



# S-111 HDF5 Validation Checks

Raphael Malyankar

On behalf of Office of Coast Survey, NOAA, USA

*5<sup>th</sup> Tides, Water Level and Currents Working Group (TWCWG5)*

*Remote VTC*

*16-18 March 2021*



IHO

# VALIDATION CHECKS



International  
Hydrographic  
Organization

- Checks verify integrity, consistency, and compliance to the product specification
- Checks for individual datasets (HDF5 files)
  - Verify the content (data objects)
  - Verify the structure of the HDF5 file
  - Verify embedded metadata
  - Verify mutual consistency of structure, content, and embedded metadata
- Checks for exchange sets
  - Exchange set format, structure, and integrity
  - Completeness – whether all components are included
  - Exchange set catalogue validation – check discovery metadata and catalogue content



Required Product Specification component	Level 1 v1.0.0	Level 2 v1-2.0.0	Level 3 >v2.0.0	Level 4 >v2.0.0	Level 5 >v2.0.0
Main Document (Defines the relevant parts of S-100 that are required for the Product Specification)	X	X	X	X	X
<i>A Default Encoding</i>	X	X	X	X	X
S-100 Compliant Feature Catalogue	X (draft)	X (updated)	X (final, from IHO GI Registry)	X	X
<i>Data Classification and Encoding Guide</i>	X (draft)	X	X (final)	X	X
S-100 Compliant Portrayal Catalogue NOTE: Not every Specification will need a Portrayal Catalogue – this should be determined as part of the development process and stakeholder feedback.		X	X	X	X
Data Quality Checks		X	X	X	X
Test Data Sets		X	X	X	X
<i>Data Validation (and test datasets)</i>		X	X	X	X
Exchange Catalogue		X	X	X	X
Encryption / Digital Signatures			X	X	X
Interoperability			x* (draft)	X* (tested)	X*
Alerts and Indications				X*	X*
Operational data					X

(X\* = ECDIS only)

- S-100 Readiness Level (S-97 1.1.0)
  - Communicates the completeness of the specification and its readiness for operational use.
  - Validation checks are a requirement for Level 2 (“the product specification can be demonstrated in prototype environments”).
- S-97 Part C – Data Quality
  - “...intended to ensure that data quality aspects are addressed in an appropriate and harmonized way for all S-100 based Product Specifications.”
  - Makes 10 recommendations for data quality measures in S-100-based products in general - product specifications can customize them.
  - The validation checks produce quality results for certain quality measures, depending on the test.



IHO

# CLASSIFICATION OF CHECKS



International  
Hydrographic  
Organization

- **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.

- **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.

- **Warning**

- An error which may be duplication or an inconsistency which will not noticeably degrade the usability of a dataset in ECDIS.



IHO

# OVERVIEW OF S-111 VALIDATION



International  
Hydrographic  
Organization

- Checks for S-111 datasets.
  - A generic check specification format is being developed at the “S-100 level”. This effort is currently focused on checks for vector formats, e.g., S-101.
- A full set of dataset validation checks for S-111 datasets has been prepared as a NOAA-sponsored effort.
  - Current version is for Edition 1.0.0, to be updated for Edition 1.1.X
  - Exchange set validation checks also need to be prepared.
- The checks incorporate feedback from the NOAA S-111 team



