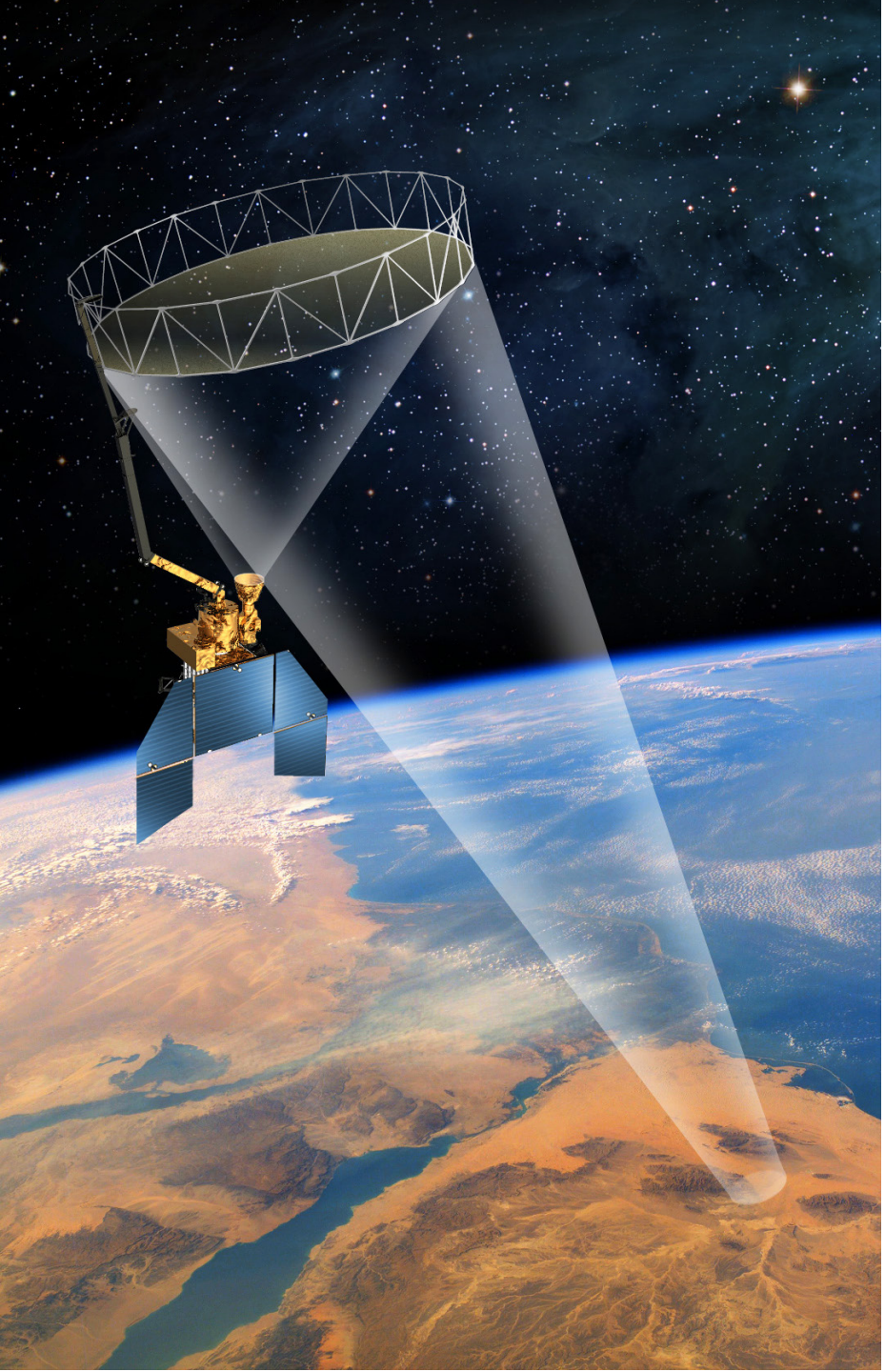


Strategies for Infusing ISO 19115 Metadata in Earth Science Data Systems

AGU 2011. Session IN31C: Advances in Metadata Generation and Applications for Earth Sciences I

Wednesday, December 7, 2011

Hook Hua, Barry Weiss

Jet Propulsion Laboratory, California Institute of Technology

Copyright 2011 California Institute of Technology.
Government sponsorship acknowledged.

The SMAP mission has not been formally approved by NASA. The decision to proceed with the mission will not occur until the completion of the National Environmental Policy Act (NEPA) process. Material in this document related to SMAP is for information purposes only.

