TRANSITIONS:
PDS3 TO PDS4

Lynn D.V. Neakrase & Matias Roybal
Atmospheres Node

Shannon Rees, Lyle Huber

OVERVIEW

- General description of the challenge
- Organization strategies
- XML schemas and Python scripting
- Comments, problems, concerns

OVERVIEW

- General description of the challenge
- Organization strategies
- XML schemas and Python scripting
- Comments, problems, concerns

THE CHALLENGE

- *Migration*
  - Move existing data into the new system
    - Modular approach eventually resulting in PDS4 compliant labels for all products
- *Generating New Data (Future Data)*
  - Working with teams for upcoming missions to archive in PDS4 from the outset

MIGRATION

- Data migration from PDS3 to PDS4 is complicated
  - ODL and XML work differently
    - Parsing PDS4-appropriate values is non trivial
      - PDS3 values, concatenation, entirely new fields
- Several approaches *based on* ***data organization***
  - ***Modular approach*** using Python programming: *general coding with modules for specific purposes*

NEW DATA

- Pipeline process will have to be more interactive
  - Nodes will provide tailored XML schemas for products
    - Mission additions can be added by science teams
- Currently working with upcoming missions to develop products with PDS4 compliant structures
  - LADEE *(ASCII tables: Product_Table_Character)*
  - MAVEN *(ASCII tables: Product_Table_Character)*

EFFORTS

- *Migration*
  - Primary goal for the first several years (now to ~2013)
  - Also serves as test grounds for more complicated data issues
- *Generating New Data*
  - First new data to be ingested after 2013 preferably after large quantities are migrated into the new system
  - Use of migration efforts to tailor data provider experience

OVERVIEW

- General description of the challenge
- **Organization strategies**
- XML schemas and Python scripting
- Comments, problems, concerns

ORGANIZATION - *MIGRATION*

- *"Mise en place"* - need to prepare for the migration
  - ***Specific schema*** (instrument/mission) – tailored from generics, with mission specific material ready for XML labels
  - ***Translating*** metadata from 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 - *MIGRATION*

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

ORGANIZATION - *MIGRATION*

- 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)

ORGANIZATION - *NEW DATA*

- Using migration code to facilitate setting up data pipelines same work effort — 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)

OVERVIEW

- General description of the challenge
- Organization strategies
- XML schemas and Python scripting
- Comments, problems, concerns

XML SCHEMAS

- The basic templates for creating PDS4 labels and documents
- Serve as a blueprint for translating the archive model into usable pieces
- ***Specific Schemas*** are derived from generic/tailored schemas, ready to create labels
  - These include Mission and Node specific additions for a given data type

MODULARIZATION

- ***Python Migration Code*** allows different types of data to be handled using the same base code

