S-100 Validation Tests sub group

10th VTC Meeting

28th October 2024

MEETING PROTOCOL

Meeting participants are kindly requested to note the following meeting protocols;

- Please keep your camera and microphone turned **"off"** if you are not talking or presenting

- If you want to make an intervention, **please turn your camera and microphone on and, raise your hand** to indicate that you wish to speak

- Don't forget to turn your microphone **"on"** before speaking, and **"off"** when finished

- Please use the **"Chat"** function to communicate an text information to the meeting

- If you have any problems connecting using Firefox or other browser – please try using Chrome.

AGENDA

- Status of draft documents

- Issues identified while preparing S-158 drafts

- Preparation for S-100 WG9 meeting in November

- Open issues

- Next meeting

CURRENT STATUS

- Following documents are out for review:
    - S-158:101
    - S-158:102
    - S-158:104
    - S-158:111
    - S-158:124
    - S-158:129

- Comprises of cover text document and spreadsheet of checks
- Received feedback from S-101, S-102 and S-129
- Will discuss handover of S-158:1xx to PS teams at S-100 WG9
    - Once handed over are the responsibility of the PS team
    - Will require a change management procedure

- Contract has been awarded to Raphael Malyankar to produce the S-100 validation checks & cross product validation, cover documents
- Funds for contract supplied by IC-ENC

- Draft ready December 2024
- Final ready February 2025
- Submission to HSSC 17

- Long standing issue of how to list multiple clauses in S-158
- Circulated 2 options in September around PS owners, implementers and RENCs
  - Machine readable vs. human readable

Agreed on option 1, using box brackets to separate clause numbers for different documents

Standards document reference	Clause reference
PS; Annex B	[4.3.2.1.1]; [5.1.28, 6.1.2.7]
PS	4.3.2.1.1
PS; Annex A	[10.1.3]; [2.4.2]
PS; Annex A	[10.1.3]; [2.4.2]
Part 10a	10a-7.2.4.1, 10a-7.2.4.2.7, 10a-7.2.4.2.8

Note:
- documents must be separated by semicolon+space
- Box brackets in clause reference must be separated by semicolon+space
- Clause references for the same document must be separated by comma+space
- No use of "and" before the last clause reference (or document)
- No box brackets if there is only one document cited

- Being able to refer to a Clause or a Table
  - Many of the S-102 checks refer to tables not a text clause
  - Propose that this is acceptable

LANGUAGE CONSISTENCY

- Can be difficult to identify if checks are duplicated in S-100 and PS validation checks as have been written in a different style.

- Likely will need to be a revision of language in the future
  - Will raise at S-100 WG

VERSIONING

Change	PS	FC	PC	S-158:1xx cov.doc. (CD)	S-158:1xx list of val. checks (VC)	Notes
New edition of PS	X.0.0	X.0.0	X.0.0	=PS+yyyymmdd	=S-158:1xx CD	All components match in E+R+C
New revision of PS either w/ or w/o model change	N.X.0	N.X.0	N.X.0	=PS+yyyymmdd	=S-158:1xx CD	All components match in E+R+C
New PS clarification not affecting validation	N.N.X	(no change)	(no change)	N.N.N+yyyymmdd	N.N.N+yyyymmdd	All components match in E+R; CD=VC
FC/PC correction w/o change to PS or validation	N.N.N	Same E.R, some type of change elsewhere	Same E.R, some type of change elsewhere	N.N.N+yyyymmdd	N.N.N+yyyymmdd	All components match in E+R; CD=VC
PS adds product-specific constraint on attribute values w/o FC/PC change	N.N.X	(no change)	(no change)	N.N.N+yyyymmdd	N.N.N+yyyymmdd	All components match in E+R; CD and VC match in E+R+C
Correct validation check w/o change to S-158:1xx cover doc., PS, FC, or PC. E.g., correct C3DI to C3IT	N.N.N	(no change)	(no change)	N.N.N+yyyymmdd	N.N.N+yyyymmdd	All components match in E+R; CD and VC match in E+R+C
Add, delete, or amend validation check w/o change to PS, FC, or PC	N.N.N	(no change)	(no change)	N.N.N+yyyymmdd	N.N.N+yyyymmdd	All components match in E+R; CD and VC match in E+R+C
Correction to S-158:1xx cover doc. only	N.N.N	(no change)	(no change)	N.N.N+yyyymmdd	N.N.N+yyyymmdd	All components match in E+R; CD and VC match in E+R+C

E= Edition, R= Revision, C=Clarification   CD= cover document, VC= validation checks

VERSIONING CONT.