Metadata Requirements – Base Reference for NASA Earth Science Data Products

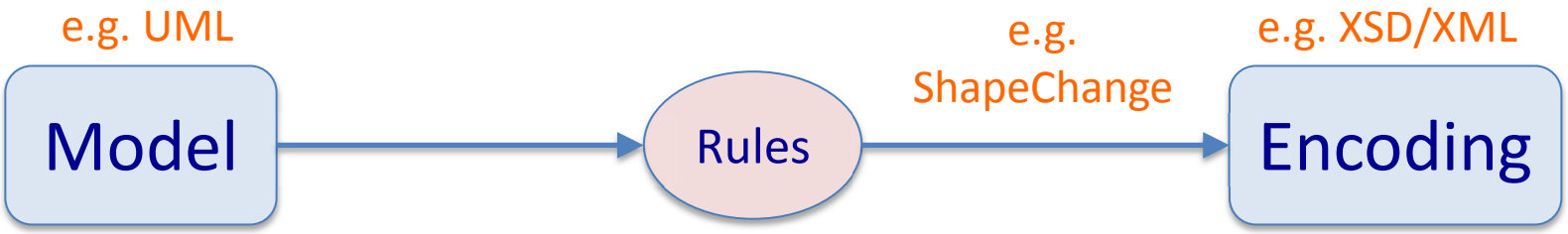
- Implementation guidelines for required metadata to be included in **science data products**.
- Assuring the **consistency** of data requirements across subsystems, and supporting the data standardization necessary for system **interoperability**
- International Organization for Standardization (ISO) Geographic Information – Metadata standard **19115 (and related standards)** shall be used to describe science data products.

- *"The **Level 1** Satellite Mission Data Systems Requirements for Science Data Management states the <<mission/instrument name(s)>> science data products metadata shall conform to ISO 19115 Geographic Information - Metadata standards and adhere to the current ESD-approved implementation."*

SMAP Use of ISO Metadata

- Soil Moisture Active Passive (**SMAP**) Mission Engineers participated in the *Metadata Evolution for NASA Earth Science Data Missions* (**MENDS**) study
- SMAP will be **first** NASA mission to implement support of ISO 19115 (and related standards)
- The effort to **ease** the implementation and utilization of the ISO 19115 metadata standard into Earth Science Data Products
  - Initial goal - Enable the proposed SMAP mission to **incorporate** ISO metadata into data products.
  - Follow-up goal – Provide **infrastructure and documented experience** that enables other NASA sponsored teams to implement and incorporate ISO 19115 metadata into their data products and data systems.

ISO Models and Encodings



Abstract Information Model

Transition from the concepts in the conceptual model to the data types in the schema

Utilizes the encoding rules to provide specific details of their application of an encoding format

- Decoupling of abstract information and specific encodings
  - For example, UML models evolved separately from encodings in XML, OWL, etc.
- Rules defining the encoding schemes

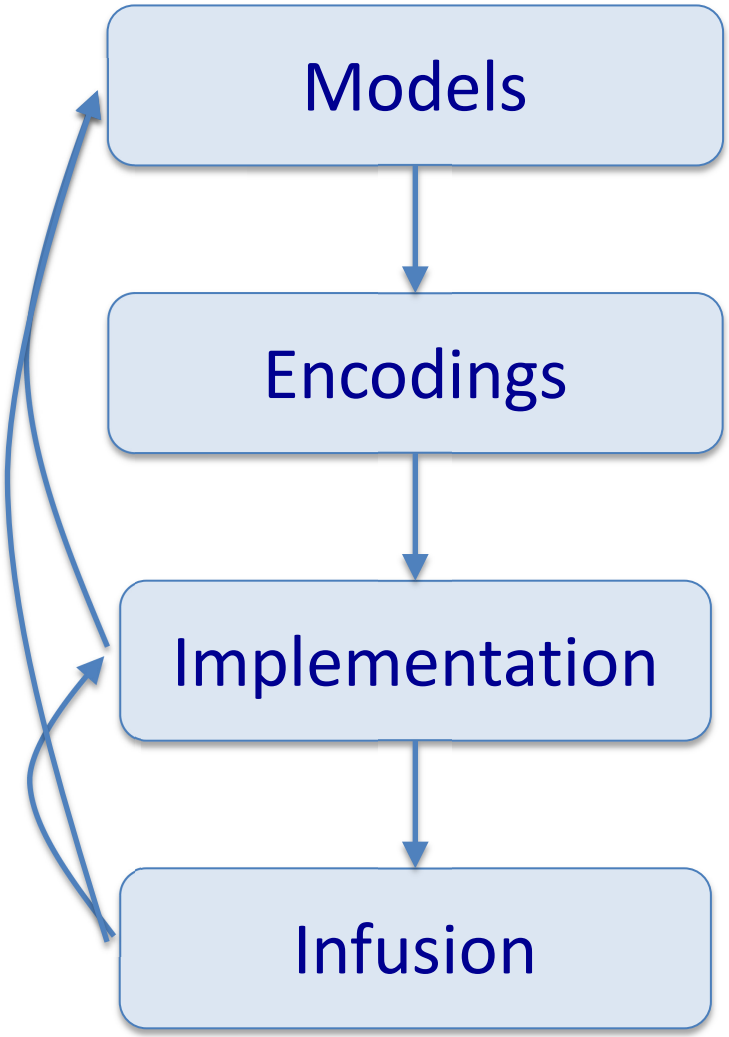
ISO Metadata Models

- ISO 19115:2003
  - Geographic information — Metadata
- ISO 19115-2:2009
  - Geographic information — Metadata Part 2: Extensions for imagery and gridded data
- ISO 19115-1
  - Geographic information — Part 1: Fundamentals
- ISO 19130:2010
  - Geographic information — Imagery sensor models for geopositioning
- ISO 19157
  - Geographic information — Data quality