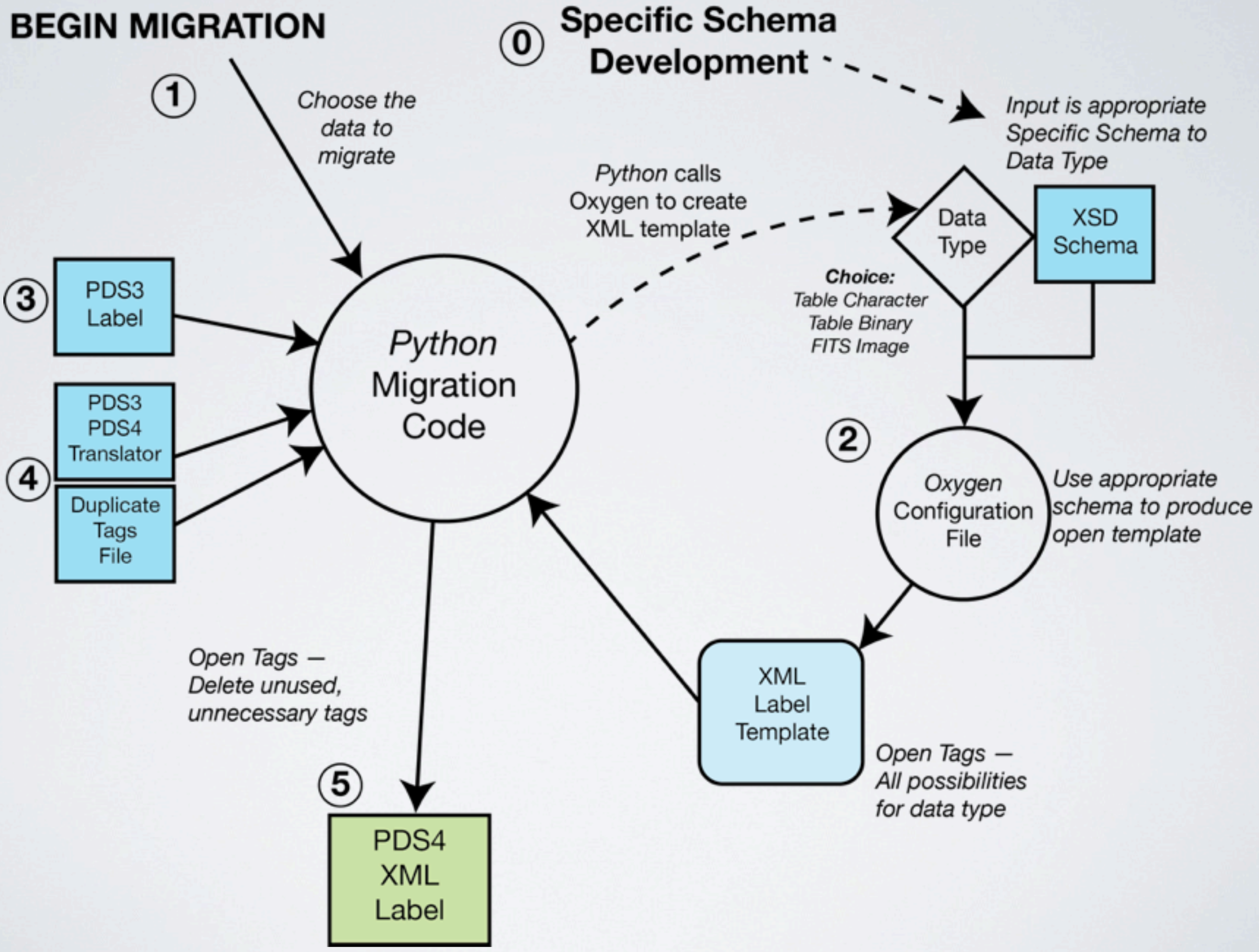
- Requires **Input Files** – **1)** *PDS3 Label* **2)** *PDS3-PDS4 Translation Table* **3)** *Duplicates Tags File* **4)** *Specific Schema*

e.g., Location Info for Duplicate Tags:
title == Identification_Area_Product
title == Observing_System

- Pointing to these files can be handled with path designations within the Python code

- Potential for handling this in a GUI or a config file input

General Walkthrough of Migration



PDS3/PDS4 LABEL COMPARISON

PDS3 Label (.LBL)

```
PDS_VERSION_ID                  = PDS3

/* FILE DATA ELEMENTS */

RECORD_TYPE                     = FIXED_LENGTH
RECORD_BYTES                    = 97
FILE_RECORDS                    = 36352


/* POINTERS TO DATA OBJECTS */

^TABLE                          = "MS000EMH_00896227783_10C6M1.TAB"

/* IDENTIFICATION DATA ELEMENTS */

DATA_SET_ID                     = "PHX-M-MET-2-PT-EDR-V1.0"
PRODUCT_ID                      = "MS000EMH_00896227783_10C6M1"
PRODUCT_TYPE                    = EDR
PRODUCT_VERSION_ID              = "V1.5 D-33236"
RELEASE_ID                      = "0001"
INSTRUMENT_HOST_ID              = PHX
INSTRUMENT_HOST_NAME            = "PHOENIX"
INSTRUMENT_ID                   = MET
INSTRUMENT_TYPE                 = "IN SITU METEOROLOGY"
LOCAL_TRUE_SOLAR_TIME           = "17:03:01"
LOCAL_MEAN_SOLAR_TIME           = "16:49:31"
MISSION_NAME                    = "PHOENIX"
MISSION_PHASE_NAME              = "PRIMARY MISSION"
PLANET_DAY_NUMBER               = 0
PRODUCER_INSTITUTION_NAME       = "YORK UNIVERSITY"
PRODUCT_CREATION_TIME           = 2008-06-05T06:55:25.877
OPS_TOKEN                       = 16#10C60000#
SPACECRAFT_CLOCK_CNT_PARTITION  = 1
SPACECRAFT_CLOCK_START_COUNT    = "896227783.309"
START_TIME                      = 2008-05-26T00:08:36.308
STOP_TIME                       = 2008-05-26T20:20:18.308
TARGET_NAME                     = MARS
TARGET_TYPE                     = PLANET

/* HISTORY DATA ELEMENTS */

SOFTWARE_NAME                   = "MET-GDS"
SOFTWARE_VERSION_ID             = "3.0.5"
PROCESSING_HISTORY_TEXT         = "CODMAC LEVEL 1 TO LEVEL 2"

/* COMMANDED PARAMETERS */
```

