S-100 TSM7 -5.6

Paper for Consideration by S-100 WG

Inconsistencies and other problems related to the HDF5 encapsulation

Submitted by:	SevenCs GmbH (Holger Bothien)
Executive Summary:	This document lists several problems that arises when trying to implement the HDF5 encapsulation of S-100
Related Documents:	S-100 4.0.0 Part 10c S-102 2.0.0 HDF5 documentation of the HDF5 Group
Related Projects:	S-102 Project team

Introduction / Background

S-100 Part 10c contains the details on the encapsulation of S-100 datasets using the HDF5 format. In S-102 a first implementation of this standard was done. It turns out that in S-100 Part 10c are some problems and inconsistencies. Furthermore, several more inconsistencies exist between the S-102 implementation and S-100 Part 10c. This altogether makes it difficult to implement the standard. It will definitely end up in different interpretation of the standard and to systems and data which will be not compatible to each other.

Analysis/Discussion

Problems / Questions in S-100 Part 10c

- chunking: Table 10c-12
  This is defined in HDF5 as a member of the property list, encoding in addition as an attribute will be redundant and leads to inconsistencies. (one inconsistency is already in the S102 document. Proposal: remove attribute.
- Attributes in feature instance group Table 10c-12
  origin and spacing is fixed to longitude, latitude with unit degrees. It should be based on the definition of the CRS. The definition of the CRS contains a definition of a coordinate system with axis that have names and units. Origin and spacing of a grid should use this units. Btw. where is the definition of the CRS in the encoding? I couldn't find it.
- Start sequence:
  What is that? It seems that the information about the sequence of the data is completely described by the sequencing rule. Proposal: remove it or describe the use case where it has some value.
- Table 10c-6 contains horizontal datum information but should contain horizontal CRS definition (which will contain a datum), Btw the code 4326 is the code of a geographic CRS
- The attribute vertical position uncertainty at Feature container groups should be optional because many data will not contain vertical coordinates.

Problems / Inconsistencies / Questions S-102

- Bounding Box and Sequence rule inconsistency (S102 B-1)
  both information are in S-10 a compound dataset under the root group of the file
  In S-100 Part 10c the bounding box is defined by attributes in the root group and sequence rule is defined by attributes of the feature container group
  The S-100 implementation of sequence rule makes more sense since it applies only to some type of data (grids). In S-102 Figure B-15 the correct attributes are described.

- Dimension of featureCode dataset seems to be wrong. (S-102 Figure B.3)
  Dimension is 2x2 but should be 2x1
- Dimension of axisName dataset seems to be wrong
  Is 2x2 but should be 2x1
- in S-102 Figure B-19 some attributes are shown that only apply to 3D grids, which we don't have in S-102 (minGridPointVertical, numPointsVertical)
- Attributes of the root group are not defined at all (see S-100 Part 10c Table 10c-6
- Naming of FeatureInstanceGroup is (feature code)_N but should be (feature code).N according to S-100 Part 10c (Table 10c-10)
- Figure B.5 shows the structure of the FeatureDescriptionDataset ***BathymetricCoverage*** fillValue, lower, and upper are of type float but should be string according to S-100 Table 10c-8. For this feature it might be ok but feature with non-numeric attributes will running into problems. See Figure B.8 and B.9 in S-102.
- The Figure B.9 has inconsistencies with 4.2.1.1.9 and with the Feature Catalogue. In addition, there is an attribute original value but in fact there are two attributes (depth and uncertainty) that can be overwritten. Also unclear is how the reference to the BathymetricCoverage instance works, with other words for which feature instance a TrackingListCoverage feature has tracked values.