Rules and Encodings

- Encoding rules for deriving XML schema of the ISO UML models
  - ISO 19109:2005, Geographic information — Rules for application schema
    - Transition from the concepts in the conceptual model to the data types in the application schema
  - ISO 19118:2011, Geographic information — Encoding
    - Specifies requirements for creating encoding rules based on UML schemas, requirements for creating encoding services, and requirements for XML-based encoding rules for neutral interchange of data.
- XSD Schema encoding
  - ISO 19139:2007, Geographic information – Metadata
    - Utilizes the encoding rules defined in ISO 19118 and provides the specific details of their application with regard to deriving XML schema for the UML models in ISO 19115
  - ISO 19139-1 (currently in development)
- OWL encoding (preliminary stages)
  - ISO 19150, Geographic information — Ontology
  - ISO 19150-2, Geographic information — Ontology — Part 2: Rules for developing ontologies in the Web Ontology Language (OWL)

Methodology



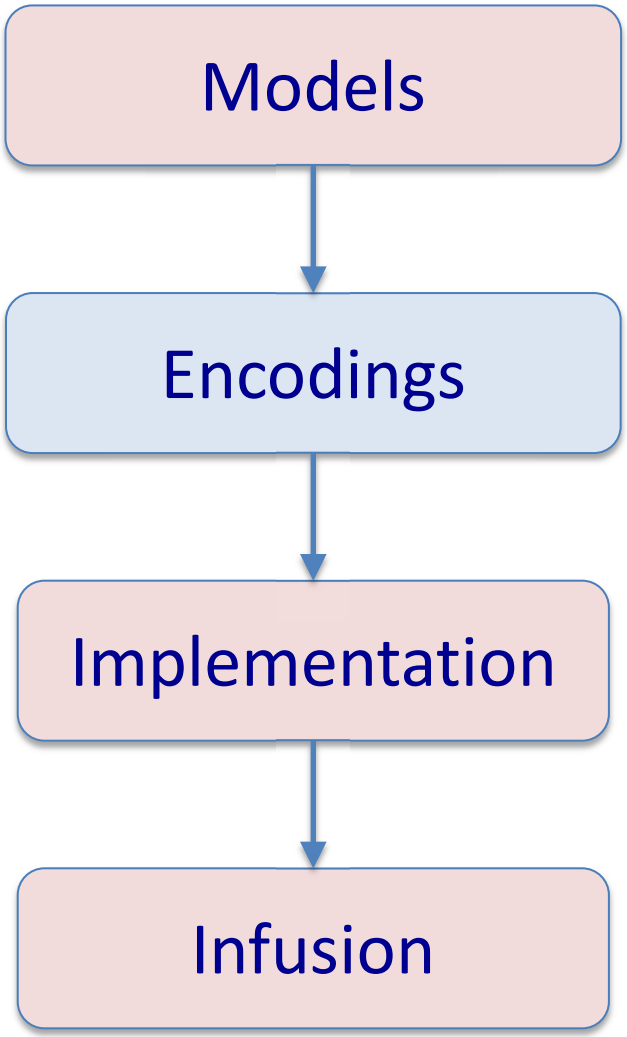
Abstract information model. Can be expressed as an ISO 19115 UML model

Rule-based encodings to apply ISO 19115 into another format such as XSD/XML

APIs for handling specific encodings.

