

MIGRATION:
PDS3 TO PDS4
(*ATMOSPHERES DEMO*)

Lynn D.V. Neakrase & Lyle Huber
Atmospheres Node

Shannon Rees, Matias Roybal

PDS4 Tech Session, JPL, 28 February - 2 March 2011

OVERVIEW

- General description of the challenge
- Organization of the data (*Things to think about*)
- XML schemas (*How much work?*)
- Python scripting (*Version caveats & pitfalls*)
- Comments, problems, concerns

THE CHALLENGE:

- Data migration from PDS3 to PDS4 is complicated
- ODL doesn't look anything like XML
- Parsing PDS4-appropriate values is non trivial
 - Old values, concatenation, entirely new fields
- Several approaches, pros and cons for each – *mostly based in **data organization***
 - Python programming – general coding with plugging in modules for specific purposes

ORGANIZATION

- “*Mise en place*” - need to prepare for the migration
 - Specific schema (instrument/mission) – tailored from generics, with mission specific material ready for XML
 - Pulling metadata out of PDS3 labels – getting rid of non-essential material (comment blocks, spaces, quotes, etc.)
 - Creating python “dictionaries” or lists (translation tables for assigning field/element values)
 - Python code for doing the migration

ORGANIZATION

- Dividing the PDS3 dataset/volume into 2 separate pieces also helps — 1) data & 2) everything else
- Data files will have similar XML structures throughout the collections *requiring minimal specific schemas*
- Documents, Catalogs, Indices, etc. have slightly different structures, *and may all be different requiring specific schemas for each*

ORGANIZATION

- Dividing the PDS3 dataset/volume into 2 separate pieces also helps — 1) data & 2) everything else

1st

- Data files will have similar XML structures throughout the collections *requiring minimal specific schemas*
(bulk amount of files)

2nd

- Documents, Catalogs, Indices, etc. have slightly different structures, *and may all be different requiring specific schemas for each*
(smaller number of files, but more variety)

XML SCHEMAS

- The basic templates for creating PDS4 labels and documents (*.xsd*)
 - Serve as a blueprint for translating the archive model into usable pieces (*.xml*)
- **Generic Schemas** are provided by the PDS
 - These include the base requirements to satisfy the model

GENERIC XML SCHEMA

BREAKDOWN

Generic_Type Declaration [1]

Identification_Area [1]

Subject_Area [1]

Name_Resolution [0..]*

Cross_Reference_Area [1]

Bibliographic_Reference [0..]*

Observing_System [1..]*

Product_Reference_Entry [0..]*

Observation_Area [1]

Mission_Area [0..]*

Node_Area [0..]*

File_Area [1]

Data_Area [1]

Data_Area_Alternate [0..*]

