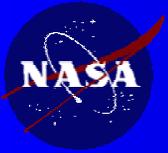


# An XML Approach to HDF-EOS5 Files

Jingli Yang<sup>1</sup>, Bob Bane<sup>1</sup>, Muhammad Rabi<sup>1</sup>,  
Zhangshi Yin<sup>1</sup>, Richard Ullman<sup>1</sup>, Robert McGrath<sup>3</sup>

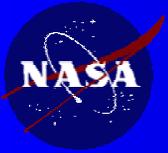
<sup>1</sup>Data Usability Group, NASA/GSFC

<sup>2</sup>NCSA, University of Illinois



# Why HDF-EOS5 XML DTD?

- XML has become important and popular for scientific data exchange
- The HDF-EOS5 XML DTD specifies a standard for using XML to describe the structure and contents of a *single* HDF-EOS5 file.
- The XML DTD can be used for a variety of purposes
  - Viewing structure and contents of HDF-EOS5 file using a Web browser
  - XML as a catalog record for locating datasets
  - XML as an intermediate to other file formats
  - Generation, validation, and reconstruction of HDF-EOS5 file
  - Storing XML in archive as machine readable formats



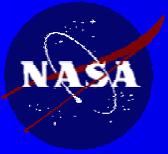
# HDF-EOS5 XML DTD

- Based on the HDF-EOS5 data model (see HDF-EOS5 Users Manual)
- Reused part of the data type description from HDF5 XML DTD
- General Statistics about the DTD
  - 146 elements
- DTD v1.0 is available on the HDF-EOS Website
  - <http://hdfeos.gsfc.nasa.gov/hdfeos/Info/index.cfm>



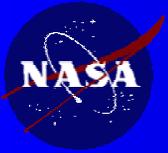
# XML Utility Tools

- Developed utility tools to convert an HDF-EOS5 into an XML file that conforms to the DTD and vice versa.
  - HDF-EOS Extractor: Convert an HDF-EOS5 file to an XML file conforms to the HDF-EOS5 XML DTD with data fields stored either inline or as links to separate Binary or ASCII files.
  - XML to HDF-EOS5 Converter: Convert an XML conforms to the HDF-EOS5 XML DTD to an HDF-EOS5 file (Grid or Swath)
  - Both utility tools are available on the HDF-EOS Website  
<http://hdfeos.gsfc.nasa.gov/hdfeos/Info/index.cfm>



# Lessons Learned

- HDF5 DTD reuse for data types is an overkill
  - HDF5 DTD description for data types has too much low level detail
  - No entity definition for those data types to include
- You can design a DTD, but you are not done until you tested it
  - Made many changes to the DTD during the testing of utility tools
- Could have started with an XML Schema



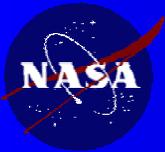
# Status

- HDF-EOS5 XML DTD v1.0 is available for comments
- Currently developing an XML Schema based on the DTD
- Need to develop a better linking mechanism to reference binary data elements (either internal or external to the file)
  - Current linking is using external file system pointers
- Will continue to maintain and enhance the XML DTD/Schema



# Sample XML File – Top Level

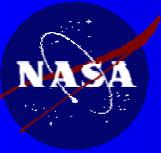
```
<?xml version="1.0" encoding="UTF-8" ?>
<!-- Sample XML file generated by XMLSPY v5 rel. 4 U (http://www.xmlspy.com) -->
<!DOCTYPE HDFEOS-File (View Source for full doctype...)>
- <HDFEOS-File Name="" Type="HDFEOS5">
- <HDFEOS>
  - <Attribute Name="">
    - <DataType>
      <IntegerType ByteOrder="" Sign="" Size="" />
    </DataType>
    <Count />
    <Value />
  </Attribute>
+ <POINTS>
+ <SWATHS>
+ <GRIDS>
</HDFEOS>
- <HDFEOS_INFORMATION>
  - <coremetadata>
    <link_to_data />
  </coremetadata>
  - <archivemetadata>
    <link_to_data />
  </archivemetadata>
  - <StructMetadata>
    <link_to_data />
  </StructMetadata>
  - <ProductMetadata>
    <link_to_data />
  </ProductMetadata>
</HDFEOS_INFORMATION>
</HDFEOS-File>
```



