



# **PDS4 Training Course**

---

**MODULE 1**

**INTRODUCTION TO PDS4**

# Welcome

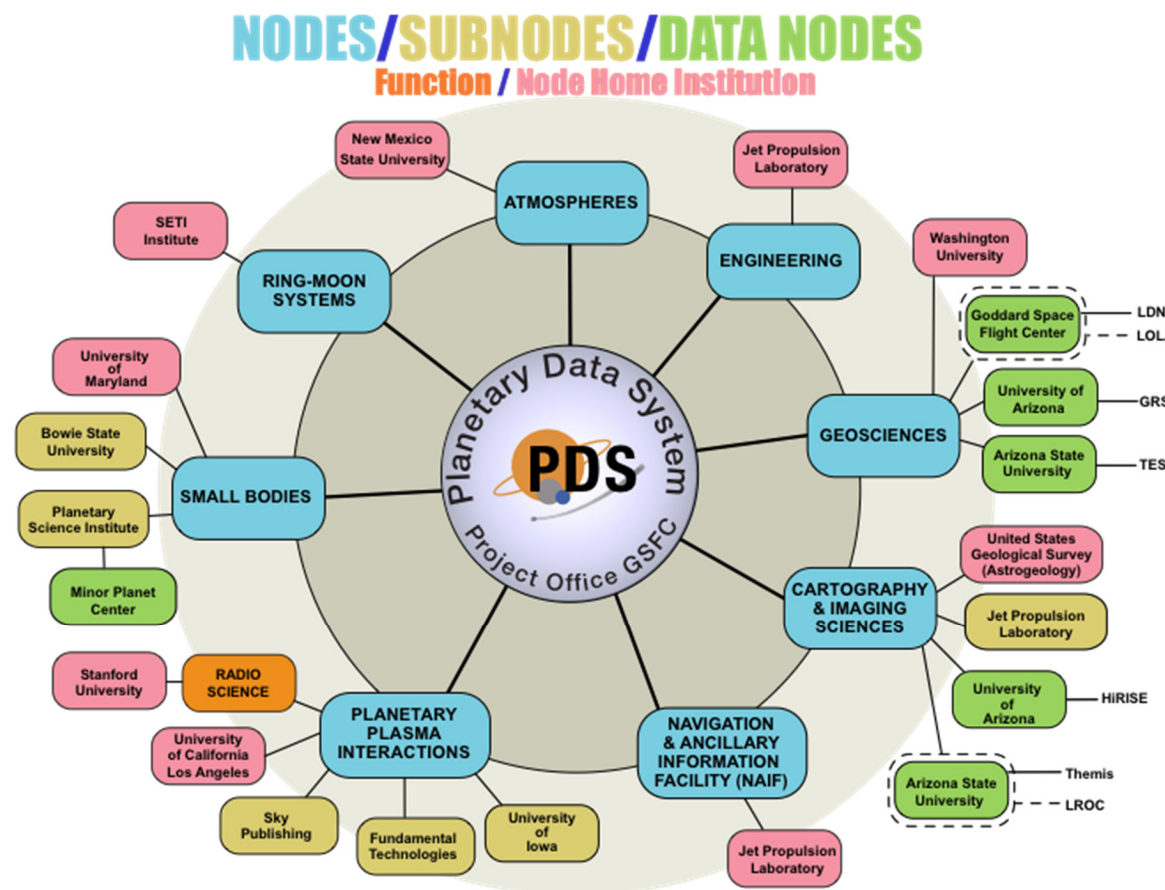
---

Welcome to the PDS4 Training Course at the DPS Fall Meeting

- Please take a thumb drive for all of today's materials
- Your instructor today is **Lynn Neakrase** from the PDS Atmospheres Node
- The first part of this course will be an introduction to PDS4 with terminology and concepts necessary for understanding the basics of the new archive standard.
- The second part of this course will present a test exercise for walking you through the initial steps of setting up a PDS4 Bundle for a simple example to give you some hands-on experience with PDS4 labels and hierarchical archive structures.
- All materials used in today's course are available on your PDS Thumb Drives.

# What is the PDS?

- PDS is NASA's Planetary Data System that is responsible for archiving planetary mission data, derived data, and supplementary analog, field, and laboratory data.
- PDS is a federation of distributed discipline nodes located all over the United States.