Product_Generic

GENERIC XML SCHEMA

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <!-- PDS4 XML/Schema for Product_Generic_0.2.0.0.d Mon Dec 13 20:27:43 PST 2010 -->
3 <!-- Generated from the PDS4 Information Model V0.2.0.0.d -->
4 <!-- *** This PDS4 product schema is a preliminary deliverable. *** -->
5 <!-- *** It is being made available for review and testing. *** -->
6 <xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
7   targetNamespace="http://pds.nasa.gov/schema/pds4/pds"
8   xmlns:pds="http://pds.nasa.gov/schema/pds4/pds"
9   elementFormDefault="qualified"
10  attributeFormDefault="unqualified"
11  version="2.0.0">
12
13 <xsd:include schemaLocation="Extended_Types_0200d.xsd">
14 <xsd:annotation>
15 <xsd:documentation>PDS (common) Data Dictionary</xsd:documentation>
16 </xsd:annotation>
17 </xsd:include>
18
19 <!-- <xsd:import namespace="http://pds.nasa.gov/schema/pds4/anyNS"
20   schemaLocation="Any_Described_Data_Object_0200d.xsd"/> -->
21
22 <xsd:complexType name="Product_Generic_Type">
23 <xsd:sequence>
24 <xsd:annotation>
25 <xsd:documentation>
26 The Product Generic class defines a template for all data products.
27 </xsd:documentation>
28 </xsd:annotation>
29 <xsd:element name="Identification_Area_Product" type="pds:Identification_Area_Product_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:I
30 <xsd:element name="Cross_Reference_Area_Product" type="pds:Cross_Reference_Area_Product_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS
31 <xsd:element name="Observation_Area" type="pds:Observation_Area_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Observation_Area -->
32 <xsd:element name="File_Area" type="pds:File_Area_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:File_Area -->
33 <xsd:element name="Data_Area_Alternate" type="pds>Data_Area_Alternate_Type" minOccurs="0" maxOccurs="unbounded"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Data_Area
34 <xsd:element name="Data_Area_Generic" type="pds>Data_Area_Generic_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Data_Area_Generic -->
35 </xsd:sequence>
36 </xsd:complexType>
37
38 <xsd:complexType name="Identification_Area_Product_Type">
39 <xsd:sequence>
40 <xsd:annotation>
41 <xsd:documentation>
42 The product identification area consists of attributes involved in the role of identifying and naming a data product.
43 </xsd:documentation>
44 </xsd:annotation>
45 <xsd:element name="logical_identifier" type="pds:logical_identifier_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area
46 <xsd:element name="version_id" type="pds:version_id_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area:version_id -->
47 <xsd:element name="product_class" type="pds:product_class_Type" fixed="Product_Generic" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Ident
48 <xsd:element name="title" type="pds:title_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area:title -->
49 <xsd:element name="alternate_title" type="pds:alternate_title_Type" minOccurs="0" maxOccurs="unbounded"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Ar
50 <xsd:element name="alternate_id" type="pds:alternate_id_Type" minOccurs="0" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area:alternate_i
51 <xsd:element name="last_modification_date_time" type="pds:last_modification_date_time_Type" minOccurs="0" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:I
52 <xsd:element name="product_subclass" type="pds:product_subclass_Type" minOccurs="0" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area:pro
53 <xsd:element name="Subject_Area" type="pds:Subject_Area_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Subject_Area -->
54 </xsd:sequence>
55 </xsd:complexType>
```

no specific information for datatypes throughout!



Product_Generic

TAILORED SCHEMAS

- **Tailored Schemas** are ready to produce simple XML labels for a standard set of products
- More detailed than the generics, but still may not contain all necessary fields for a mission (etc.)
- A library of tailored schemas are also provided by the PDS
 - *Product_Table_Character, Product_Image_Grayscale,*

TAILORED XML SCHEMA

BREAKDOWN

Table_Character_Type Declaration [1]

added information to generic

Identification_Area

Subject_Area

Name_Resolution

Cross_Reference_Area

Bibliographic_Reference

Observing_System

Product_Reference_Entry

Observation_Area

Mission_Area

Node_Area

File_Area

Data_Area

Data_Area_Alternate

specific to Table_Character

Product_Table_Character

TAILORED XML SCHEMA

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <!-- PDS4 XML/Schema for Product_Table_Character_0.2.0.0.d Mon Dec 13 20:27:43 PST 2010 -->
3 <!-- Generated from the PDS4 Information Model V0.2.0.0.d -->
4 <!-- *** This PDS4 product schema is a preliminary deliverable. *** -->
5 <!-- *** It is being made available for review and testing. *** -->
6 <xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
7   targetNamespace="http://pds.nasa.gov/schema/pds4/pds"
8   xmlns:pds="http://pds.nasa.gov/schema/pds4/pds"
9   elementFormDefault="qualified"
10  attributeFormDefault="unqualified"
11  version="2.0.0">
12
13 <xsd:include schemaLocation="Extended_Types_0200d.xsd">
14 <xsd:annotation>
15 <xsd:documentation>PDS (common) Data Dictionary</xsd:documentation>
16 </xsd:annotation>
17 </xsd:include>
18
19 <!-- <xsd:import namespace="http://pds.nasa.gov/schema/pds4/anyNS"
20   schemaLocation="Any_Described_Data_Object_0200d.xsd"/> -->
21
22 <xsd:complexType name="Product_Table_Character_Type">
23 <xsd:sequence>
24 <xsd:annotation>
25 <xsd:documentation>
26   The Product Table Character class defines a product consisting of at least one character table and other associated data objects and metadata.
27 </xsd:documentation>
28 </xsd:annotation>
29 <xsd:element name="Identification_Area_Product" type="pds:Identification_Area_Product_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:I
30 <xsd:element name="Cross_Reference_Area_Product" type="pds:Cross_Reference_Area_Product_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS
31 <xsd:element name="Observation_Area" type="pds:Observation_Area_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Observation_Area -->
32 <xsd:element name="File_Area" type="pds:File_Area_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:File_Area -->
33 <xsd:element name="Data_Area_Table_Character" type="pds:Data_Area_Table_Character_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Data_
34 <xsd:element name="Data_Area_Alternate" type="pds:Data_Area_Alternate_Type" minOccurs="0" maxOccurs="unbounded"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Data_Area
35 </xsd:sequence>
36 </xsd:complexType>
37
38 <xsd:complexType name="Identification_Area_Product_Type">
39 <xsd:sequence>
40 <xsd:annotation>
41 <xsd:documentation>
42   The product identification area consists of attributes involved in the role of identifying and naming a data product.
43 </xsd:documentation>
44 </xsd:annotation>
45 <xsd:element name="logical_identifier" type="pds:logical_identifier_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area
46 <xsd:element name="version_id" type="pds:version_id_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area:version_id -->
47 <xsd:element name="product_class" type="pds:product_class_Type" fixed="Product_Table_Character" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:P
48 <xsd:element name="title" type="pds:title_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area:title -->
49 <xsd:element name="alternate_title" type="pds:alternate_title_Type" minOccurs="0" maxOccurs="unbounded"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Ar
50 <xsd:element name="alternate_id" type="pds:alternate_id_Type" minOccurs="0" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area:alternate_i
51 <xsd:element name="last_modification_date_time" type="pds:last_modification_date_time_Type" minOccurs="0" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:I
52 <xsd:element name="product_subclass" type="pds:product_subclass_Type" minOccurs="0" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area:pro
53 <xsd:element name="Subject_Area" type="pds:Subject_Area_Type" minOccurs="1" maxOccurs="1"> </xsd:element> <!-- RA_0001_NASA_PDS_1:PDS:Subject_Area -->
54 </xsd:sequence>
55 </xsd:complexType>
```

added specific information for
Table_Character files



Product_Table_Character

SPECIFIC SCHEMAS

- ***Specific Schemas*** are modified from tailored schemas by the inclusion of one or more dictionaries
- These would include, *node* or *mission* dictionaries with special rules from the nodes or instrument specific details not included in the generic dictionaries
- All additional dictionaries conform to the PDS4 model but can be modified for specific purposes

SPECIFIC SCHEMAS

- Only need a *specific schema* for each individual datatype for a mission

Example

- *Phoenix* (ATM) data are all TABLEs, so we can get away with **one** specific schema to produce **all** data labels
- All instrument specific elements are added to the Mission section of the schema (can be substituted)

SPECIFIC XML SCHEMA

BREAKDOWN

Table_Character_Type Declaration [1]

Identification_Area [1]

Subject_Area [1]

Name_Resolution [0..]*

Cross_Reference_Area [1]

Bibliographic_Reference [0..]*

Observing_System [1..]*

Product_Reference_Entry [0..]*

Observation_Area [1]

PHOENIX_Mission_Area [1]

Added a Phoenix Mission Dictionary

Node_Area [0..]*

Removed Node Area (not used)

File_Area [1]

Data_Area [1]

Data_Area_Alternate [0..*]