Utilize APIs within executables that generate the Earth science data products

Lessons Learned



Abstract models need to be “adapted” to Earth Science domains

Relative lack of generic and reusable API implementations

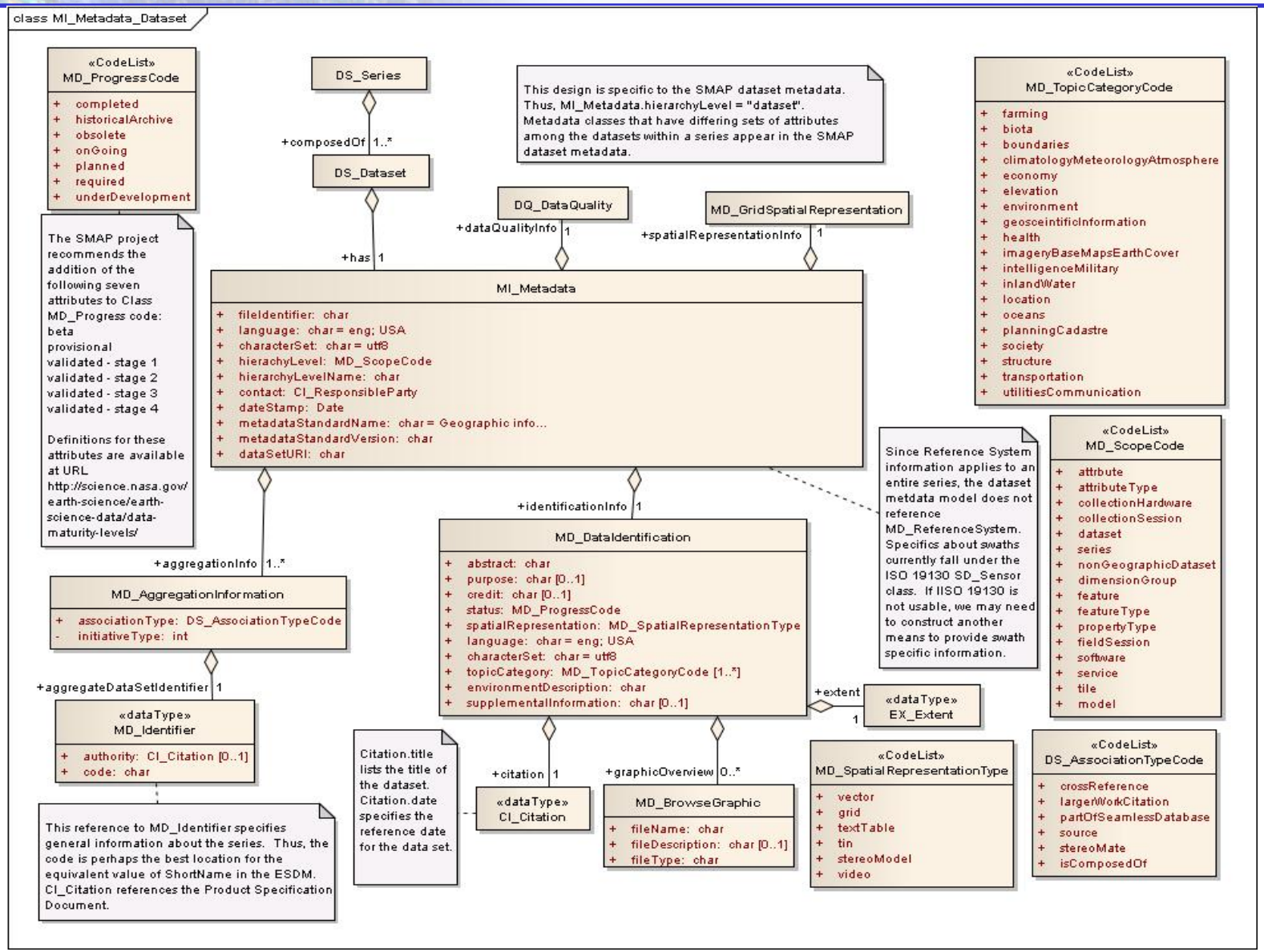
Design how metadata is generated, propagated through data system, and accessed.

Organization of ISO Representation

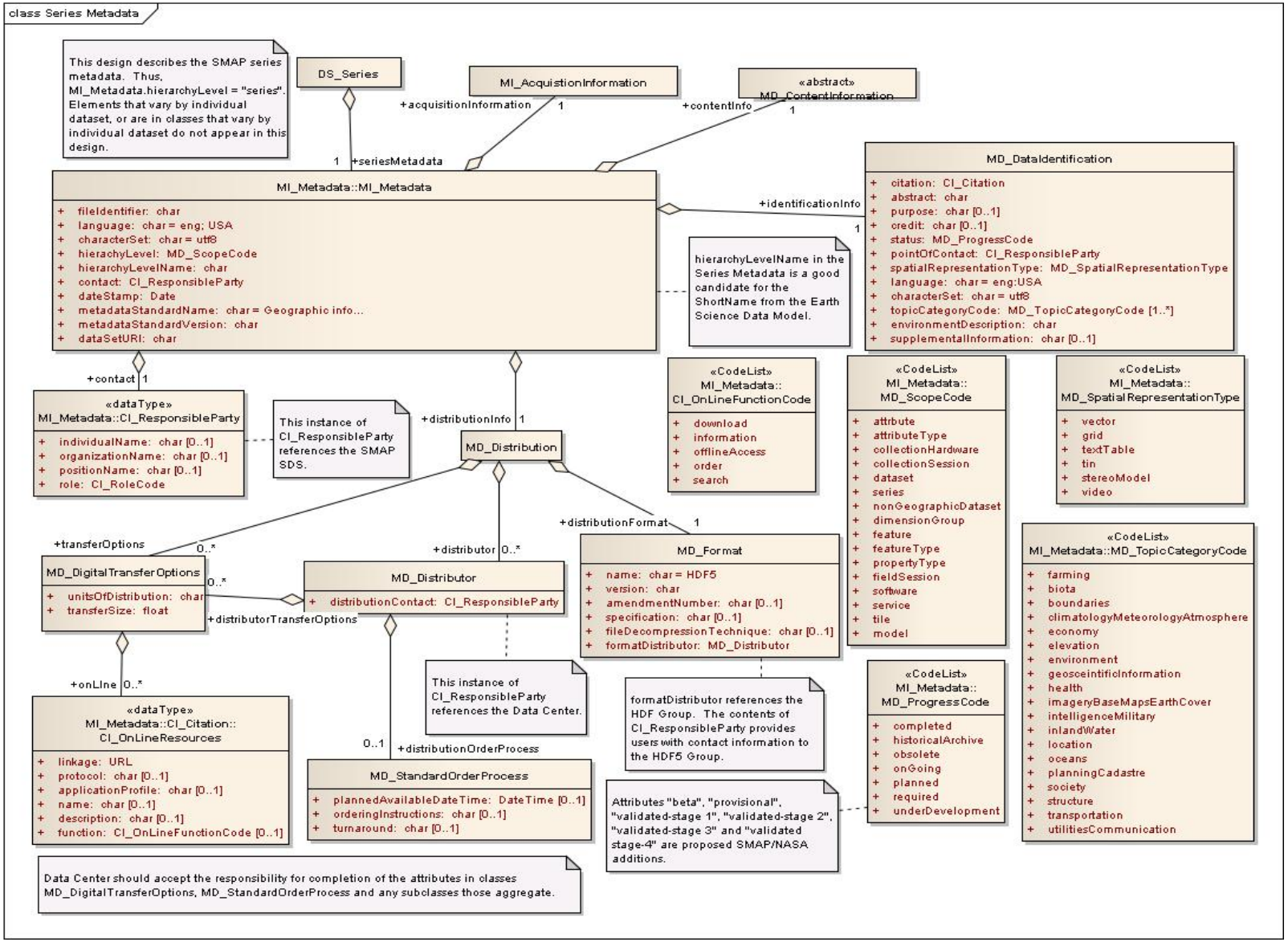
- Dataset metadata
  - Applies to each **granule** in a collection
  - Granule metadata is "*dynamic*".