General comment on structure documentation

- The use of screen shots for documentation is not appropriate.
- It does not show all information (General information, attributes, content)
- Different version of the tool produces different images
- A better while standardized method is the DDL. It is defined by the HDF5 group, ASCII format and can be easily created by the ***h5dump*** tool

Here is an example, note that it neither a complete S-102 file nor It might be free of errors. Nevertheless, I believe that with this format we can describe much more exact the structure of an HDF5 file.

```
HDF5 "s102test.h5" {
GROUP "/" {
   GROUP "BathymetryCoverage" {
      ATTRIBUTE "commonPointRule" {
         DATATYPE  H5T_STD_I8LE
         DATASPACE  SIMPLE { ( 1 ) / ( 1 ) }
         DATA {
         (0): 1
         }
      }
      ATTRIBUTE "dataCodingFormat" {
         DATATYPE  H5T_STD_I8LE
         DATASPACE  SIMPLE { ( 1 ) / ( 1 ) }
         DATA {
         (0): 2
         }
      }
      ATTRIBUTE "dimension" {
         DATATYPE  H5T_STD_I8LE
         DATASPACE  SIMPLE { ( 1 ) / ( 1 ) }
         DATA {
         (0): 2
         }
      }
      ATTRIBUTE "horizontalPositionUncertainty" {
         DATATYPE  H5T_IEEE_F32LE
```

```
      DATASPACE  SIMPLE { ( 1 ) / ( 1 ) }
      DATA {
      (0): -1
      }
   }
   ATTRIBUTE "interpolationType" {
      DATATYPE  H5T_STD_I8LE
      DATASPACE  SIMPLE { ( 1 ) / ( 1 ) }
      DATA {
      (0): 5
      }
   }
   ATTRIBUTE "numInstances" {
      DATATYPE  H5T_STD_I32LE
      DATASPACE  SIMPLE { ( 1 ) / ( 1 ) }
      DATA {
      (0): 2
      }
   }
   ATTRIBUTE "sequencingRule.scanDirection" {
      DATATYPE  H5T_STRING {
         STRSIZE 1024;
         STRPAD H5T_STR_NULLTERM;
         CSET H5T_CSET_ASCII;
         CTYPE H5T_C_S1;
      }
      DATASPACE  SIMPLE { ( 1 ) / ( 1 ) }
      DATA {
      (0): "-latitude,longitude"
      }
   }
   ATTRIBUTE "sequencingRule.type" {
      DATATYPE  H5T_STD_I8LE
      DATASPACE  SIMPLE { ( 1 ) / ( 1 ) }
      DATA {
      (0): 1
      }
   }
   ATTRIBUTE "verticalUncertainty" {
      DATATYPE  H5T_IEEE_F32LE
      DATASPACE  SIMPLE { ( 1 ) / ( 1 ) }
      DATA {
      (0): -1
      }
   }
}
GROUP "Group_F" {
   DATASET "BathymetryCoverage" {
      DATATYPE  H5T_COMPOUND {
         H5T_STRING {
            STRSIZE 1024;
            STRPAD H5T_STR_NULLTERM;
            CSET H5T_CSET_ASCII;
            CTYPE H5T_C_S1;
         } "code";
         H5T_STRING {
            STRSIZE 1024;
            STRPAD H5T_STR_NULLTERM;
            CSET H5T_CSET_ASCII;
            CTYPE H5T_C_S1;
         } "name";
         H5T_STRING {
            STRSIZE 1024;
            STRPAD H5T_STR_NULLTERM;
            CSET H5T_CSET_ASCII;
            CTYPE H5T_C_S1;
         } "uom.name";
         H5T_IEEE_F32LE "fillValue";
         H5T_STRING {
            STRSIZE 1024;
            STRPAD H5T_STR_NULLTERM;
            CSET H5T_CSET_ASCII;
            CTYPE H5T_C_S1;
         } "dataType";
```

```
      H5T_IEEE_F32LE "lower";
      H5T_IEEE_F32LE "upper";
      H5T_STRING {
         STRSIZE 1024;
         STRPAD H5T_STR_NULLTERM;
         CSET H5T_CSET_ASCII;
         CTYPE H5T_C_S1;
      } "closure";
   }
   DATASPACE  SIMPLE { ( 2, 1 ) / ( 2, 1 ) }
   DATA {
   (0,0): {
         "depth",
         "depth",
         "metres",
         1e+06,
         "H5T_NATIVE_FLOAT",
         -12000,
         12000,
         "closedInterval"
      },
   (1,0): {
         "uncertainty",
         "uncertainty",
         "metres",