PHX_Product_Table_Character

SPECIFIC XML SCHEMA

The screenshot shows the XML Editor interface with the following XSD schema content:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <!-- PDS4 XML/Schema for Product_Table_Character_0.2.0.0.d Mon Dec 13 20:27:43 PST 2010 -->
3 <!-- Generated from the PDS4 Information Model V0.2.0.0.d -->
4 <!-- *** This PDS4 product schema is a preliminary deliverable. *** -->
5 <!-- *** It is being made available for review and testing. *** -->
6 <xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
7   targetNamespace="http://pds.nasa.gov/schema/pds4/pds"
8   xmlns:pds="http://pds.nasa.gov/schema/pds4/pds"
9   xmlns:phxmd="file:/home/shannon/Documents/WORK/phxmd"
10  xmlns:compps3="http://pds.nasa.gov/schema/pds4/compps3"
11  elementFormDefault="qualified"
12  attributeFormDefault="unqualified"
13  version="2.0.0">
14
15  <xsd:import namespace="http://pds.nasa.gov/schema/pds4/compps3" schemaLocation="Data_Dict_compps3_0200d.xsd"/>
16  <xsd:import namespace="file:/home/shannon/Documents/WORK/phxmd" schemaLocation="PHXMD.xsd"/>
17  <xsd:include schemaLocation="Extended_Types_0200d.xsd">
18    <xsd:annotation>
19      <xsd:documentation>PDS (common) Data Dictionary</xsd:documentation>
20    </xsd:annotation>
21  </xsd:include>
22
23  <!-- <xsd:import namespace="http://pds.nasa.gov/schema/pds4/anyNS"
24    schemaLocation="Any_Described_Data_Object_0200d.xsd"/> -->
25
26  <xsd:complexType name="Product_Table_Character_Type">
27    <xsd:sequence>
28      <xsd:annotation>
29        <xsd:documentation>
30          The Product Table Character class defines a product consisting of at least one character table and other associated data objects and metadata.
31        </xsd:documentation>
32      </xsd:annotation>
33      <xsd:element name="Identification_Area_Product" type="pds:Identification_Area_Product_Type" minOccurs="1" maxOccurs="1"/> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area_Product -->
34      <xsd:element name="Cross_Reference_Area_Product" type="pds:Cross_Reference_Area_Product_Type" minOccurs="1" maxOccurs="1"/> <!-- RA_0001_NASA_PDS_1:PDS:Cross_Reference_Area_Product -->
35      <xsd:element name="Observation_Area" type="pds:Observation_Area_Type" minOccurs="1" maxOccurs="1"/> <!-- RA_0001_NASA_PDS_1:PDS:Observation_Area -->
36      <xsd:element name="File_Area" type="pds:File_Area_Type" minOccurs="1" maxOccurs="1"/> <!-- RA_0001_NASA_PDS_1:PDS:File_Area -->
37      <xsd:element name="Data_Area_Table_Character" type="pds:Data_Area_Table_Character_Type" minOccurs="1" maxOccurs="1"/> <!-- RA_0001_NASA_PDS_1:PDS:Data_Area_Table_Character -->
38    </xsd:sequence>
39  </xsd:complexType>
40
41  <xsd:complexType name="Identification_Area_Product_Type">
42    <xsd:sequence>
43      <xsd:annotation>
44        <xsd:documentation>
45          The product identification area consists of attributes involved in the role of identifying and naming a data product.
46        </xsd:documentation>
47      </xsd:annotation>
48      <xsd:element name="logical_identifier" type="pds:logical_identifier_Type" minOccurs="1" maxOccurs="1"/> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area:logical_identifier -->
49      <xsd:element name="version_id" type="pds:version_id_Type" minOccurs="1" maxOccurs="1"/> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area:version_id -->
50      <xsd:element name="product_class" type="pds:product_class_Type" fixed="Product_Table_Character" minOccurs="1" maxOccurs="1"/> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area:product_class -->
51      <xsd:element name="title" type="pds:title_Type" minOccurs="1" maxOccurs="1"/> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area:title -->
52      <xsd:element name="last_modification_date_time" type="pds:last_modification_date_time_Type" minOccurs="0" maxOccurs="1"/> <!-- RA_0001_NASA_PDS_1:PDS:Identification_Area:last_modification_date_time -->
53      <xsd:element name="Subject_Area" type="pds:Subject_Area_Type" minOccurs="1" maxOccurs="1"/> <!-- RA_0001_NASA_PDS_1:PDS:Subject_Area -->
54    </xsd:sequence>
55  </xsd:complexType>
```

Only need a schema for each different datatype

PHX_Product_Table_Character

SPECIFIC XML SCHEMA - MISSION AREA

```
XPath 2.0
Product_Table_Character_0200d.xsd
145
146 <xsd:complexType name="Mission_Area_Type">
147   <xsd:sequence>
148     <xsd:annotation>
149       <xsd:documentation>
150         The mission area allow the insertion of mission specific metadata.
151       </xsd:documentation>
152     </xsd:annotation>
153     <!-- When creating a specific XML schema, remove the 'xsd:any' element. You may insert any described nondigital object, one or more times. -->
154     <xsd:element name="PHX_Local_True_Solar_Time" type="compds3:local_true_solar_time_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
155     <xsd:element name="PHX_Local_Mean_Solar_Time" type="compds3:local_mean_solar_time_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
156     <xsd:element name="PHX_Mission_Phase_Name" type="compds3:mission_phase_name_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
157     <xsd:element name="PHX_Planet_Day_Number" type="compds3:planet_day_number_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
158     <xsd:element name="PHX_Filter_Name" type="phxmd:filter_name_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
159     <xsd:element name="PHX_Commanded_Parameters" type="pds:Commanded_Parameters_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
160     <xsd:element name="PHX_History_Data_Elements" type="pds:History_Data_Elements_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
161     <xsd:element name="PHX_Object_Header" type="pds:Object_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
162     <!-- <xsd:element name="Any_NonDigital_Object" type="pds:Any_NonDigital_Object_Type" minOccurs="0" maxOccurs="unbounded" /> </xsd:element -->
163   </xsd:sequence>
164 </xsd:complexType>
165
166 <xsd:complexType name="PHX_History_Data_Elements_Type">
167   <xsd:sequence>
168     <xsd:element name="PHX_Software_Name" type="compds3:software_name_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
169     <xsd:element name="PHX_Software_Version_Id" type="compds3:software_version_id_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
170     <xsd:element name="PHX_Processing_History" type="compds3:processing_history_text_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
171   </xsd:sequence>
172 </xsd:complexType>
173
174 <xsd:complexType name="PHX_Commanded_Parameters_Type">
175   <xsd:sequence>
176     <xsd:element name="PHX_Ops_Token_Activity" type="phxmd:ops_token_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
177     <xsd:element name="PHX_Instrument_Mode_Id" type="compds3:instrument_mode_id_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
178     <xsd:element name="PHX_Detector_Id" type="compds3:detector_id_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
179     <xsd:element name="PHX_Pressure_Threshold" type="phxmd:pressure_threshold_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
180     <xsd:element name="PHX_Temperature_Threshold" type="phxmd:temperature_threshold_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
181     <xsd:element name="PHX_Period_Duration" type="phxmd:period_duration_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
182     <xsd:element name="PHX_Period_Number" type="phxmd:period_number_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
183     <xsd:element name="PHX_Integration_Duration" type="compds3:integration_duration_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
184     <xsd:element name="PHX_Integration_Number" type="phxmd:integration_number_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
185   </xsd:sequence>
186 </xsd:complexType>
187
188 <xsd:complexType name="PHX_Object_Header_Type">
189   <xsd:sequence>
190     <xsd:element name="PHX_object_type" type="phxmd:object_type_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
191     <xsd:element name="PHX_object_record_bytes" type="phxmd:object_record_bytes_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
192     <xsd:element name="PHX_object_records" type="phxmd:object_records_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
193     <xsd:element name="PHX_object_description_HEADER1" type="phxmd:object_description_Type" minOccurs="1" maxOccurs="1" /> </xsd:element>
194   </xsd:sequence>
195 </xsd:complexType>
196
197
```

PHX_Product_Table_Character

PDS3 METADATA

- We strip down the PDS3 labels, pulling translatable values out of the label dropping all non-essential pieces
 - Comment blocks, blank lines, etc.
- This can be handled directly within the Python code we're using to migrate the data
 - Python concept of “dictionaries” and “lists”

PDS3 METADATA

PDS3 Label