- Series metadata
  - Applies to a **collection** of data granules.
  - Collection metadata is "*quasi-static*".
  - "Open" and "closed" collections.

SMAP Dataset (Granule) Top-Level Model



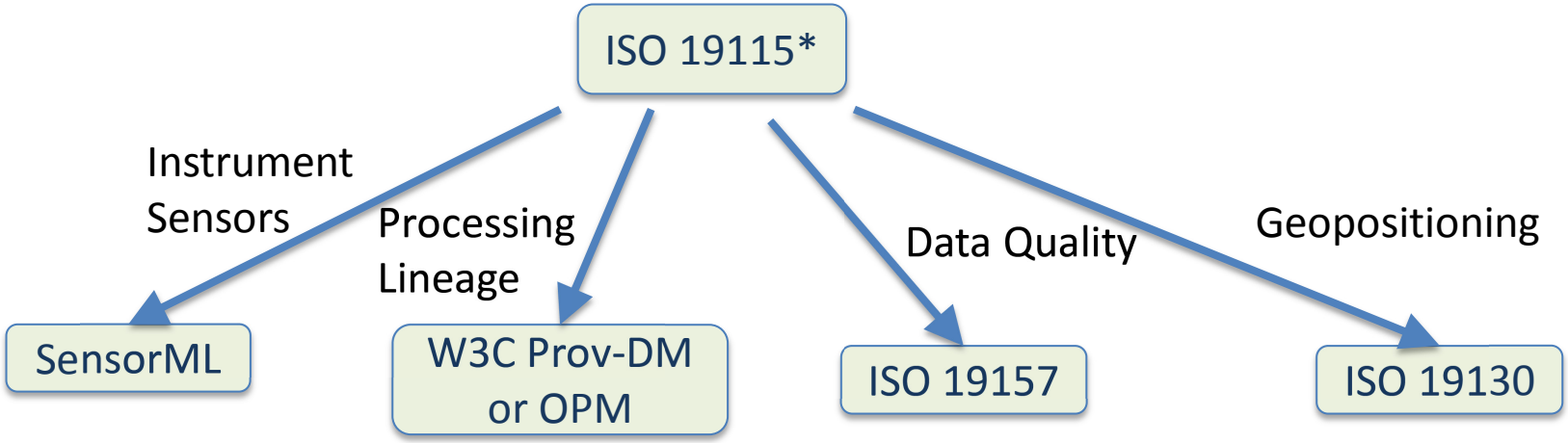
SMAP Series (Collection) Top-Level Model



"Profiles", "Extensions", "Flavors", "Conventions", etc.

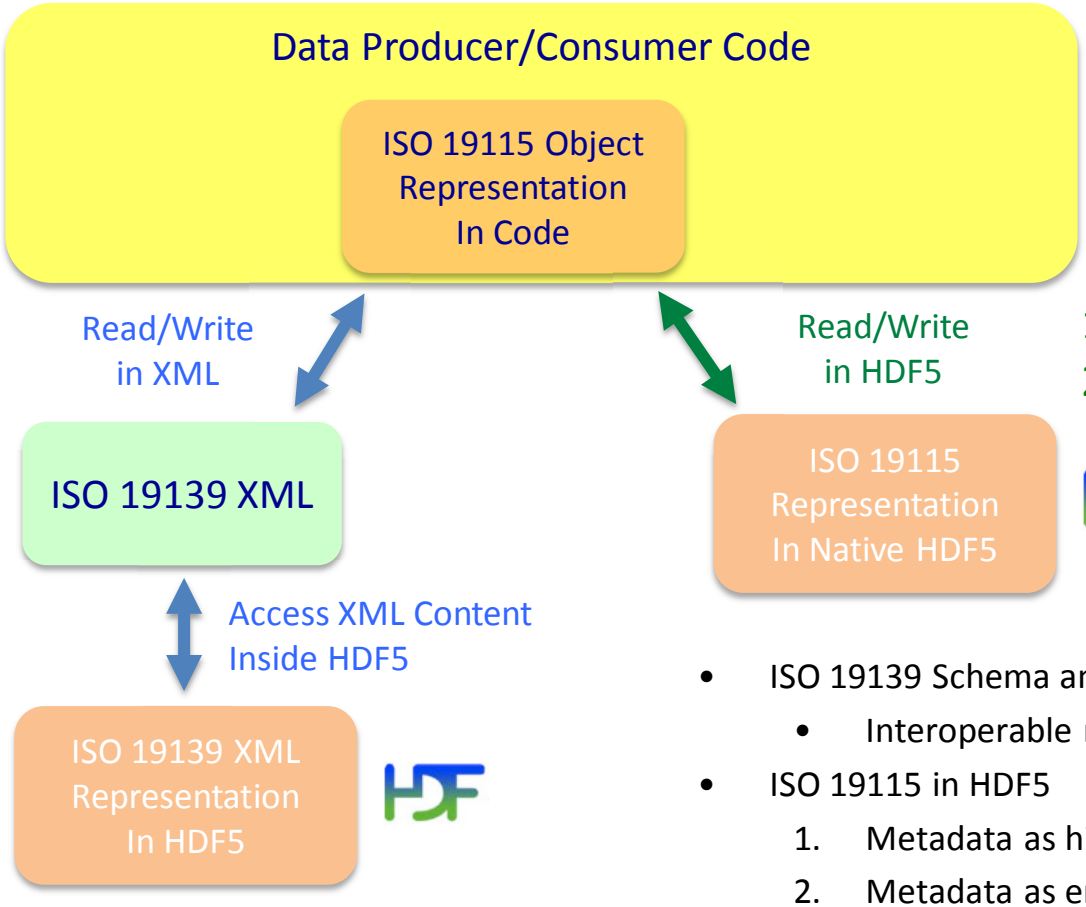
- ISO model enables many utilization methods
- MENDS Recommendation #2: *"The MENDS team recommends that NASA ESDIS create NASA-specific **conventions** for use of the ISO 19115 standard."*
- Additional attributes
  - Product-Specific Attributes (PSA)
  - For example, *Algorithmic Parameters and Thresholds*
    - LI_Lineage/LE_ProcessingStep/LE_Processing/ProcessingInformation
- CodeLists
  - For example, MD_ProgressCode
- Major goal is the generation of NASA conventions for dealing with domain-specific constructs