PDS4 Label (.xml)

PHOENIX EXAMPLE

```
<Product_Table_Character xsi:schemaLocation="http://pds.nasa.gov/schema/pds4/pds http://pds.nasa.gov/s
  <Identification_Area_Product>
    <logical_identifier>URN:NASA:PDS:PHX_MET-RAW:MS000EML_00896227783_10C6M1</logical_identifier>
    <version_id>1.0</version_id>
    <product_class>Product_Table_Character</product_class>
    <title>PHOENIX MET Experiment</title>
    <Subject_Area>
      <target_name>MARS</target_name>
      <instrument_name>MET</instrument_name>
      <instrument_host_name>PHOENIX</instrument_host_name>
    </Subject_Area>
  </Identification_Area_Product>
  <Cross_Reference_Area_Product>
    <Observing_System>
      <title>MET</title>
      <Observing_System_Component>
        <observing_system_component_type>Instrument</observing_system_component_type>
        <Observing_System_Reference_Entry>
          <lid_reference>URN:NASA:PDS:instrument.MET+PHX</lid_reference>
          <reference_association_type>has_instrument</reference_association_type>
        </Observing_System_Reference_Entry>
      </Observing_System_Component>
    </Observing_System>
  </Cross_Reference_Area_Product>
  <Observation_Area>
    <start_date_time>2008-05-26T00:08:36.308</start_date_time>
    <stop_date_time>2008-05-26T20:20:18.308</stop_date_time>
    <spacecraft_clock_start_count>896227783.309</spacecraft_clock_start_count>
    <spacecraft_clock_stop_count>N/A</spacecraft_clock_stop_count>
    <Mission_Area>
      <PHX_Local_True_Solar_Time>17:03:01</PHX_Local_True_Solar_Time>
      <PHX_Local_Mean_Solar_Time>16:49:31</PHX_Local_Mean_Solar_Time>
      <PHX_Mission_Phase_Name>PRIMARY MISSION</PHX_Mission_Phase_Name>
      <PHX_Planet_Day_Number>0</PHX_Planet_Day_Number>
      <PHX_Commanded_Parameters>
        <PHX_Ops_Token_Activity>16#10C6#</PHX_Ops_Token_Activity>
        <PHX_Instrument_Mode_Id>5</PHX_Instrument_Mode_Id>
        <PHX_Detector_Id>3</PHX_Detector_Id>
        <PHX_Pressure_Threshold>0</PHX_Pressure_Threshold>
        <PHX_Temperature_Threshold>0</PHX_Temperature_Threshold>
        <PHX_Period_Duration>512</PHX_Period_Duration>
        <PHX_Period_Number>142</PHX_Period_Number>
      </PHX_Commanded_Parameters>
      <PHX_History_Data_Elements>
        <PHX_Software_Name>MET-GDS</PHX_Software_Name>
        <PHX_Software_Version_Id>3.0.5</PHX_Software_Version_Id>
        <PHX_Processing_History>CODMAC LEVEL 1 TO LEVEL 2</PHX_Processing_History>
      </PHX_History_Data_Elements>
    </Mission_Area>
  </Observation_Area>
```