```
1 PDS_VERSION_ID          = PDS3
2
3 /* FILE DATA ELEMENTS */
4
5 RECORD_TYPE              = FIXED_LENGTH
6 RECORD_BYTES             = 97
7 FILE_RECORDS             = 36352
8
9
10
11 /* POINTERS TO DATA OBJECTS */
12
13 ^TABLE                   = "MS000EMH_00896227783_10C6M1.TAB"
14
15
16 /* IDENTIFICATION DATA ELEMENTS */
17
18 DATA_SET_ID             = "PHX-M-MET-2-PT-EDR-V1.0"
19 PRODUCT_ID               = "MS000EMH_00896227783_10C6M1"
20 PRODUCT_TYPE             = EDR
21 PRODUCT_VERSION_ID      = "V1.5 D-33236"
22 RELEASE_ID              = "0001"
23 INSTRUMENT_HOST_ID      = PHX
24 INSTRUMENT_HOST_NAME    = "PHOENIX"
25 INSTRUMENT_ID           = MET
26 INSTRUMENT_TYPE         = "IN SITU METEOROLOGY"
27 LOCAL_TRUE_SOLAR_TIME   = "17:03:01"
28 LOCAL_MEAN_SOLAR_TIME   = "16:49:31"
29 MISSION_NAME            = "PHOENIX"
30 MISSION_PHASE_NAME      = "PRIMARY MISSION"
31 PLANET_DAY_NUMBER       = 0
32 PRODUCER_INSTITUTION_NAME = "YORK UNIVERSITY"
33 PRODUCT_CREATION_TIME   = 2008-06-05T06:55:25.877
34 OPS_TOKEN               = 16#10C60000#
35 SPACECRAFT_CLOCK_CNT_PARTITION = 1
36 SPACECRAFT_CLOCK_START_COUNT = "896227783.309"
37 START TIME              = 2008-05-26T00:08:36.308
```

PDS3 Relevant Information

Anything Relevant to PDS4

Tag and Value

RECORD_TYPE	FIXED LENGTH
RECORD_BYTES	97
FILE_RECORDS	36352
^TABLE	MS000EMH_....
DATA_SET_ID	PHX-M-MET-2...

etc.

PYTHON “DICTIONARY”

- Once the data is ripped from the PDS3 labels, we create a Python “dictionary”
 - This works as a translation table that shows Python how to map PDS3 values to appropriate values in PDS4 — should be 1:1 or 2:1 (etc.) to avoid duplication*

*NOTE: (example)

- *<local_identifier>* poses a problem with multiple locations with different meanings (i.e., 1:2 mapping)
 - Results in overwriting

PYTHON "DICTIONARY"