Metadata Delegation



- ISO 19115 represents generic geographic metadata.
  - For sub-areas, can delegate to other more specific metadata formats
- Leverage constructs for
  - Record and RecordType
    - Enables *typed* arbitrary information embedded into the metadata
  - XLink
    - Hyperlinks in XML documents
  - Citations
    - Link to external URIs or files for artifacts such as ATBD, PDFs and documentation
- Benefit: more comprehensive
- Drawback: more complexity

Metadata Handling

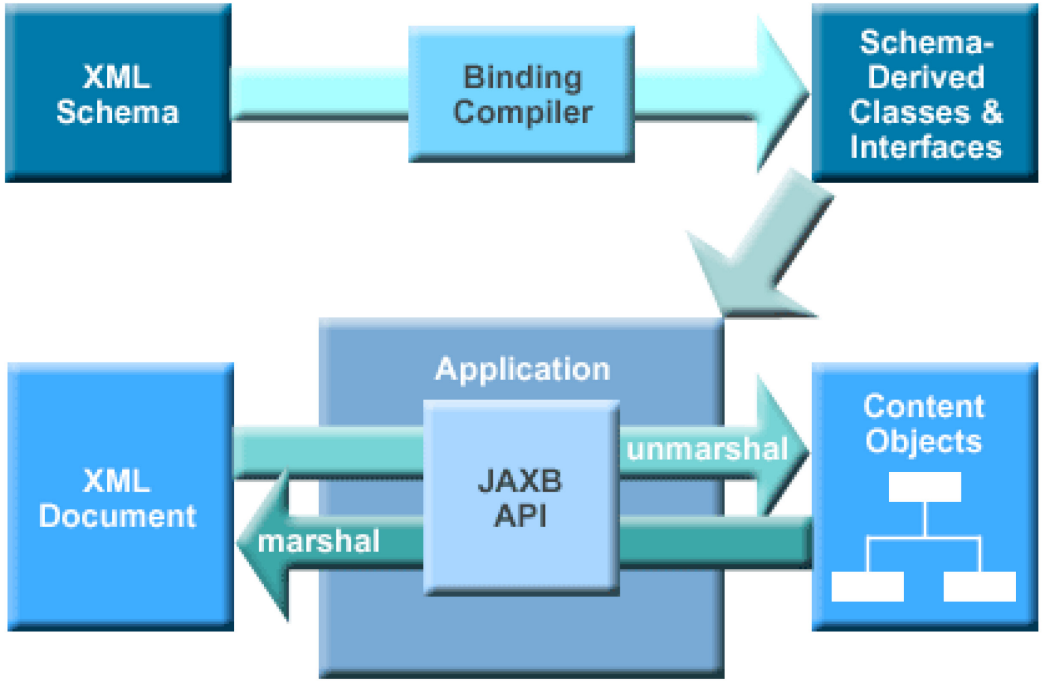


1. XSLT: 19139 XML to HDF's XML
2. 19139 XML mapping using HDF APIs

- ISO 19139 Schema and XML
  - Interoperable metadata interchange format
- ISO 19115 in HDF5
  1. Metadata as hierarchical group/attributes
  2. Metadata as embedded ISO 19139 XML stream
     - Compatible with legacy formats such as HDF4 and NetCDF3
- Conversion cross-walks are possible