         1e+06,
         "H5T_NATIVE_FLOAT",
         -12000,
         12000,
         "closedInterval"
      }
   }
}
DATASET "TrackingListCoverage" {
   DATATYPE  H5T_COMPOUND {
      H5T_STRING {
         STRSIZE 1024;
         STRPAD H5T_STR_NULLTERM;
         CSET H5T_CSET_ASCII;
         CTYPE H5T_C_S1;
      } "code";
      H5T_STRING {
         STRSIZE 1024;
         STRPAD H5T_STR_NULLTERM;
         CSET H5T_CSET_ASCII;
         CTYPE H5T_C_S1;
      } "name";
      H5T_STRING {
         STRSIZE 1024;
         STRPAD H5T_STR_NULLTERM;
         CSET H5T_CSET_ASCII;
         CTYPE H5T_C_S1;
      } "uom.name";
      H5T_IEEE_F32LE "fillValue";
      H5T_STRING {
         STRSIZE 1024;
         STRPAD H5T_STR_NULLTERM;
         CSET H5T_CSET_ASCII;
         CTYPE H5T_C_S1;
      } "dataType";
      H5T_IEEE_F32LE "lower";
      H5T_IEEE_F32LE "upper";
      H5T_STRING {
         STRSIZE 1024;
         STRPAD H5T_STR_NULLTERM;
         CSET H5T_CSET_ASCII;
         CTYPE H5T_C_S1;
      } "closure";
   }
   DATASPACE  SIMPLE { ( 5, 1 ) / ( 5, 1 ) }
   DATA {
   (0,0): {
         "X",
         "X",
```

```
            "N/A",
            0,
            "H5T_NATIVE_INT",
            0,
            0,
            "closedInterval"
         },
      (1,0): {
            "Y",
            "Y",
            "N/A",
            0,
            "H5T_NATIVE_INT",
            0,
            0,
            "closedInterval"
         },
      (2,0): {
            "originalValue",
            "Original Value",
            "metres",
            0,
            "H5T_NATIVE_FLOAT",
            -12000,
            12000,
            "closedInterval"
         },
      (3,0): {
            "trackCode",
            "Track Code",
            "N/A",
            0,
            "H5T_NATIVE_INT",
            0,
            0,
            "closedInterval"
         },
      (4,0): {
            "listSeries",
            "List Series",
            "N/A",
            0,
            "H5T_NATIVE_INT",
            0,
            0,
            "closedInterval"
         }
      }
   }
   DATASET "featureCode" {
      DATATYPE  H5T_STRING {
         STRSIZE 1024;
         STRPAD H5T_STR_NULLTERM;
         CSET H5T_CSET_ASCII;
         CTYPE H5T_C_S1;
      }
      DATASPACE  SIMPLE { ( 2, 1 ) / ( 2, 1 ) }
      DATA {
      (0,0): "BathymetryCoverage",
      (1,0): "TrackingListCoverage"
      }
   }
}
}
}
```

Conclusions

Although the majority of S-100 Part 10c is in a good shape some details leads to difficulties in the implementations.

S-102 uses this standard in a general understandable way but many details are not consistent with S-100 Part 10c and will lead to implementation issues and a variety of data. This is something standards should avoid and not promote.

Recommendations

- Correct and/or clarify S-100 Part 10c to remove inconsistencies and limitations.
- Correct S-102 in order to bring it to full compliance with the main standard.
- In our opinion a small expert group should propose the necessary changes based on our observation and on reviews from the other team members.
- A more general discussion should be started on how the quality of standards could be improved before publishing them. Possible mechanism could be:
    - Test implementations
    - Expert reviews (active reviews not silent acceptance)

Action Required of [S-100 WG/S-102 PT]

The TSM meeting is invited to:

a. endorse the paper
b. discuss the topics
c. conclude on the action list
d. decide on the next steps forward