```
754
755     data_dict = {'logical_identifer':'logical_identifer', 'version_id':'version_id', 'product_class':'product_class', 'title':'title', 'last_modific
ation_date_time':'PRODUCT_CREATION_TIME', 'target_name':'TARGET_NAME', 'instrument_name':('INSTRUMENT_ID', 'INSTRUMENT_NAME'), 'instrument_host_name':'INST
RUMENT_HOST_NAME', 'local_identifer':'REFERENCE_KEY_ID', 'reference_text':'REFERENCE_DESC', 'title_OBSERVINGSYSTEM':('INSTRUMENT_ID', 'INSTRUMENT_NAME'),
'observing_system_component_type':'observing_system_component_type', 'lid_reference':'lid_reference', 'reference_association_type':'reference_association
_type', 'start_date_time':'START_TIME', 'stop_date_time':'STOP_TIME', 'spacecraft_clock_start_count':('SPACECRAFT_CLOCK_START_COUNT', 'N/A'), 'spacecraft_
clock_stop_count':('SPACECRAFT_CLOCK_STOP_COUNT', 'N/A'), 'PHX_Local_True_Solar_Time':'LOCAL_TRUE_SOLAR_TIME', 'PHX_Local_Mean_Solar_Time':'LOCAL_MEAN_SOL
AR_TIME', 'PHX_Mission_Phase_Name':'MISSION_PHASE_NAME', 'PHX_Planet_Day_Number':'PLANET_DAY_NUMBER', 'PHX_Filter_Name':('FILTER_NAME', 'DELETEME'), 'PHX_
Ops-Token_Activity':'OPS_TOKEN_ACTIVITY', 'PHX_Instrument_Mode_Id':'INSTRUMENT_MODE_ID', 'PHX_Detector_Id':'DETECTOR_ID', 'PHX_Pressure_Threshold':'PRESSU
RE_THRESHOLD', 'PHX_Temperature_Threshold':'TEMPERATURE_THRESHOLD', 'PHX_Period_Duration':'PERIOD_DURATION', 'PHX_Period_Number':'PERIOD_NUMBER', 'PHX_Int
egration_Duration':'INTEGRATION_DURATION', 'PHX_Integration_Number':'INTEGRATION_NUMBER', 'PHX_Software_Name':'SOFTWARE_NAME', 'PHX_Software_Version_Id':
'SOFTWARE_VERSION_ID', 'PHX_Processing_History':'PROCESSING_HISTORY_TEXT', 'PHX_object_type':'HEADER', 'PHX_object_record_bytes':'header_bytes', 'PHX_obje
ct_records':'RECORDS', 'PHX_object_description':'DESCRIPTION', 'local_identifer_FILEAREA':'file_local_identifer', 'file_local_identifer':'file_local_id
entifer', 'creation_date_time':'time', 'file_name':'^TABLE', 'file_size':'size', 'max_record_bytes':'ROW_BYTES', 'records':'ROWS', 'md5_checksum':'check
sum', 'local_identifer_TABLECHARACTER':'Table_local_identifer', 'records_TABLECHARACTER':'ROWS', 'encoding_type':'INTERCHANGE_FORMAT', 'fields':'COLUMNS
', 'record_bytes':'ROW_BYTES', 'offset':'START_BYTE0', 'field_name':'NAME0', 'field_number':'COLUMN_NUMBER0', 'field_data_type':'DATA_TYPE0', 'field_locat
ion':'START_BYTE0', 'field_length':'BYTES0', 'field_format':('FORMAT0', 'N/A'), 'field_min_logical':'field_min_logical0', 'field_max_logical':'field_max_lo
gical0', 'field_value_offset':'START_BYTE0', 'field_unit':'UNIT0', 'field_description':'DESCRIPTION0', 'field_name01':'NAME1', 'field_number01':'COLUMN_NU
MBER1', 'field_data_type01':'DATA_TYPE1', 'field_location01':'START_BYTE1', 'field_length01':'BYTES1', 'field_format01':('FORMAT1', 'N/A'), 'field_min_log
ical01':'field_min_logical1', 'field_max_logical01':'field_max_logical1', 'field_value_offset01':'START_BYTE1', 'field_unit01':'UNIT1', 'field_description0
1':'DESCRIPTION1', 'field_name02':'NAME2', 'field_number02':'COLUMN_NUMBER2', 'field_data_type02':'DATA_TYPE2', 'field_location02':'START_BYTE2', 'field_le
ngth02':'BYTES2', 'field_format02':('FORMAT2', 'N/A'), 'field_min_logical02':'field_min_logical2', 'field_max_logical02':'field_max_logical2', 'field_value
_offset02':'START_BYTE2', 'field_unit02':'UNIT2', 'field_description02':'DESCRIPTION2', 'field_name03':'NAME3', 'field_number03':'COLUMN_NUMBER3', 'field_d
ata_type03':'DATA_TYPE3', 'field_location03':'START_BYTE3', 'field_length03':'BYTES3', 'field_format03':('FORMAT3', 'N/A'), 'field_min_logical03':'field_m
in_logical3', 'field_max_logical03':'field_max_logical3', 'field_value_offset03':'START_BYTE3', 'field_unit03':'UNIT3', 'field_description03':'DESCRIPTION3
', 'field_name04':'NAME4', 'field_number04':'COLUMN_NUMBER4', 'field_data_type04':'DATA_TYPE4', 'field_location04':'START_BYTE4', 'field_length04':'BYTES4
', 'field_format04':('FORMAT4', 'N/A'), 'field_min_logical04':'field_min_logical4', 'field_max_logical04':'field_max_logical4', 'field_value_offset04':'STAR
T_BYTE4', 'field_unit04':'UNIT4', 'field_description04':'DESCRIPTION4', 'field_name05':'NAME5', 'field_number05':'COLUMN_NUMBER5', 'field_data_type05':'DA
TA_TYPE5', 'field_location05':'START_BYTE5', 'field_length05':'BYTES5', 'field_format05':('FORMAT5', 'N/A'), 'field_min_logical05':'field_min_logical5', 'f
ield_max_logical05':'field_max_logical5', 'field_value_offset05':'START_BYTE5', 'field_unit05':'UNIT5', 'field_description05':'DESCRIPTION5', 'field_name06
':'NAME6', 'field_number06':'COLUMN_NUMBER6', 'field_data_type06':'DATA_TYPE6', 'field_location06':'START_BYTE6', 'field_length06':'BYTES6', 'field_format
06':('FORMAT6', 'N/A'), 'field_min_logical06':'field_min_logical6', 'field_max_logical06':'field_max_logical6', 'field_value_offset06':'START_BYTE6', 'fiel
d_unit06':'UNIT6', 'field_description06':'DESCRIPTION6', 'field_name07':'NAME7', 'field_number07':'COLUMN_NUMBER7', 'field_data_type07':'DATA_TYPE7', 'fie
ld_location07':'START_BYTE7', 'field_length07':'BYTES7', 'field_format07':('FORMAT7', 'N/A'), 'field_min_logical07':'field_min_logical7', 'field_max_logica
l07':'field_max_logical7', 'field_value_offset07':'START_BYTE7', 'field_unit07':'UNIT7', 'field_description07':'DESCRIPTION7', 'field_name08':'NAME8', 'fi
eld_number08':'COLUMN_NUMBER8', 'field_data_type08':'DATA_TYPE8', 'field_location08':'START_BYTE8', 'field_length08':'BYTES8', 'field_format08':('FORMAT8
', 'N/A'), 'field_min_logical08':'field_min_logical8', 'field_max_logical08':'field_max_logical8', 'field_value_offset08':'START_BYTE8', 'field_unit08':'UNI
T8', 'field_description08':'DESCRIPTION8', 'field_name09':'NAME9', 'field_number09':'COLUMN_NUMBER9', 'field_data_type09':'DATA_TYPE9', 'field_location09':
'START_BYTE9', 'field_length09':'BYTES9', 'field_format09':('FORMAT9', 'N/A'), 'field_min_logical09':'field_min_logical9', 'field_max_logical09':'field_max
_logical9', 'field_value_offset09':'START_BYTE9', 'field_unit09':'UNIT9', 'field_description09':'DESCRIPTION9', 'comment':'comment', 'PHX_object_descripti
on_HEADER1':'des_header'}
756
```

This is basically a translation table

PYTHON CAVEATS

- USE Python 2.x.x (NOT 3.x.x) *using 2.6.5-2.7.1*
 - Not backward compatible! (Two separate development trees)
 - Allows use of **LXML** Python package*
- LXML is a free package of routines that allows Python to work directly with XML (specifically the etree routine allows navigation through the XML tree structures)

LXML

- Allows generation of XML from schema while populating it with appropriate values from PDS3 label
- Written modularly to allow export to different missions/nodes
- We delete unused, *non-required* elements from the label at the end
- Allows for handling of `<local_identifier>` problem (renaming elements at the end)

MODULARIZATION

- Mission Area can be added to individual schema as needed (same for a Node Area)
- XML Input File – we've broken down some of the parts specific to individual data collections and use that as an input for the Python scripts