XML Data Binding

- Represents information in an XML document as an object in memory
  - No more XML parsing as in DOM or SAX
- Leverages the model in an XSD Schema to create classes and interfaces that adhere to the information structure defined by the schema.
- Enables serialization/deserialization of XML instances to/from code, with **explicit** methods.
- Supports expected **variations** in the ISO model



XML Data Binding Tools

- C++
  - **CodeSynthesis XSD**
    - An open-source, cross-platform W3C XML Schema to C++ data binding compiler.
  - **CodeSynthesis XSD/e: XML for Light-Weight C++ Applications**
    - An open-source, dependency-free XML Schema to C++ compiler for mobile, embedded, and light-weight applications
- C
  - **Mel**
    - An open source C language XML Binding tool with *limited* XSD Schema support
- Java
  - **Java Architecture for XML Binding (JAXB)**
- Python
  - **PyXB**

Code Synthesis XSD

- An open-source, cross-platform W3C XML Schema to C++ data binding compiler.
- Free for non-commercial use
- Supports two XML Schemas to C++ mappings:
  - in-memory C++/Tree
  - stream-oriented C++/Parser.
- http://codesynthesis.com/products/xsd/

```
<contact>
  <name>John Doe</name>
  <email>j@doe.com</email>
  <phone>555 12345</phone>
</contact>
```

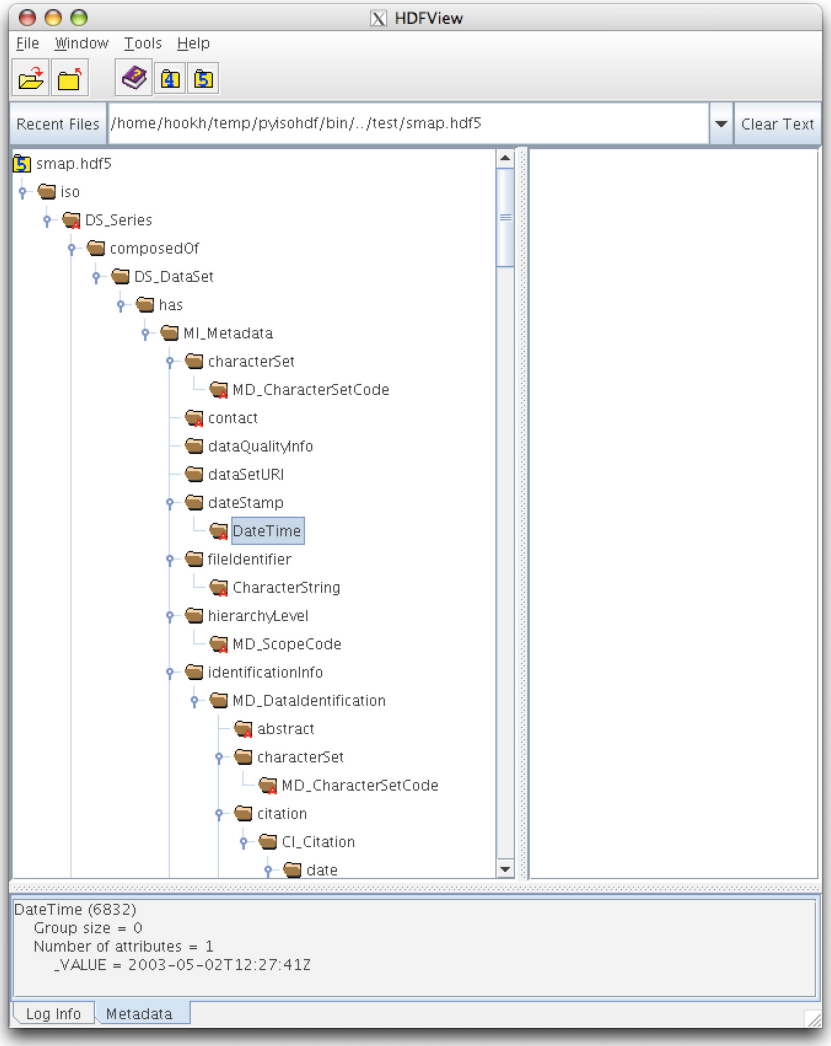
```
auto_ptr<Contact> c = contact ("c.xml");
cout << c->name () << ", "
     << c->email () << ", "
     << c->phone () << endl;
```