PDS3/PDS4 LABEL COMPARISON

PDS3 Label (.LBL)

```
PDS_VERSION_ID                  = PDS3

/* FILE DATA ELEMENTS */

RECORD_TYPE                     = FIXED_LENGTH
RECORD_BYTES                    = 97
FILE_RECORDS                    = 36352


/* POINTERS TO DATA OBJECTS */

^TABLE                          = "MS000EMH_00896227783_10C6M1.TAB"

/* IDENTIFICATION DATA ELEMENTS */

DATA_SET_ID                     = "PHX-M-MET-2-PT-EDR-V1.0"
PRODUCT_ID                      = "MS000EMH_00896227783_10C6M1"
PRODUCT_TYPE                    = EDR
PRODUCT_VERSION_ID              = "V1.5 D-33236"
RELEASE_ID                      = "0001"
INSTRUMENT_HOST_ID              = PHX
INSTRUMENT_HOST_NAME            = "PHOENIX"
INSTRUMENT_ID                   = MET
INSTRUMENT_TYPE                 = "IN SITU METEOROLOGY"
LOCAL_TRUE_SOLAR_TIME           = "17:03:01"
LOCAL_MEAN_SOLAR_TIME           = "16:49:31"
MISSION_NAME                    = "PHOENIX"
MISSION_PHASE_NAME              = "PRIMARY MISSION"
PLANET_DAY_NUMBER               = 0
PRODUCER_INSTITUTION_NAME       = "YORK UNIVERSITY"
PRODUCT_CREATION_TIME           = 2008-06-05T06:55:25.877
OPS_TOKEN                       = 16#10C60000#
SPACECRAFT_CLOCK_CNT_PARTITION  = 1
SPACECRAFT_CLOCK_START_COUNT    = "896227783.309"
START_TIME                      = 2008-05-26T00:08:36.308
STOP_TIME                       = 2008-05-26T20:20:18.308
TARGET_NAME                     = MARS
TARGET_TYPE                     = PLANET

/* HISTORY DATA ELEMENTS */

SOFTWARE_NAME                   = "MET-GDS"
SOFTWARE_VERSION_ID             = "3.0.5"
PROCESSING_HISTORY_TEXT         = "CODMAC LEVEL 1 TO LEVEL 2"

/* COMMANDED PARAMETERS */
```

*PDS4 Label (.xml)**

PHOENIX EXAMPLE

```
<Product_Table_Character xsi:schemaLocation="http://pds.nasa.gov/schema/pds4/pd http://pds.nasa.gov/s
  <Identification_Area_Product>
    <logical_identifier>URN:NASA:PDS:PHX_MET-RAW:MS000EML_00896227783_10C6M1</logical_identifier>
    <version_id>1.0</version_id>
    <product_class>Product_Table_Character</product_class>
    <title>PHOENIX MET Experiment</title>
    <Subject_Area>
      <target_name>MARS</target_name>
      <instrument_name>MET</instrument_name>
      <instrument_host_name>PHOENIX</instrument_host_name>
    </Subject_Area>
  </Identification_Area_Product>
  <Cross_Reference_Area_Product>
    <Observing_System>
      <title>MET</title>
      <Observing_System_Component>
        <observing_system_component_type>Instrument</observing_system_component_type>
        <Observing_System_Reference_Entry>
          <lid_reference>URN:NASA:PDS:instrument.MET+PHX</lid_reference>
          <reference_association_type>has_instrument</reference_association_type>
        </Observing_System_Reference_Entry>
      </Observing_System_Component>
    </Observing_System>
  </Cross_Reference_Area_Product>
  <Observation_Area>
    <start_date_time>2008-05-26T00:08:36.308</start_date_time>
    <stop_date_time>2008-05-26T20:20:18.308</stop_date_time>
    <spacecraft_clock_start_count>896227783.309</spacecraft_clock_start_count>
    <spacecraft_clock_stop_count>N/A</spacecraft_clock_stop_count>
    <Mission_Area>
      <PHX_Local_True_Solar_Time>17:03:01</PHX_Local_True_Solar_Time>
      <PHX_Local_Mean_Solar_Time>16:49:31</PHX_Local_Mean_Solar_Time>
      <PHX_Mission_Phase_Name>PRIMARY MISSION</PHX_Mission_Phase_Name>
      <PHX_Planet_Day_Number>0</PHX_Planet_Day_Number>
      <PHX_Commanded_Parameters>
        <PHX_Ops_Token_Activity>16#10C6#</PHX_Ops_Token_Activity>
        <PHX_Instrument_Mode_Id>5</PHX_Instrument_Mode_Id>
        <PHX_Detector_Id>3</PHX_Detector_Id>
        <PHX_Pressure_Threshold>0</PHX_Pressure_Threshold>
        <PHX_Temperature_Threshold>0</PHX_Temperature_Threshold>
        <PHX_Period_Duration>512</PHX_Period_Duration>
        <PHX_Period_Number>142</PHX_Period_Number>
      </PHX_Commanded_Parameters>
      <PHX_History_Data_Elements>
        <PHX_Software_Name>MET-GDS</PHX_Software_Name>
        <PHX_Software_Version_Id>3.0.5</PHX_Software_Version_Id>
        <PHX_Processing_History>CODMAC LEVEL 1 TO LEVEL 2</PHX_Processing_History>
      </PHX_History_Data_Elements>
    </Mission_Area>
  </Observation_Area>
```