Change	PS	FC	PC	S-158:1xx cov.doc. (CD)	S-158:1xx list of val. checks (VC)	Notes
Add, remove, or amend S-158:100 check w/o revision to S-100 or S-1xx PS	N.N.N	(no change)	(no change)	No change or N.N.X+yyyymmdd (depends on nature of change)	No change or N.N.X+yyyymmdd (align to CD)	All components match in E+R; CD and VC match in E+R+C
New edition of S-100	X.0.0 or N.X.0	X.0.0 or N.X.0	X.0.0 or N.X.0	=PS+yyyymmdd	=S-158:1xx CD	All components match in E+R+C
Revise structure of checks in S-158	N.N.N	(no change)	(no change)	N.N.N+yyyymmdd	N.N.N+yyyymmdd	All components match in E+R+C
Add product-specific field to list of checks	N.N.N	(no change)	(no change)	N.N.X+yyyymmdd	N.N.X+yyyymmdd (align to CD)	All components match in E+R; CD=VC

- The S-158:1xx cover document and list of checks always have the same edition+revision as the PS, FC, and PC.
- The S-158:1xx cover document and list of checks always have the same edition+revision+clarification.
- An editorial change to the S-158:1xx cover document does not require a new list of checks, and vice versa.

Rule for selection of checks
Implementations use the list of checks which has the same Ed+Rev as the Ed+Rev part of the PS version specified in the dataset. Ties are broken by picking the list of checks which has the latest build date.

General notes
New editions or revisions of S-100 will always require an update to S-158:100 at least to update the reference to S-10, even if there is no change to validation requirements. (The S-100 edition is mentioned on the title page of S-158:100, e.g., "Aligned to S-100 Edition 5.2.0".)

CHECK CLASSIFICATION

International Hydrographic Organization

- Use of current S-58 Classifications – Critical Error, Error & Warning – ENC and ECDIS specific.
    - Will need rewording for S-100

1.2 Check Classification

The check classification is intended to ensure that published ENC data is free of errors which would affect the use of an ENC in ECDIS. In some cases it has been necessary to diverge from the strength of wording used in the S-57 ENC Product Specification or the Use of the Object Catalogue for ENC. In such cases the impact on the user has been the overriding factor for consideration. The classifications have the following meanings:

C	Critical Error	An error which would make an ENC 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 ENC 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 an ENC in ECDIS.

At a minimum validation software must group validation reports using these categories. They may also support sub-grouping of related checks such as those relating to geometric validity or attribute consistency. Software may allow checks of type Error or Warning to be deselected completely or by such categories.

- GitHub Issue open on whether to change Critical Error to just Critical to avoid confusion #52

IHO

International Hydrographic Organization

- Linking which S-100 validation checks or Parts are relevant to different Product Specifications?

5 Other Applicable Checks

5.1 Generic S-100 checks

S-101 datasets and exchange sets must also be validated using the following subset of the generic S-100 validation checks defined in S-158:100:

Table 5.1 - Applicablity of generic S-100 checks

Document reference in S-158:100 list	Checks	Apply to	Remarks
Part 1	All	Product Specification	No direct implementation on datasets or exchange sets
Part 2 / 2a	All	Product Specification	No direct implementation on datasets or exchange sets
Part 4a	All	Exchange catalogue	
Part 4b	All	Product Specification	No direct implementation on datasets or exchange sets
Part 5 / 5a	S100_Dev0069	Product Specification	No direct implementation on datasets or exchange sets
	S100_Dev0077 S100_Dev0468 S100_Dev0161 S100_Dev0162 S100_Dev0163 S100_Dev0164 S100_Dev0165 S100_Dev0166 S100_Dev0167 S100_Dev0168 S100_Dev0169 S100_Dev0170 S100_Dev0171	Datasets	
Part 6	S100_Dev0172	Datasets	