IHO

# S-100 GENERIC VALIDATION CHECKS – UNDER DEVELOPMENT



S-100 Generic Validation Checks			Draft 0.3 20200720				
Pa	Check ID	S-100 Refere	Short Name	Detailed Description	Replaces S-58 Check	DQ Theme	Rationale/Requirement
4	S100_4a_001	4a	InvalidAgencyCode	If the producer agency code is not a value present in the XXX register.	1518a	Domain Consistency	If a defined listing of valid producer codes is established values must be from that list. Need to consider the point in time as the list may change.
4	S100_4a_002	4a	CatalogStructure	If the structure and content of the CATALOG.XML file is invalid.	1017-20b	Metadata Consistency	Schema validation and schematrons required to support this check. <b>Specific to exchange set transfer.</b>
4	S100_4a_003	4a	MissingSupportFile	Check that each support file included in the SupportFileDiscovery Metadata is present in the fileLocation specified.	1011	Completeness	
4	S100_4a_004	4a	DigitalSignatures	If the digital signature values are not present and valid in the DatasetDiscoveryMetadata or SupportFileDiscoveryMetadata.	1016	Format Consistency	Valid probably needs to be specified somehow.
4	S100_4a_005	4a	InvalidDatasetName	If the filename of the dataset does not conform to the DPS.	531	Format Consistency	
4	S100_4a_006	4a	InvalidSupportFilename	If the filename of any support file provided in the exchange set does not conform to the DPS.	1015	Format Consistency	
4	S100_4a_007	4a	MissingCatalogFile	If the CATALOG.XML file is not present.	1012	Metadata Consistency	<b>Specific to exchange set transfer.</b>
4	S100_4a_008	4a	CatalogDatasetInconsistent	If any of the same values listed in the CATALOG.XML and present in the ISO 8211 are not identical.	531	Metadata Consistency	Table of ISO 8211 fields and metadata elements is DPS specific e.g. Dataset Title (DSTL) must match Resource Title.
5	S100_5_001	5-A	ProhibitedObject	For each object which is present in the dataset but not present in the S-XXX Feature Catalogue.	504, 545	Domain Consistency	Datasets must only contain the objects listed in the FC (feature and information types). <b>Also need check for update consistency and how codes are defined in the dataset.</b>
5	S100_5_002	5-A	ProhibitedGeometry	For each object which is present in the dataset which references a geometry which is not permitted in the S-XXX Feature Catalogue. (includes no geometry)	12, 20a, 1797	Conceptual Consistency	Feature types must only have the permitted geometric primitives. <b>(noGeometry is a valid option for this check)</b>
5	S100_5_003	5-A	ProhibitedAttribute	For each attribute present in the dataset which is not present in the S-XXX FC.	511, 546, 547, 1567+various	Conceptual Consistency	Attributes present in the dataset must be present in the FC, includes simple and complex.
5	S100_5_004	5-A	InvalidAttributeBinding	For each attribute binding which is either not present in the S-XXX FC or does not conform to the multiplicity of the attribute binding.	8, 507, 547, 1679	Domain Consistency	Objects present in the dataset must only have attributes permitted in the Feature Catalogue. Also must only have the permitted multiplicity of attribute instances.

Image courtesy Thomas Richardson (S-101 PT Chair/S-100 generic validation check team lead).

The generic S-100 validation checks definition effort is focused on S-101.



Column	Description
Quality Measure (S-97 Part C)	Quality measure from S-97 Part C.
Check No.	Identifying number for check
Prerequisite check(s)	Checks which must be passed before this check can be executed.
Context test (IF ...) Initialization (SET ...)	<p>Combination of test conditions and initialization statements.</p> <ul style="list-style-type: none"> <li>• Check the existence of an HDF5 attribute or group</li> <li>• Test the value of a metadata attribute.</li> <li>• Initialization statements set the value of parameters used in the specific test in that row.</li> </ul>
Check condition	<p>Specification of check condition, written in structured English.</p> <p>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 fails.
Check solution	Solution to be applied to correct the failure.
Category/Severity	Whether check failure is a critical, error, or warning issue.
Post-condition	<p>Action to be executed if the check fails. For example:</p> <ul style="list-style-type: none"> <li>• Set a global flag to control check processing</li> <li>• Set a variable used in later checks (e.g., context variable to store the value of <i>dataCodingFormat</i>).</li> </ul>
S-100 Ed. 4.0.0 reference	S-100 reference for check
S-111 Ed. 1.0.1 reference	S-111 reference
Comments	Explanatory remarks or additional notes.

Red text denotes check information specific to the HDF5 format.