```
pythonxmlinput.xml x
1 <?xml version="1.0" encoding="UTF-8"?>
2 <Table_Input_File xmlns="file:/home/shannon/Documents/WORK/build1_versiond/py"
3   xmlns:ns2="file:/home/shannon/Documents/WORK/build1_versiond/ll"
4   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
5   xsi:schemaLocation="file:/home/shannon/Documents/WORK/build1_versiond/py file:/home/shannon/Documents/WORK/build1_versiond/MATIAS_PYTHON_SCHEMA.xsd">
6   <file_location>file_location/home</file_location>
7   <data_type>Raw</data_type>
8   <min_max_column_numbers>8,33,22</min_max_column_numbers>
9   <column_header>column_header0</column_header>
10  <comma_delimited>comma_delimited0</comma_delimited>
11  <comment>comment0</comment>
12 </Table_Input_File>
13
```


PYTHON ROUTINE

```
73 # This is version 1 of a script that converts a PDS3 table lable file into the equivalent PDS4 label xml file. It uses a schema made from the generic Pro-
duct Table Character schema. This schema has added extensions to the names of duplicate tag names to make them unique. It also requires a pre processing
xml file with specific information about the label file. In this file you will answer a few specific questions about the table file. The comments in this
script will start with # and end at the #### mark at the end of the block that is being described.
74 # This handles the input xml file. It asks for the file location as a raw_input. The try and except statement in this section handles an error if the fil-
e name is wrong or does not exist. It creates a dictionary with the values from the file. The xml tag is the key and the text in the tag is the value. It
is using the lxml module to parse the xml file.
75 xml = raw_input('Enter input file with path: ')
76 input_dict = {}
77 from lxml import etree
78 try:
79     context = etree.iterparse((xml))
80     print "\033[92mCONVERSION IN PROGRESS\033[0m"
81 except:
82     print "\033[91mINVALID FILE NAME OR PATH\033[0m"
83     while True:
84         xml = raw_input('Renter input file with path: ')
85         try:
86             context = etree.iterparse((xml))
87             print "\033[92mCONVERSION IN PROGRESS\033[0m"
88             break
89         except:
90             print "\033[91mINVALID FILE NAME OR PATH\033[0m"
91
92 for action, elem in context:
93     text = elem.text
94     tag = elem.tag
95     input_dict[tag] = text
96 ####
97
98 # Here the python module glob is imported. This modules allows us to search through a directory and find all the files that meet the search criteria. Wil-
d cards can be used (*). It uses the value for file_location in the input file to search for label files. It puts all these files into a list called INPU-
T.
99 import glob
100 INPUT = glob.glob(input_dict['file_location'])
101 INPUT.sort()
102 ####
103 if len(INPUT) == 0:
104     print "\033[91mTHE '<file_location>' ELEMENT IN THE INPUT FILE HAS AN INVALID PATH OR FILE NAME\033[0m"
105 ####
106
107 # The values that were entered in the min_max_column_numbers field in the xml file input file are put into a list to be used later in the script.
108 mm_list = []
109 raw_mm = input_dict['min_max_column_numbers']
110 for x in range(raw_mm.count(',')+1):
111     mm_list += [raw_mm.split(',')[x]]
```


PDS3/PDS4 LABEL COMPARISON

PDS3 Label (.LBL)

```
1 PDS_VERSION_ID          = PDS3
2
3 /* FILE DATA ELEMENTS */
4
5 RECORD_TYPE             = FIXED_LENGTH
6 RECORD_BYTES           = 97
7 FILE_RECORDS           = 36352
8
9
10
11 /* POINTERS TO DATA OBJECTS */
12
13 ^TABLE                  = "MS000EMH_00896227783_10C6M1.TAB"
14
15
16 /* IDENTIFICATION DATA ELEMENTS */
17
18 DATA_SET_ID            = "PHX-M-MET-2-PT-EDR-V1.0"
19 PRODUCT_ID              = "MS000EMH_00896227783_10C6M1"
20 PRODUCT_TYPE            = EDR
21 PRODUCT_VERSION_ID     = "V1.5 D-33236"
22 RELEASE_ID              = "0001"
23 INSTRUMENT_HOST_ID     = PHX
24 INSTRUMENT_HOST_NAME   = "PHOENIX"
25 INSTRUMENT_ID          = MET
26 INSTRUMENT_TYPE        = "IN SITU METEOROLOGY"
27 LOCAL_TRUE_SOLAR_TIME  = "17:03:01"
28 LOCAL_MEAN_SOLAR_TIME  = "16:49:31"
29 MISSION_NAME           = "PHOENIX"
30 MISSION_PHASE_NAME     = "PRIMARY MISSION"
31 PLANET_DAY_NUMBER      = 0
32 PRODUCER_INSTITUTION_NAME = "YORK UNIVERSITY"
33 PRODUCT_CREATION_TIME  = 2008-06-05T06:55:25.877
34 OPS_TOKEN               = 16#10C60000#
35 SPACECRAFT_CLOCK_CNT_PARTITION = 1
36 SPACECRAFT_CLOCK_START_COUNT = "896227783.309"
37 START_TIME              = 2008-05-26T00:08:36.308
38 STOP_TIME               = 2008-05-26T20:20:18.308
39 TARGET_NAME             = MARS
40 TARGET_TYPE             = PLANET
41
42 /* HISTORY DATA ELEMENTS */
43
44 SOFTWARE_NAME           = "MET-GDS"
45 SOFTWARE_VERSION_ID     = "3.0.5"
46 PROCESSING_HISTORY_TEXT = "CODMAC LEVEL 1 TO LEVEL 2"
47
48 /* COMMANDED PARAMETERS */
```

PDS4 XML Label (.xml)