	S100_Dev0171		
Part 6	S100_Dev0172 S100_Dev0173 S100_Dev0174	Datasets	
Part 7	All checks except those for arc, circle, and spline primitives or Level 3b geometry	Datasets	S-101 uses Level 3a geometry. S-101 does not use arc, spline, and circle spatial primitives
Part 8	None	N/A	Part 8 does not apply to S-101
Part 9 / 9a / 13	?	Product Specification	Validation checks for Portrayal Catalogue
Part 10a	All checks except those applicable to arc, circle, and spline primitives	Dataset	S-101 does not use arc, spline, and circle spatial primitives
Part 10b / 10c	None	N/A	Parts 10b and 10c do not apply to S-101
Part 11	S100_Dev0466	Dataset	There is only one Part 11 generic check, for dataset size
Part 15	?	?	
Part 17	All checks	Exchange catalogue	

DIFFERENCE BETWEEN DATASET CHECKS AND STANDARDS CHECKS

The S-100 validation checks contain a mixture of checks for two very different use cases.

- Checks for products, which are used by data producers to validate datasets .
- Checks to validate standards. These are used by standards organizations, such as the IHO to validate that their standards, i.e. product specifications, feature catalogues, etc. are compliant with S-100. An example is check S100_Dev0009.

S100_Dev0009		Warning	Feature Catalogue not in registry	The Feature Catalogue conforms to the Registry	Add feature catalogue to entry for product specification in IHO GI Registry	Part 2

These two types of checks have very different user groups.

- It is not the responsibility of the data producer to validate components of the standard, e.g., feature catalogues. Once signed, it is assumed that the feature catalogue has been validated and is correct.
- Validation software designed to validate datasets should not be required to validate components of the standard. There will be separate software applications for points 1 and 2, since the user groups are very different. As such, it would be a very good idea to somehow mark validation checks related to the verification of a standard.

Proposal: Raise as topic at S-100 WG with potential options

e.g. different document, separate sections in S-158:100 or tag to identify.

I would like to provide some recommendations and comments as follow:

1. it is recommended to swap the positions of column "Data Quality Measure" and column "Classification"。

2. There is an ordering in data quality evaluation which is recommended to be followed in the template. When evaluating data quality, the usual ordering is:

a)   Logical consistency/Format consistency: The very first to be evaluated is the readability (or interpretability) of the data to decide whether it is possible to decode/read/understand the data or not. Not interpretable data should be reported and ignored in the further evaluation. The result of the format consistency should describe which parts of the data are not readable.

b)   Other Logical consistency (concept consistency, domain consistency and topological consistency): Decide if the rules set up for the data set are followed. Parts of the data set not conforming to the rules should be ignored in the further evaluation.

c)   Completeness (Commission and omission): The next step in the evaluation is the feature (not attribute or item) existence aspect covered by completeness. To evaluate this, the features in the actual data set and the ground truth data are compared, and commissions and omissions reported.

d)   Accuracy (positional, thematic and temporal aspects): The last step in the evaluation covers the accuracy aspect, measuring the deviation between actual and ground truth feature properties. These measurements can be based only on parts of the data set present in both the actual data set and the universe of discourse.

3. A draft guidelines on the data quality evaluation of S-100 products has been developed by DQWG and presented in HSSC16. Noting the link between S-158 and the Data quality evaluation of S-100 products, HSSC invited the DQWG and the S-100WG to discuss the possible **development of a guidance on data quality evaluation of S-100 products** (Action HSSC16/72). I will try to share the draft guidelines in S-100WG9.

(paper submitted to S-100 WG 9-08.02)

Dev ID	Check ID	Classification	Check Message	Check Description	Check Solution	Standards document reference	Clause reference	Data Quality Measure	Introduced	Modified	Deleted

Separately it has been raised if we retain DQ Measure?

