the value of parameters used in the specific test in that row.</p> <p>The scope of the test condition or initialization is limited to the check described in that row.</p>
Check condition description	<p>Specification of check condition, written in structured English.</p> <p>The conditions are written so that if the condition evaluates to TRUE it indicates an error or issue in the dataset.</p>
Check message	Message to emit if check condition fails (evaluates to TRUE).
Check solution	Solution to be applied to correct the failure.
Category	Whether check failure is a critical, error, or warning issue. See Table 1.
Post-condition	<p>Action to be executed if the check condition evaluates to TRUE (i.e., if the check fails).</p> <p>This action will generally either set a global flag to control check processing (e.g., "SET TERMINATE=TRUE") or set a variable in the processing context which is used in later checks (e.g., set a context variable to store the value of the metadata attribute <i>dataCodingFormat</i>).</p>
S-100 Ed. 4.0.0 reference	Reference to place in S-100 where more information about the check can be found, e.g., lists of allowed values for enumerations.
S-111 Ed. 1.0.1 reference	Reference to place in S-111 where more information about the check can be found, e.g., allowed values for attributes of enumeration types.
Comments	Explanatory remarks or additional notes.

### 8.2.1 Phase initialization

Certain parameters need to be initialized before processing of the phase begins. The required initialization statements are indicated in a sub-head row at the beginning of each phase.

### 8.2.2 Check specifications

The individual checks are specified in an Excel spreadsheet accompanying this document.

Words in angle brackets <> indicate the content is a parameter which must be substituted by the appropriate value. E.g., <FX> (Phase 1) should be replaced by the appropriate feature code (the code "SurfaceCurrent" for S-111 1.0.x).

Bold type indicates a literal name (e.g., **Group\_F.featureCode** means the HDF5 array named "featureCode" in the HDF5 group named "Group\_F").