Quality Measure (see S-97 Part C)	Check No.	Prerequisite check(s)	Context test (IF ...) or initialization (SET ...)	Check condition description	Check message	Check solution	Category	Post-condition	S-100 Ed. 4.0.0 reference	S-111 Ed. 1.0.1 reference	Comments		
<b>Phase 2: Validate Feature Container Groups</b>													
	Initialization: TERMINATE=FALSE FTYPE=name of feature container group			The context for all checks in this phase is a feature container group named FTYPE. In S-111 the value of FTYPE=SurfaceCurrent, the name of the S-111 grid feature									
Logical Consistency / Format Consistency OR Conceptual Consistency	P2-29			IF attribute dimension is not found or not of the correct datatype or does not have the value specified in the product specification	Missing or invalid feature container attribute dimension	Add or correct attribute dimension	C	IF CHECK_SUCCESS SET context.DIM =dimension	Table 10c-10	Table 12.2	Not found: format consistency; Wrong datatype or value: conceptual consistency		
⋮													
Logical Consistency / Format Consistency OR Conceptual Consistency	P2-39			IF dataset axisNames is not found or does not have the correct rank and datatype, or has length ≠ DIM	Feature container group does not have the axisNames dataset, or the axisNames dataset does not have the correct dimensions	Add or correct the axisNames dataset	C		Table 10c-9	Clause 10.2.8; Table 10.2	Not found: format consistency; Wrong value: conceptual consistency		



Part	Check ID	S-100 Reference	Short Name	Detailed Description	Check Message	Severity	DQ Theme	Rationale/Requirement
4	S100_4a_001	4a	InvalidAgencyCode	If the producer agency code is not a value present in the IHO Data Producer register.	...	C	Domain Consistency	
4	S100_4a_002	4a	CatalogStructure	If the structure and content of the CATALOG.XML file is invalid.	...	C	Metadata Consistency	
4	S100_4a_004	4a	DigitalSignatures	If the digital signature values are not present and valid in the DatasetDiscoveryMetadata.	...	C	Format Consistency	
4	S100_4a_005	4a	InvalidDatasetName	If the filename of the dataset does not conform to the Product Specification.	...	C	Format Consistency	

The specific exchange set checks for S-111 are still to be prepared.