- The Excel format is working for now, but we all recongise this will not work longer term

- Suggestions for a new Registry, a database, Metanorma, xml format

- Different use cases

- The documents cross multiple working groups, will there be an agreed format for all or can it differ

- Reuse of validation checks for other product specifications
  - Inland ENCs and AMLs may want to replicate some of the S-101 checks

- Will raise at S-100 WG, proposing to recommend to HSSC that ICE-PT assist

- Restrictions may be needed for WLA
  - Likely these will need to be handled through validation checks
  - Refinement will likely be needed
  - e.g. alignment of display scales, datums across products

- Will be discussed at S-100 WG either in S-98 or Validation section, possibly both
- Need to establish how we will coordinate across working groups & PTs

- Status of production of validation checks incl. cross product validation
  - WLA checks
- Template and feedback from DQWG & do we retain DQ Measure
- Issues encountered e.g. citations, versioning, language
- Check classification
- Potential need for splitting datasets and standards validation checks
- Handover and change process (Managed Impact Study)
- Longer term format of validation checks
- Issue with Part 6 checks – is anyone using this Part?

CHANGE PROPOSAL SUBMITTED

- Remove Part 7 statement that all geometric primitives must be part of at least one GM_Aggregate

Change Proposal

Semantics section contains the text "All geometric primitives (GM_Primitive) must be part of at least one GM_Aggregate (see ISO 19107 clause 8.10.1)."

This statement is not supported by the UML diagram 7-3 and it is also not possible to find a relevant statement in ISO.
ISO 19107:2003 describes GM_Primitive in clause 6.3.10 but that does not contain the statement from S-100 7-4.2.11.1
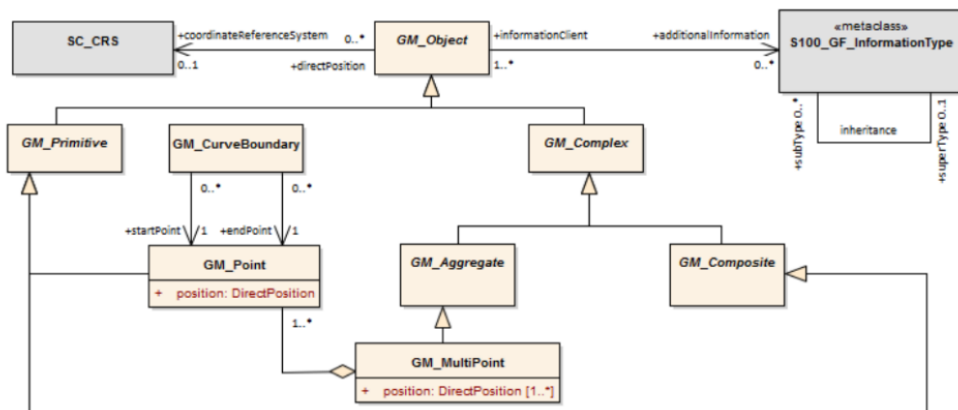
Part of diagram 7-3

This was identified by the S-100 Validation sub group and can be tracked in the GitHub issue
Part 7 check S100_Dev0088 · Issue #22 · iho-ohi/S-100-Validation-Checks (github.com)

Change Proposal Justification

Propose removing the statement from Part 7 as to provide consistency with the UML diagram as no supporting ISO source can be found for the proposal.

- S100 WG 04-11

https://github.com/iho-ohi/S-100-Validation-Checks/issues

30 open, 16 issues related to the 10c checks (been open since February)

10c checks:
- Where there is no disagreement I will give two weeks and then I am closing them accepting initial comments from Raphael
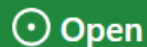- Six with disagreement which we can decide on now.

Many of the other issues can be resolved

S158:100 0.1.0 Dev0271 10c Check #41

⊙ Open   LizHahessy opened this issue on Feb 26 · 0 comments

LizHahessy commented on Feb 26   (Collaborator) •••

"Check that all numeric values represented as strings within the Feature Information Group are valid numbers or empty (0-length) strings."

Comment (RM) - (NB: Might have to modify the 0-length requirement depending on implementer experience.)

Comment (FH) - If an amendment needs to be made, it should be to the 10c-9.5 clause first.

☺

10c – 9.5

All the numeric values in the feature description dataset are string representations of numeric values; for example, "-9999.0" not the float value -9999.0. Applications are expected to parse the strings to obtain the numeric value. Inapplicable entries are represented by null values or the empty (0-length) string.

Propose leave as is for now

ISSUE 35

S158:100 0.1.0 Dev0253 10c Check #35

⊙ Open **LizHahessy** opened this issue on Feb 26 · 0 comments

LizHahessy commented on Feb 26   [Collaborator] •••

"Values of HDF5 datatypes must conform to the constraints defined by table 10c-1."