PHOENIX EXAMPLE

PHOENIX (Atmospheres)

5 Instrument Bundles

Phoenix Mission Data Dictionary (Geosciences)

Bundle_MET

INSTRUMENT (MET)
INVENTORY_MET.xml
README.txt

Collection_context
- *context.xml*
- Instrument_host
- Instrument
- Investigation

Collection_document
- *document.xml*
- PDS3 Documents
- Index
- ERRATA

Collection_data_raw
- *data_raw.xml*
- Sol directories
  - .XML, .TAB

Collection_data_reduced
- *data_reduced.xml*
- Sol directories
  - .XML, .TAB

Collection_schema
- *schema.xml*
- Specific Schema (.xsd)

Bundle_LIDAR

INSTRUMENT (LIDAR)
INVENTORY_LIDAR.xml
README.txt

Collection_context
- *context.xml*
- Instrument_host
- Instrument
- Investigation

Collection_document
- *document.xml*
- PDS3 Documents
- Index
- ERRATA

Collection_data_raw
- *data_raw.xml*
- Sol directories
  - .XML, .TAB

Collection_data_reduced
- *data_reduced.xml*
- Sol directories
  - .XML, .TAB

Collection_schema
- *schema.xml*
- Specific Schema (.xsd)

Bundle_ASE

INSTRUMENT (ASE)
INVENTORY_ASE.xml
README.txt

Collection_context
- *context.xml*
- Instrument_host
- Instrument
- Investigation

Collection_document
- *document.xml*
- PDS3 Documents
- Index
- ERRATA

Collection_data_raw
- *data_raw.xml*
- XML, .TAB

Collection_data_reduced
- *data_reduced.xml*
- XML, .TAB

Collection_schema
- *schema.xml*
- Specific Schema (.xsd)

Bundle_AO

INSTRUMENT (AO)
INVENTORY_AO.xml
README.txt

Collection_context
- *context.xml*
- Instrument_host
- Instrument
- Investigation

Collection_document
- *document.xml*
- PDS3 Documents
- Index
- ERRATA

Collection_data_derived
- *data_derived.xml*
- .XML, .TAB

Collection_schema
- *schema.xml*
- Specific Schema (.xsd)

Bundle_TT

INSTRUMENT (TT)
INVENTORY_TT.xml
README.txt

Collection_context
- *context.xml*
- Instrument_host
- Instrument
- Investigation

Collection_document
- *document.xml*
- PDS3 Documents
- Index
- ERRATA

Collection_data_derived
- *data_derived.xml*
- .XML, .TAB

Collection_schema
- *schema.xml*
- Specific Schema (.xsd)

REGISTRATION & HARVEST

- Local Node registration...
  - Registry working at NMSU*
    - Each Node should have local registries by the time we are all migrating data
- Sean harvests...
  - Either directly from the data OR from local registry

* *Needs to be updated, haven't registered the e-versions yet*

CONCLUSION

- PDS4 is working, and we are making fewer adjustments at this point
- Atmos Python scripts are currently useful for ***Table_Character***, ***Table_Binary***, and ***FITS Images*** *e-version*
- The modular approach to the Python code should lessen the prep time for each "new" dataset
- These different routines have been grouped into a downloadable ***PDS Python library*** that should work as a plugin

CONCLUSION

- Planning the transition for upcoming new missions
  - LADEE, MAVEN *Predominantly Table_Character files*