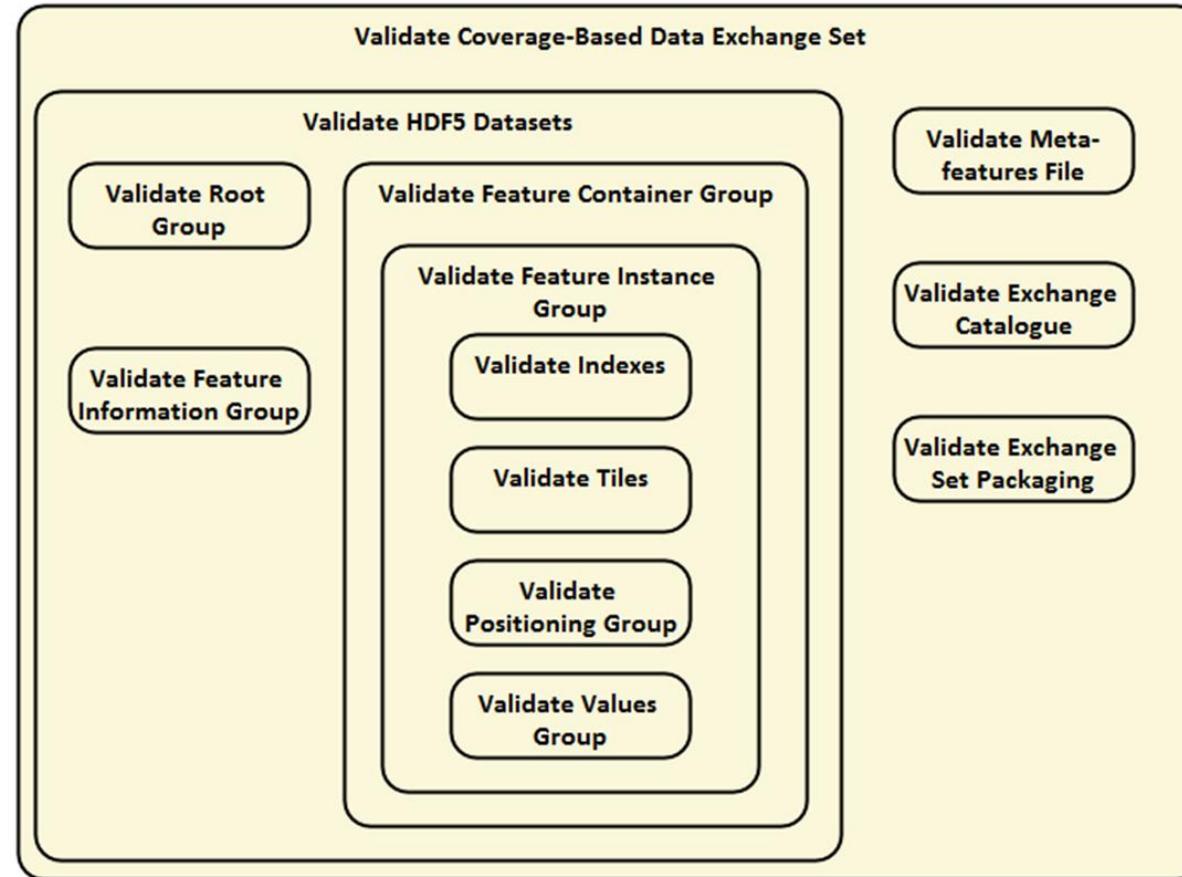


IHO

# GENERAL OVERVIEW OF S-100 HDF5 EXCHANGE SET VALIDATION

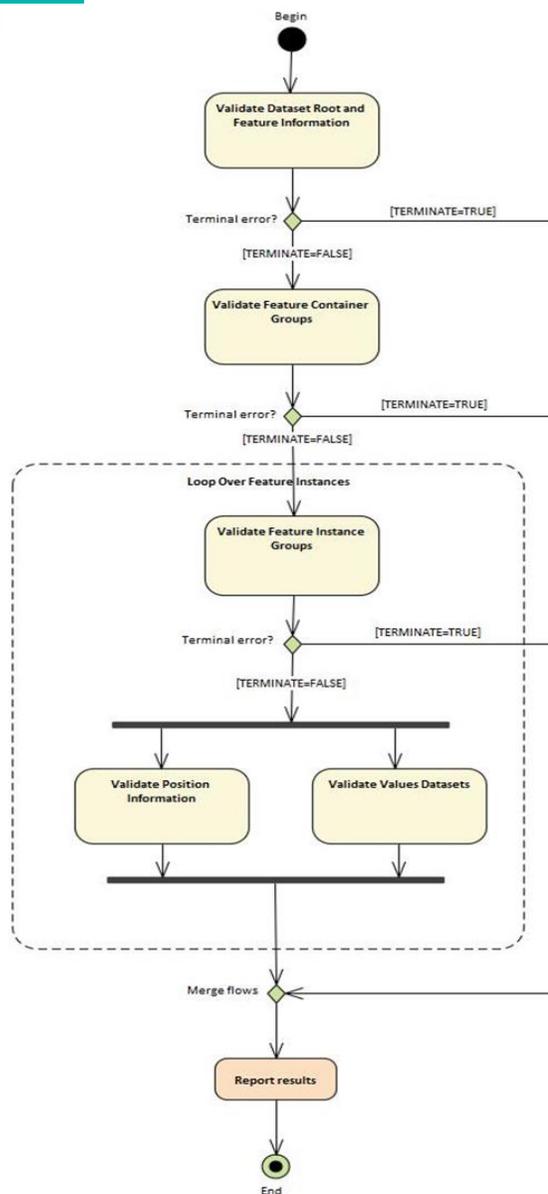


International  
Hydrographic  
Organization





# DATASET VALIDATION PROCESSING FLOW



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.