Comment (RM) - Applies to datasets only indirectly. Potentially superseded by lower-level and product-specific checks. Consider deletion.

Comment (FH) - I can't think of a use case where it would even make sense for a product specification to need to alter these constraints. Perhaps a product specification could add further constraints, but to allow a product specification to relax S-100 level constraints is not a good idea in my opinion.

☺

Propose leave as is for now

Table 10c-1 – Equivalences between S-100 and HDF5 datatypes

S-100 Attribute Value Types	HDF5 Datatype Class	Constraint on HDF5 datatype
real	Float	32 or 64-bit floating point
integer	Integer	1, 2, or 4-byte signed and unsigned integers
text (CharacterString in S-100 metadata)	String	variable-length string
enumeration	Enumeration	Numeric codes must be 1 or 2-byte unsigned integers, range $[1, 2^8 – 1]$ or $[1, 2^{16} - 1]$
date	(Character) String, length=8	Date format according to Table 1-2 (Part 1); that is, complete representation, basic format, as specified by ISO 8601
time	(Character) Variable-length string	Time format according to Table 1-2 (Part 1); that is, complete representation, basic format as specified by ISO 8601. UTC indicated by "Z" suffix; local time by absence of suffix. The zone offset format is also permitted); for example, 123000+0100
dateTime	(Character) (variable length string)	Date-time format as specified by ISO 8601. EXAMPLES: 19850412T101530Z 19850412T101530-0500
boolean	(Integer)	1-byte unsigned, Values: 1 (TRUE); 0 (FALSE)
S100_Codelist	Compound (Enumeration, variable-length string)	Exactly one of the components is allowed; the other must be the numeric value 0 or the empty (0-length) string according to its data type
URI, URL, URN	String (variable-length)	Format specified in RFC 3986 (URI, URL) or RFC 2141 (URN)
S100_TruncatedDate	String, length=8	Format as in Part 1 Table 1-2
value record (Part 8)	Compound	Datatypes of components must be according to value attribute types in the Application Schema. The "value record" corresponds to the value(s) record in Part 8 Figures 8-22, 8-23, 8-24 and 8-25
external object reference	String	Format: extObjRef:<fileName>:<recordIdentifier> where <fileName> is the base name of the ISO 8211 or GML file, and <recordIdentifier> is the record identifier of the vector object record within that file. The extension part of the file name is not used. The record identifier is the gml:id for GML datasets, or the record identification number (RCID) for ISO 8211 datasets. The file must be present in the same exchange set

S158:100 0.1.0 Dev0252 10c Check #34
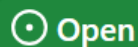
⊙ Open LizHahessy opened this issue on Feb 26 · 0 comments

LizHahessy commented on Feb 26   Collaborator ···

"HDF5 data classes not listed in table 10c-1 shall not be used."

Comment (RM) - Applies to datasets only indirectly. Potentially superseded by lower-level and product-specific checks. Consider deletion.

Comment (FH) - This is almost a direct quote from 10c-7. If the check is to be deleted, then the clause in S-100 should be deleted or may require clarification.

☺

10c-7 Data types

Predefined HDF5 data types include Integer, Float, String, and Enumeration, but there are no HDF5 equivalents to the S-100 data types Boolean, S100_Codelist or S100_TruncatedDate. The latter types are mapped to the HDF5 constructs specified in the Table below. The S-100 data types Date, DateTime, and Time are mapped to HDF5 strings due to potential problems with portability across different processor architectures of HDF5 Time formats. In S-100 HDF5 data products, S-100 data types defined in Part 3 are mapped to equivalent HDF5 data types. These equivalences are summarized in Table 10c-1 below. HDF5 datatype classes not mentioned in this Table shall not be used.

Propose leave as is for now

S158:100 0.1.0 Dev0251 10c Check #32

⊙ Open **LizHahessy** opened this issue on Feb 26 · 0 comments

LizHahessy commented on Feb 26

"Each named object has at least one link to it."

Comment (RM) - Delete, descriptive rather than prescriptive and inherent in HDF5

Comment (FH) - Is it possible to have a named object that doesn't have a link to it?

☺

10c-5.1.2 Group

An HDF5 group is analogous to a file system directory. Abstractly, a group contains zero or more objects, and every object must be a member of at least one group. The root group is a special case; it may not be a member of any group.