```
1 <Product_Table_Character xsi:schemaLocation="http://pds.nasa.gov/schema/pds4/pds http://pds.nasa.gov/s
2 <Identification_Area_Product>
3   <logical_identifier>URN:NASA:PDS:PHX_MET-RAW:MS000EML_00896227783_10C6M1</logical_identifier>
4   <version_id>1.0</version_id>
5   <product_class>Product_Table_Character</product_class>
6   <title>PHOENIX MET Experiment</title>
7   <Subject_Area>
8     <target_name>MARS</target_name>
9     <instrument_name>MET</instrument_name>
10    <instrument_host_name>PHOENIX</instrument_host_name>
11  </Subject_Area>
12 </Identification_Area_Product>
13 <Cross_Reference_Area_Product>
14   <Observing_System>
15     <title>MET</title>
16     <Observing_System_Component>
17       <observing_system_component_type>Instrument</observing_system_component_type>
18       <Observing_System_Reference_Entry>
19         <lid_reference>URN:NASA:PDS:instrument.MET+PHX</lid_reference>
20         <reference_association_type>has_instrument</reference_association_type>
21       </Observing_System_Reference_Entry>
22     </Observing_System_Component>
23   </Observing_System>
24 </Cross_Reference_Area_Product>
25 <Observation_Area>
26   <start_date_time>2008-05-26T00:08:36.308</start_date_time>
27   <stop_date_time>2008-05-26T20:20:18.308</stop_date_time>
28   <spacecraft_clock_start_count>896227783.309</spacecraft_clock_start_count>
29   <spacecraft_clock_stop_count>N/A</spacecraft_clock_stop_count>
30   <Mission_Area>
31     <PHX_Local_True_Solar_Time>17:03:01</PHX_Local_True_Solar_Time>
32     <PHX_Local_Mean_Solar_Time>16:49:31</PHX_Local_Mean_Solar_Time>
33     <PHX_Mission_Phase_Name>PRIMARY MISSION</PHX_Mission_Phase_Name>
34     <PHX_Planet_Day_Number>0</PHX_Planet_Day_Number>
35     <PHX_Commanded_Parameters>
36       <PHX_Ops-Token_Activity>16#10C6#</PHX_Ops-Token_Activity>
37       <PHX_Instrument_Mode_Id>5</PHX_Instrument_Mode_Id>
38       <PHX_Detector_Id>3</PHX_Detector_Id>
39       <PHX_Pressure_Threshold>0</PHX_Pressure_Threshold>
40       <PHX_Temperature_Threshold>0</PHX_Temperature_Threshold>
41       <PHX_Period_Duration>512</PHX_Period_Duration>
42       <PHX_Period_Number>142</PHX_Period_Number>
43     </PHX_Commanded_Parameters>
44     <PHX_History_Data_Elements>
45       <PHX_Software_Name>MET-GDS</PHX_Software_Name>
46       <PHX_Software_Version_Id>3.0.5</PHX_Software_Version_Id>
47       <PHX_Processing_History>CODMAC LEVEL 1 TO LEVEL 2</PHX_Processing_History>
48     </PHX_History_Data_Elements>
49   </Mission_Area>
50 </Observation_Area>
```


PDS3/PDS4 LABEL COMPARISON

PDS3 Label (.LBL)

```
1 PDS_VERSION_ID          = PDS3
2
3 /* FILE DATA ELEMENTS */
4
5 RECORD_TYPE             = FIXED_LENGTH
6 RECORD_BYTES            = 97
7 FILE_RECORDS            = 36352
8
9
10
11 /* POINTERS TO DATA OBJECTS */
12
13 ^TABLE                  = "MS000EMH_00896227783_10C6M1.TAB"
14
15
16 /* IDENTIFICATION DATA ELEMENTS */
17
18 DATA_SET_ID            = "PHX-M-MET-2-PT-EDR-V1.0"
19 PRODUCT_ID              = "MS000EMH_00896227783_10C6M1"
20 PRODUCT_TYPE            = EDR
21 PRODUCT_VERSION_ID     = "V1.5 D-33236"
22 RELEASE_ID             = "0001"
23 INSTRUMENT_HOST_ID     = PHX
24 INSTRUMENT_HOST_NAME   = "PHOENIX"
25 INSTRUMENT_ID          = MET
26 INSTRUMENT_TYPE        = "IN SITU METEOROLOGY"
27 LOCAL_TRUE_SOLAR_TIME  = "17:03:01"
28 LOCAL_MEAN_SOLAR_TIME  = "16:49:31"
29 MISSION_NAME           = "PHOENIX"
30 MISSION_PHASE_NAME     = "PRIMARY MISSION"
31 PLANET_DAY_NUMBER      = 0
32 PRODUCER_INSTITUTION_NAME = "YORK UNIVERSITY"
33 PRODUCT_CREATION_TIME  = 2008-06-05T06:55:25.877
34 OPS_TOKEN              = 16#10C60000#
35 SPACECRAFT_CLOCK_CNT_PARTITION = 1
36 SPACECRAFT_CLOCK_START_COUNT = "896227783_309"
37 START_TIME             = 2008-05-26T00:08:36.308
38 STOP_TIME              = 2008-05-26T20:20:18.308
39 TARGET_NAME            = MARS
40 TARGET_TYPE            = PLANET
41
42 /* HISTORY DATA ELEMENTS */
43
44 SOFTWARE_NAME          = "MET-GDS"
45 SOFTWARE_VERSION_ID    = "3.0.5"
46 PROCESSING_HISTORY_TEXT = "CODMAC LEVEL 1 TO LEVEL 2"
47
48 /* COMMANDED PARAMETERS */
```

PDS4 XML Label (.xml)