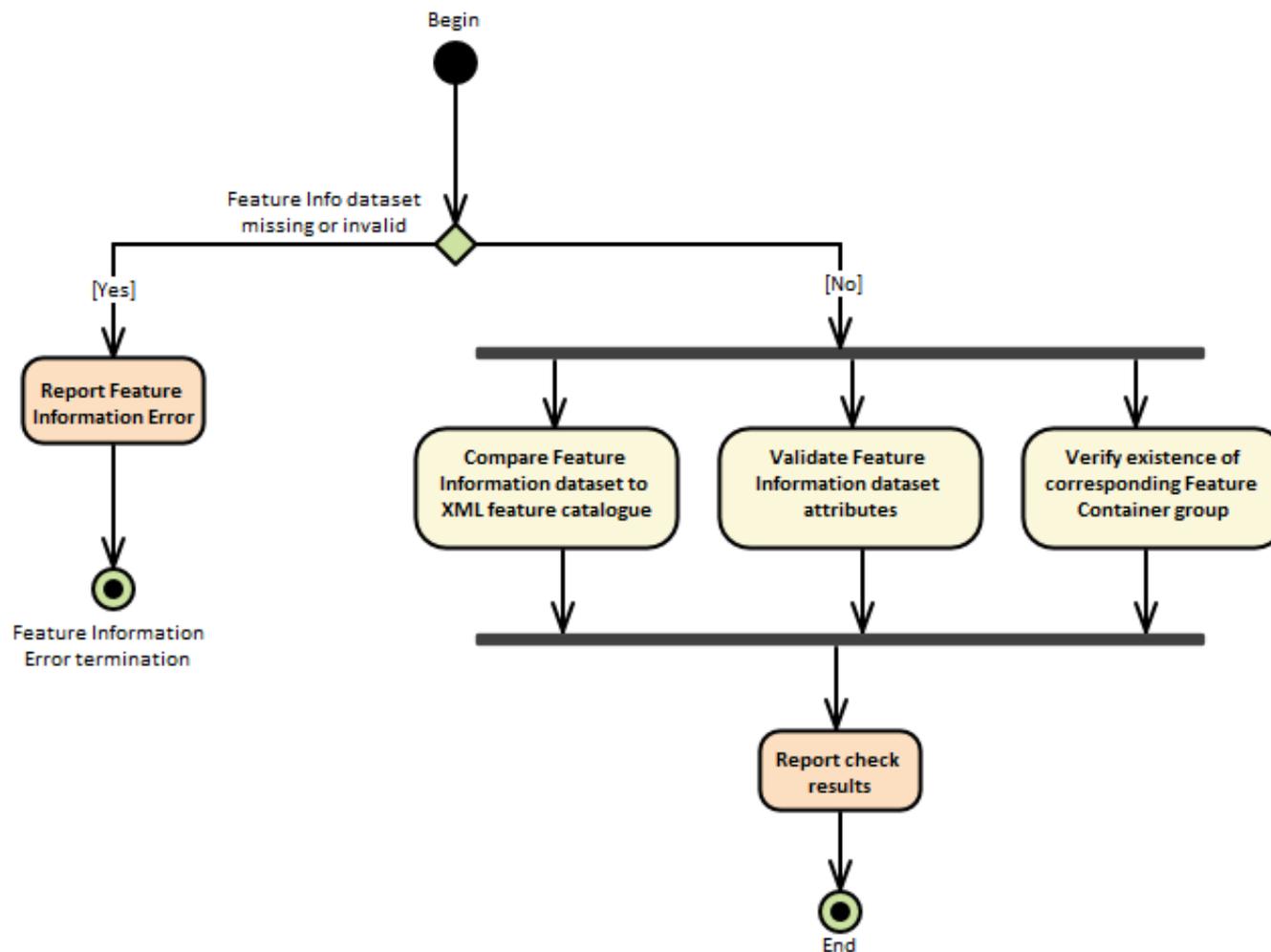


IHO

# VALIDATION FLOW FOR FEATURE INFORMATION GROUP



International  
Hydrographic  
Organization



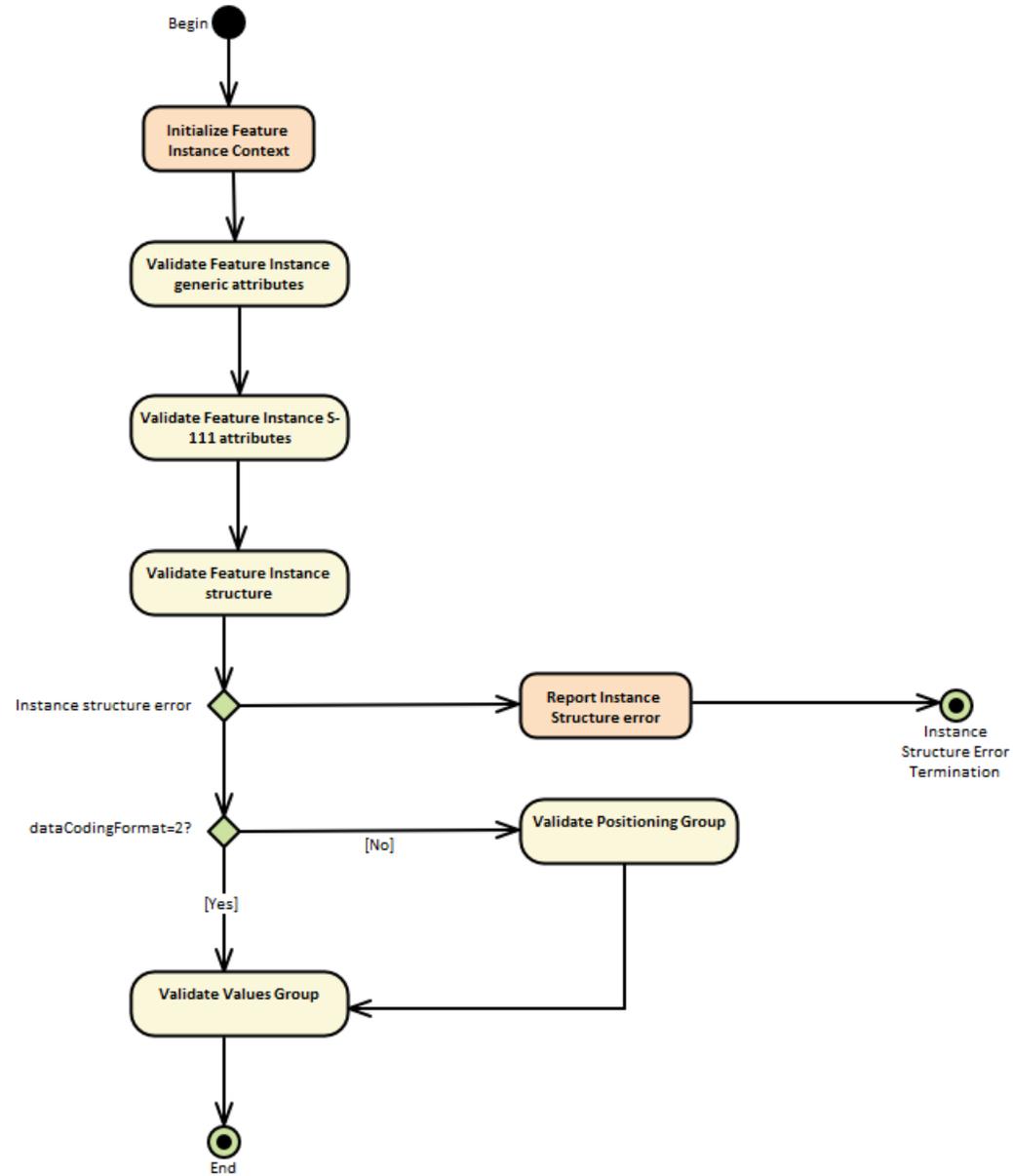


IHO

# VALIDATION FLOW FOR FEATURE INSTANCE GROUP



International  
Hydrographic  
Organization





IHO

# DOCUMENTATION OF VALIDATION CHECKS FOR THE PRODUCT SPECIFICATION



International  
Hydrographic  
Organization

- Derived from S-58 (S-57 ENC validation checks), adapted for the HDF5 data format.
  - Differences due to the fact that the HDF5 format is used for coverages rather than vector objects.
  - Differences due to the inherent hierarchical structure of the HDF5 format.
  - The applicability of certain checks depends on the success of other checks or on the values of metadata attributes.
  - Some checks require initialization of context
- The main components of the S-111 Validation Checks component of the product specification are:
  - A “core specification” for validation, describing:
    - Check classification(s): e.g., critical/error/warning.
    - Description of check syntax.
    - Operators (comparison, logical, and spatial).
    - Check processing flow (for HDF5 in particular, because the structure is more complex).