- Python code could be adapted to produce labels with proper input files
  - Instead of migrating PDS3 labels, could be retooled to create labels from data coming in for the mission
  - Started the planning process for these missions

QUESTIONS/COMMENTS

BACKUP SLIDES

GENERIC XML SCHEMA

BREAKDOWN

Generic_Type Declaration [1]

Identification_Area_Product [1]

Subject_Area [1]

Name_Resolution [0..*]

Cross_Reference_Area_Product [1]

Bibliographic_Reference [0..*]

Observing_System [1..*]

Product_Reference_Entry [0..*]

Observation_Area [1]

Mission_Area [0..*]

Node_Area [0..*]

File_Area_Observational [1]

File [1]

Data_Type [0..*]

Table_Character
Table_Binary
Array_2D_Image (FITS Image)

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 pre-tailored schemas are also provided by the PDS (edited by discipline nodes)

    - *Product_Table_Character, Product_Image_Grayscale,*

TAILORED XML SCHEMA

BREAKDOWN

Table_Character_Type Declaration [1]

added information to generic

Identification_Area_Product [1]

Subject_Area [1]

Name_Resolution [0..*]

Cross_Reference_Area_Product [1]

Bibliographic_Reference [0..*]

Observing_System [1..*]

Product_Reference_Entry [0..*]

Observation_Area [1]

Mission_Area [0..*]

Node_Area [0..*]

File_Area_Observational [1]

File [1]

Table_Character [0..*]

Specifics for Table_Character...

specific to Table_Character

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

PHX = *5 Instruments:*
MET, LIDAR, ASE, AO, TT

- 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_Product [1]

Subject_Area [1]

Name_Resolution [0..*]

Cross_Reference_Area_Product [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_Observational [1]

File [1]

Table_Character [1] — *Only one table per file for this mission*

Specifics for Table_Character...

PHX_Product_Table_Character

PDS3 METADATA

PDS3 Label

```
PDS_VERSION_ID                  = PDS3

/* FILE DATA ELEMENTS */

RECORD_TYPE                     = FIXED_LENGTH
RECORD_BYTES                    = 97
FILE_RECORDS                    = 36352


/* POINTERS TO DATA OBJECTS */

^TABLE                          = "MS000EMH_00896227783_10C6M1.TAB"

/* IDENTIFICATION DATA ELEMENTS */

DATA_SET_ID                     = "PHX-M-MET-2-PT-EDR-V1.0"
PRODUCT_ID                      = "MS000EMH_00896227783_10C6M1"
PRODUCT_TYPE                    = EDR
PRODUCT_VERSION_ID              = "V1.5 D-33236"
RELEASE_ID                      = "0001"
INSTRUMENT_HOST_ID              = PHX
INSTRUMENT_HOST_NAME            = "PHOENIX"
INSTRUMENT_ID                   = MET
INSTRUMENT_TYPE                 = "IN SITU METEOROLOGY"
LOCAL_TRUE_SOLAR_TIME           = "17:03:01"
LOCAL_MEAN_SOLAR_TIME           = "16:49:31"
MISSION_NAME                    = "PHOENIX"
MISSION_PHASE_NAME              = "PRIMARY MISSION"
PLANET_DAY_NUMBER               = 0
PRODUCER_INSTITUTION_NAME       = "YORK UNIVERSITY"
PRODUCT_CREATION_TIME           = 2008-06-05T06:55:25.877
OPS_TOKEN                       = 16#10C60000#
SPACECRAFT_CLOCK_CNT_PARTITION  = 1
SPACECRAFT_CLOCK_START_COUNT    = "896227783.309"
START_TIME                      = 2008-05-26T00:08:36.308
```

PDS3 Relevant Information

Relevant data for PDS4 labels

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

etc.

PYTHON SCRIPTING

- We take PDS3 labels, pull translatable values out of the label dropping all non-essential pieces and place it in an external file
- The external file can be handled directly within the Python code we're using to migrate the data
  - File serves to populate a Python "dictionary", that serves as a translation table

pds4_tag == *pds3_keyword*

<data_set_name> == DATA_SET_ID

LXML

Shareware Python Routine Library

- Adds functionality to Python
- Allows generation of XML from schema while populating it with appropriate values from PDS3 label
- Allows direct manipulation of XML tags within Python code
- **Python code is the version 2.x.x branch