```
1 <Product_Table_Character xsi:schemaLocation="http://pds.nasa.gov/schema/pds4/pds http://pds.nasa.gov/s
2 <Identification_Area_Product>
3   <logical_identifier>URN:NASA:PDS:PHX_MET-RAW:MS000EML_00896227783_10C6M1</logical_identifier>
4   <version_id>1.0</version_id>
5   <product_class>Product_Table_Character</product_class>
6   <title>PHOENIX MET Experiment</title>
7   <Subject_Area>
8     <target_name>MARS</target_name>
9     <instrument_name>MET</instrument_name>
10    <instrument_host_name>PHOENIX</instrument_host_name>
11  </Subject_Area>
12 </Identification_Area_Product>
13 <Cross_Reference_Area_Product>
14   <Observing_System>
15     <title>MET</title>
16     <Observing_System_Component>
17       <observing_system_component_type>Instrument</observing_system_component_type>
18       <Observing_System_Reference_Entry>
19         <lid_reference>URN:NASA:PDS:instrument.MET+PHX</lid_reference>
20         <reference_association_type>has_instrument</reference_association_type>
21       </Observing_System_Reference_Entry>
22     </Observing_System_Component>
23   </Observing_System>
24 </Cross_Reference_Area_Product>
25 <Observation_Area>
26   <start_date_time>2008-05-26T00:08:36.308</start_date_time>
27   <stop_date_time>2008-05-26T20:20:18.308</stop_date_time>
28   <spacecraft_clock_start_count>896227783.309</spacecraft_clock_start_count>
29   <spacecraft_clock_stop_count>N/A</spacecraft_clock_stop_count>
30   <Mission_Area>
31     <PHX_Local_True_Solar_Time>17:03:01</PHX_Local_True_Solar_Time>
32     <PHX_Local_Mean_Solar_Time>16:49:31</PHX_Local_Mean_Solar_Time>
33     <PHX_Mission_Phase_Name>PRIMARY MISSION</PHX_Mission_Phase_Name>
34     <PHX_Planet_Day_Number>0</PHX_Planet_Day_Number>
35     <PHX_Commanded_Parameters>
36       <PHX_Ops-Token_Activity>16#10C6#</PHX_Ops-Token_Activity>
37       <PHX_Instrument_Mode_Id>5</PHX_Instrument_Mode_Id>
38       <PHX_Detector_Id>3</PHX_Detector_Id>
39       <PHX_Pressure_Threshold>0</PHX_Pressure_Threshold>
40       <PHX_Temperature_Threshold>0</PHX_Temperature_Threshold>
41       <PHX_Period_Duration>512</PHX_Period_Duration>
42       <PHX_Period_Number>142</PHX_Period_Number>
43     </PHX_Commanded_Parameters>
44     <PHX_History_Data_Elements>
45       <PHX_Software_Name>MET-GDS</PHX_Software_Name>
46       <PHX_Software_Version_Id>3.0.5</PHX_Software_Version_Id>
47       <PHX_Processing_History>CODMAC LEVEL 1 TO LEVEL 2</PHX_Processing_History>
48     </PHX_History_Data_Elements>
49   </Mission_Area>
50 </Observation_Area>
```


DOCUMENT FILES

- Once we have the working schema producing XML labels for the data, the rest of the Bundle needs to be processed
- Set up schemas for rest of the document types to migrate into XML Bundle (*archive bundle*)
 - PDF, text, etc.
- Organization needs to be different for the archive bundle than in PDS3

DOCUMENT FILES

- Currently, our document schemas are previous version *c* and are in the process of being upgraded (to *d* or higher)
- Similar process to data files, we have Python scripts to handle the migration

**NOTE:

- Bundle architecture is the biggest change from PDS3 to PDS4
- ➔ • Finishing the Bundle needs multiple specific schemas and this can be *very time intensive*

ARCHIVE BUNDLE

PDS4 Organization

17 FEB 2011

PHOENIX EXAMPLE

PHOENIX (Atmospheres)

5 Instrument Bundles

Phoenix Mission Data Dictionary (Geosciences)

Bundle_MET

INSTRUMENT (MET) <i>INVENTORY_MET.xml/.TAB</i>
README
Collection_context PDS3 Catalog Files Instrument_host Instrument personnel
Collection_document PDS3 Documents Index SIS Files ERRATA
Collection_data_raw Sol directories .XML, .TAB
Collection_data_reduced Sol directories .XML, .TAB
Collection_schema Specific Schema (.xsd)

Bundle_LIDAR

INSTRUMENT (LIDAR) <i>INVENTORY_LIDAR.xml/.TAB</i>
README
Collection_context PDS3 Catalog Files Instrument_host Instrument personnel
Collection_document PDS3 Documents Index SIS Files ERRATA
Collection_data_raw Sol directories .XML, .TAB
Collection_data_reduced Sol directories .XML, .TAB
Collection_schema Specific Schema (.xsd)

Bundle_ASE

INSTRUMENT (ASE) <i>INVENTORY_ASE.xml/.TAB</i>
README
Collection_context PDS3 Catalog Files Instrument_host Instrument personnel
Collection_document PDS3 Documents Index SIS Files ERRATA
Collection_data_raw XML, .TAB
Collection_data_reduced XML, .TAB
Collection_schema Specific Schema (.xsd)

Bundle_AO

INSTRUMENT (AO) <i>INVENTORY_AO.xml/.TAB</i>
README
Collection_context PDS3 Catalog Files Instrument_host Instrument personnel
Collection_document PDS3 Documents Index SIS Files ERRATA
Collection_data_derived .XML, .TAB
Collection_schema Specific Schema (.xsd)

Bundle_TT

INSTRUMENT (TT) <i>INVENTORY_TT.xml/.TAB</i>
README
Collection_context PDS3 Catalog Files Instrument_host Instrument personnel
Collection_document PDS3 Documents Index SIS Files ERRATA
Collection_data_derived .XML, .TAB
Collection_schema Specific Schema (.xsd)

REGISTRATION & HARVEST

- Local Node registration...
 - Registry working at NMSU*
 - Each Node will probably have this setup at some point
- Sean harvests...
 - Either directly from the data OR from local registry

CONCLUSION

- For the ATMOS, this process should be relatively straightforward at this point
- Python scripts are being modified for Table_Binary now and FITS next
- Planning modular approach to the Python to lessen the prep time for each “new” dataset
 - Unclear how reusable the ancillary schemas will be
 - These Python adaptations could be useful for the rest of PDS...

QUESTIONS/COMMENTS