Group membership is actually implemented via link objects. See the Figure below. A link object is owned by a group and points to a named object. Each link has a name, and each link points to exactly one object. ==Each named object has at least one and possibly many links to it.==
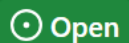
Propose leave as is for now

S158:100 0.1.0 Dev0249 10C check #31

⊙ Open   **LizHahessy** opened this issue on Feb 26 · 0 comments

LizHahessy commented on Feb 26 · edited ▾

"Within a single HDF5 file, all HDF5 objects are uniquely defined."

Comment (RM) - Delete, descriptive rather than prescriptive and inherent in HDF5

Comment (FH) - I think to check that object identifiers are unique within a HDF5 is valid.

10c-5.1.1 File

Abstractly, an HDF5 file is a container for an organized collection of objects. The objects are groups, datasets, and other objects as defined below. The objects are organized as a rooted, directed graph. Every HDF5 file has at least one object, the root group. See the figure below. All objects are members of the root group or descendents of the root group.

HDF5 objects have a unique identity *within a single HDF5 file* and can be accessed only by its names within the hierarchy of the file. HDF5 objects in different files do not necessarily have unique identities, and it is not possible to access a permanent HDF5 object except through a file.

When the file is created, the *file creation properties* specify settings for the file. The file creation properties include version information and parameters of global data structures. When the file is opened, the *file access properties* specify settings for the current access to the file. File access properties include parameters for storage drivers and parameters for caching and garbage collection. The file creation properties are set permanently for the life of the file, and the file access properties can be changed by closing and reopening the file.

An HDF5 file can be "mounted" as part of another HDF5 file. This is analogous to Unix file system mounts. The root of the mounted file is attached to a group in the mounting file, and all the contents can be accessed as if the mounted file were part of the mounting file.

Propose ?

IHO — International Hydrographic Organization

S158:100 0.1.0 Dev0247 HDF5 data 10C check #29
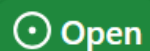
⊙ **Open**   **LizHahessy** opened this issue on Feb 26 · 0 comments

LizHahessy commented on Feb 26 · edited ▾

"HDF5 constructs requiring the use of an HDF5 library later than version 1.8.8 must not be used."

Comment (FH) - HDF5 1.8.8 is quite old now.
Comment (RM) - that implementations may use later HDF5 libraries compatible with version 1.8.8.
Comment (FH) - If that is the case, then should the S-100 clause be removed?

10c-3 Conformance

The S-100 HDF5 data format conforms to release 1.8.8 of HDF5.

10c-5.3 Prohibited HDF5 constructs

Constructs which cannot be processed using the standard libraries of the HDF5 release specified in this Part must not be used. This means specifically that HDF5 constructs which require the use of a library for a later release than that specified in this Part must not be used.

Propose leave as is for now

NEXT STEPS

- Continue producing a set of tests from Parts of S-100 and cross product checks

- Coordinate Cross-Product Validation with S-98/S-164 lead, DQWG and S-1xx PS owners

- Prepare papers and validation session for S-100 WG9

- Start resolving open issues on GitHub

S100WG9 and PROJECT TEAMS TIMETABLE

- Version 1.0 -

	November 4	November 5	November 6		November 7		November 8	
Morning	S100WG Plenary	S100WG Plenary – ECDIS Implementation Topics (S-98/S-164)	S100WG Plenary – Data Producer Implementation (Validation)		S100WG Plenary		S-102 PT Meeting – Breakout	ICE-PT – Breakout
Meeting Room	Main Room	Main Room	Main Room		Main Room		Room 1	Room 2
Afternoon	S100WG Plenary	S100WG Plenary – ECDIS Implementation Topics (S-98/S-164)	S100WG Plenary – Data Producer Implementation (Conversion Workshop)	S-129 PT Meeting – Breakout room	S-102 PT Meeting	S-129 PT Meeting – Breakout room	S-102 PT Meeting – Breakout	ICE-PT - Breakout
Meeting Room	Main Room	Main Room	Main Room	Room 1	Main Room	Room 1	Room 1	Room 2

S-100WG9, Genoa, Italy – 04/11/24 – 08/11/24

Suggest next VTC meeting is January

Any Questions?