Example ISO Metadata Representation

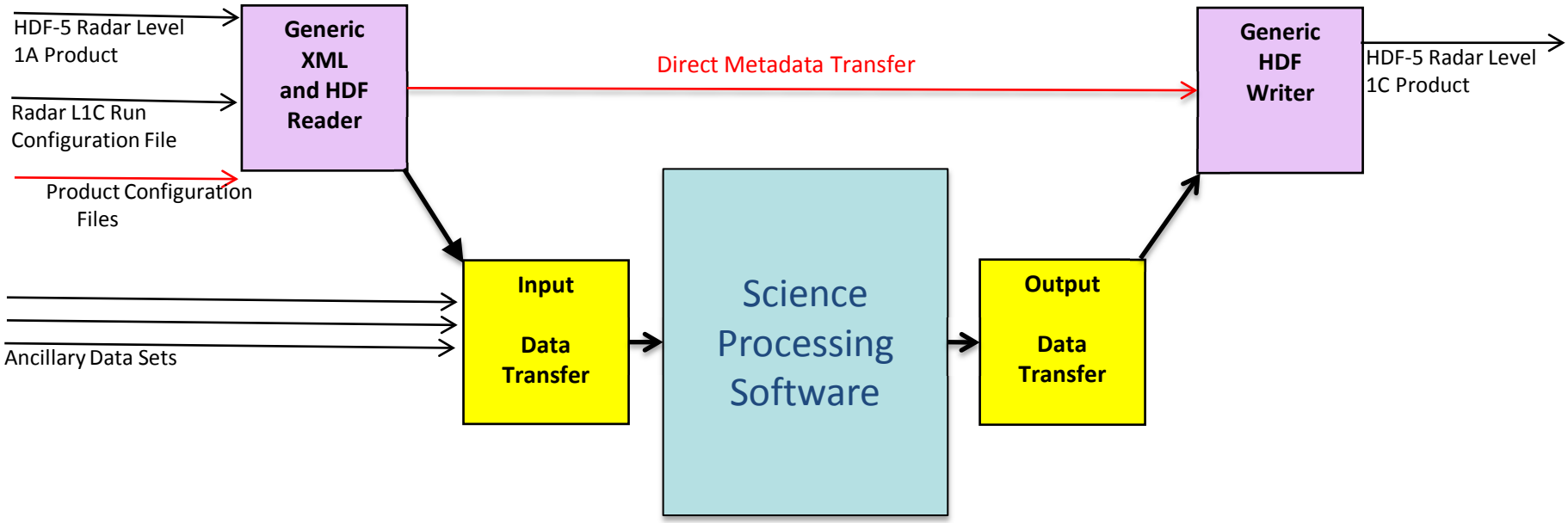
XML view of SMAP Data Product Metadata


```
<gmd:composedOf>
        <gmd:DS_DataSet>
            <gmd:has>
                <gmi:MI_Metadata>
                    <gmd:fileIdentifier>
<gco:CharacterString>SMAP_L2_SM_AP_01767_D_20030502T122741_001.h5</gco:Chara
cterString>
                    </gmd:fileIdentifier>
                    <gmd:language>
                        <gco:CharacterString>eng</gco:CharacterString>
                    </gmd:language>
                    <gmd:characterSet>
                        <gmd:MD_CharacterSetCode
codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml#M
D_CharacterSetCode" codeListValue="UTF8">UTF8</gmd:MD_CharacterSetCode>
                    </gmd:characterSet>
                    <gmd:hierarchyLevel>
                        <gmd:MD_ScopeCode
codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml#M
D_ScopeCode" codeListValue="dataset">dataset</gmd:MD_ScopeCode>
                    </gmd:hierarchyLevel>
                    <gmd:contact xlink:href="#seriesMetadataContact"/>
                    <gmd:dateStamp>
                        <gco:DateTime>2014-09-25T03:00:00Z</gco:DateTime>
                    </gmd:dateStamp>
                    <gmd:metadataStandardName>
                        <gco:CharacterString>ISO 19115-2 Geographic
information — Metadata — Part 2: Extensions for imagery and gridded
                        data</gco:CharacterString>
                     </gmd:metadataStandardName>
                    <gmd:metadataStandardVersion>
                        <gco:CharacterString>ISO 19115-2:2009-02-
15</gco:CharacterString>
                    </gmd:metadataStandardVersion>
```


HDF5 view of SMAP Data Product Metadata



Metadata Mapping Pass-Through



- Many metadata values transfer directly from the input product to the output product
- **XPaths** provide a simple means to *implicitly* codify automated mapping
  - XPaths can be used to drive automated mapping calls.

Challenges

- Generation of a SMAP ISO Model
  - How that might fit with a **NASA specific flavor** of ISO
  - **Coordination** of the effort with other NASA entities and other missions
- **Few** existing ISO support tools and libraries
- Ongoing **updates and changes** to ISO models
  - Prefer approach that minimizes costs of changes
- Generation of tools that
  - Automate **generation** of ISO metadata
  - Automate capability to **read** ISO metadata
  - Enable users to seek and **interpret** available metadata

Conclusions

- Effort will meet the challenge for SMAP as the first of the **Decadal Survey Missions**
  - SMAP will be **first** NASA mission to implement ISO 19115.
  - Provides the necessary **tools** to implement ISO.
  - Decisions about **standard use are still forthcoming**. Need to converge as closely as possible with developing standard.
- Effort will encourage **collaboration** across NASA ESDIS
  - Select optimal technical approach to enable modifications to the metadata schema and reuse by other NASA organizations
  - Regularly review approach and implementation with outside organizations to ensure applicability

BACKUP SLIDES

XML Schema Definition (XSD)

- ISO 19139 provides a set of XSD encodings representing the 19115 model
- XSD provides:
  - A full description of the XML structure
  - Specification of permissible values in an XML document
- Enables validation
  - Can be used to **validate *the structure and the value*** of an XML metadata instance
- Can leverage existing UML and XSD documents of ISO 19115
  - gco: Geographic Common extensible markup language
  - gmd: Geographic MetaData extensible markup language
  - gmx: Geographic Metadata XML Schema
  - gss: Geographic Spatial Schema extensible markup language
  - gsr: Geographic Spatial Referencing extensible markup language
  - gts: Geographic Temporal Schema extensible markup language