# Sample XML File - Swaths

```
<?xml version="1.0" encoding="UTF-8" ?>
<!-- Sample XML file generated by XMLSPY v5 rel. 4 U
<!DOCTYPE HDFEOS-File (View Source for full doctype...)>
- <HDFEOS-File Name="" Type="HDFEOS5">
- <HDFEOS>
- <Attribute Name="">
- <DataType>
- <IntegerType ByteOrder="" Sign="" Size="" />
</DataType>
<Count />
<Value />
</Attribute>
+ <POINTS>
- <SWATHS>
- <Swath Name="">
- <Attribute Name="">
- <DataType>
- <IntegerType ByteOrder="" Sign="" Size="" />
</DataType>
<Count />
<Value />
</Attribute>
<Dimension Name="" Size="" />
- <GeoField Name="">
- <Attribute Name="">
- <DataType>
- <IntegerType ByteOrder="" Sign="" Size="" />
</DataType>
<Count />
<Value />
</Attribute>
- <DataType>
- <IntegerType ByteOrder="" Sign="" Size="" />
</DataType>
- <DimList>
- <DimName Size="" />
</DimList>
- <MaxDimList>
- <DimName Size="" />
</MaxDimList>
- <FieldStorage>
- <Merge Code="" />
- <Compression Code="">
- <CompParameter />
</Compression>
</FieldStorage>
<link_to_data />
</GeoField>
</HDFEOS-File>
```

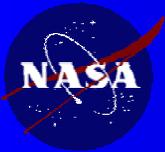
```
- <DataField Name="">
- <Attribute Name="">
- <DataType>
- <IntegerType ByteOrder="" Sign="" Size="" />
</DataType>
<Count />
<Value />
</Attribute>
- <DataType>
- <IntegerType ByteOrder="" Sign="" Size="" />
</DataType>
- <DimList>
- <DimName Size="" />
</DimList>
- <MaxDimList>
- <DimName Size="" />
</MaxDimList>
- <FieldStorage>
- <Merge Code="" />
- <Compression Code="">
- <CompParameter />
</Compression>
</FieldStorage>
<link_to_data />
</DataField>
- <DimensionMapping>
- <RegularDimensionMapping>
- <GeoDimension />
- <DataDimension />
- <Offset />
- <Increment />
</RegularDimensionMapping>
</DimensionMapping>
</Swath>
</SWATHS>
+ <GRIDS>
</HDFEOS>
+ <HDFEOS_INFORMATION>
</HDFEOS-File>
```



# Sample XML File - Grids

```
<?xml version="1.0" encoding="UTF-8" ?>
<!-- Sample XML file generated by XMLSPY v5 rel. 4 U (http://www.
<!DOCTYPE HDFEOS-File (View Source for full doctype...)>
- <HDFEOS-File Name="" Type="HDFEOS5">
- <HDFEOS>
- <Attribute Name="">
- <DataType>
  <IntegerType ByteOrder="" Sign="" Size="" />
</DataType>
  <Count />
  <Value />
</Attribute>
+ <POINTS>
+ <SWATHS>
- <GRIDS>
- <Grid Name="">
- <Attribute Name="">
- <DataType>
  <IntegerType ByteOrder="" Sign="" Size="" />
</DataType>
  <Count />
  <Value />
</Attribute>
- <Parameters>
  <YDim />
  <XDim />
  <UpperLeftPoint X="" Y="" />
  <LowerRightPoint />
- <Projection>
  <GEO Code="HE5_GCTP_GEO" Name="Geographic" ID="0" />
</Projection>
  <SphereCode />
  <ZoneCode />
  <PixelRegistration />
  <OriginType Code="" />
</Parameters>
  <Dimension Name="" Size="" />
- <GridDataField Name="">
- <Attribute Name="">
- <DataType>
  <IntegerType ByteOrder="" Sign="" Size="" />
</DataType>
  <Count />
  <Value />
</Attribute>
```

```
- <DataType>
  <IntegerType ByteOrder="" Sign="" Size="" />
</DataType>
- <DimList>
  <DimName Size="" />
</DimList>
- <MaxDimList>
  <DimName Size="" />
</MaxDimList>
- <FieldStorage>
  <Merge Code="" />
  - <Compression Code="">
    <CompParameter />
  </Compression>
</FieldStorage>
  <TileDimList Code="" Rank="" />
  <link_to_data />
</GridDataField>
</Grid>
</GRIDS>
</HDFEOS>
+ <HDFEOS_INFORMATION>
</HDFEOS-File>
```



# Sample XML File - Points

```
<?xml version="1.0" encoding="UTF-8" ?>
<!-- Sample XML file generated by XMLSPY v5 rel. 4
<!DOCTYPE HDFEOS-File (View Source for full doctype...)>
- <HDFEOS-File Name="" Type="HDFEOS5">
- <HDFEOS>
  - <Attribute Name="">
    - <DataType>
      <IntegerType ByteOrder="" Sign="" Size="" />
    </DataType>
    <Count />
    <Value />
  </Attribute>
  - <POINTS>
    - <Point Name="">
      - <Attribute Name="">
        - <DataType>
          <IntegerType ByteOrder="" Sign="" Size="" />
        </DataType>
        <Count />
        <Value />
      </Attribute>
      - <level Name="">
        - <PointField Order="">
          <FieldName />
          - <DataType>
            <IntegerType ByteOrder="" Sign="" Size="" />
          </DataType>
          <link_to_data />
        </PointField>
      </level>
    </Point>
  </POINTS>
  + <SWATHS>
  + <GRIDS>
</HDFEOS>
+ <HDFEOS_INFORMATION>
</HDFEOS-File>
```