# What is PDS4?

---

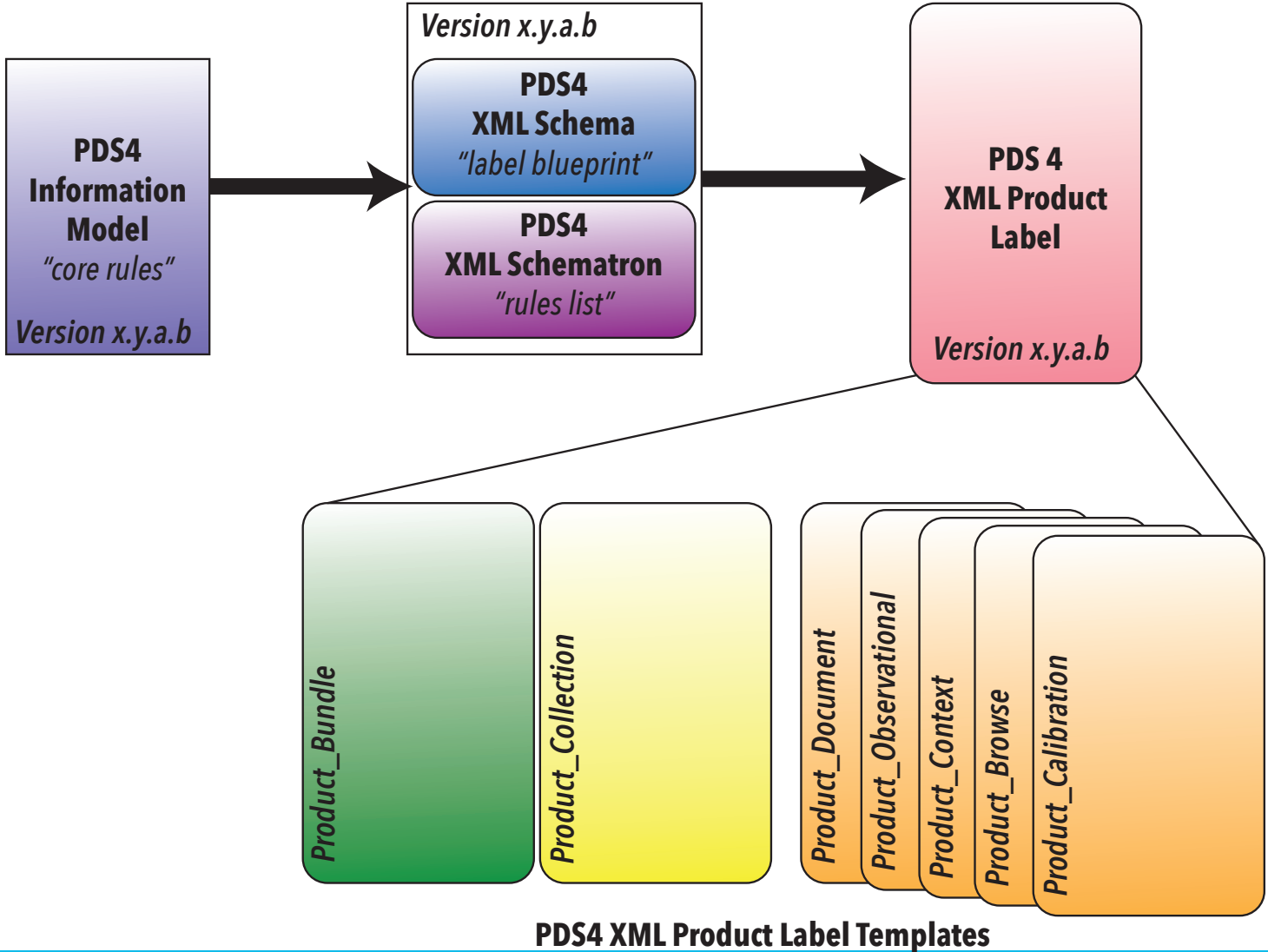
- PDS4 is the latest **Archive Standard** used to organize and register your data so that it can be used by the public and planetary science community. PDS4 is **not** a format!
- PDS4 allows metadata to be organized into product labels which sit with your data and can be used to access the data more efficiently and effectively.
- PDS4 at its heart is a formal **Information Model (IM)** that dictates all the rules and regulations for everything that is acceptable in the archive. The IM can be expressed in many ways to produce labels – PDS uses the international standard of **eXtensible Markup Language (XML)** to express the rules and regulations. XML is similar to HTML as a markup language and many text editors are *XML-Aware* and preserve the indented look of XML. Browsers are typically also *XML-Aware* but some (like Safari) may require a plug-in to display properly.

# PDS4 Structures

---

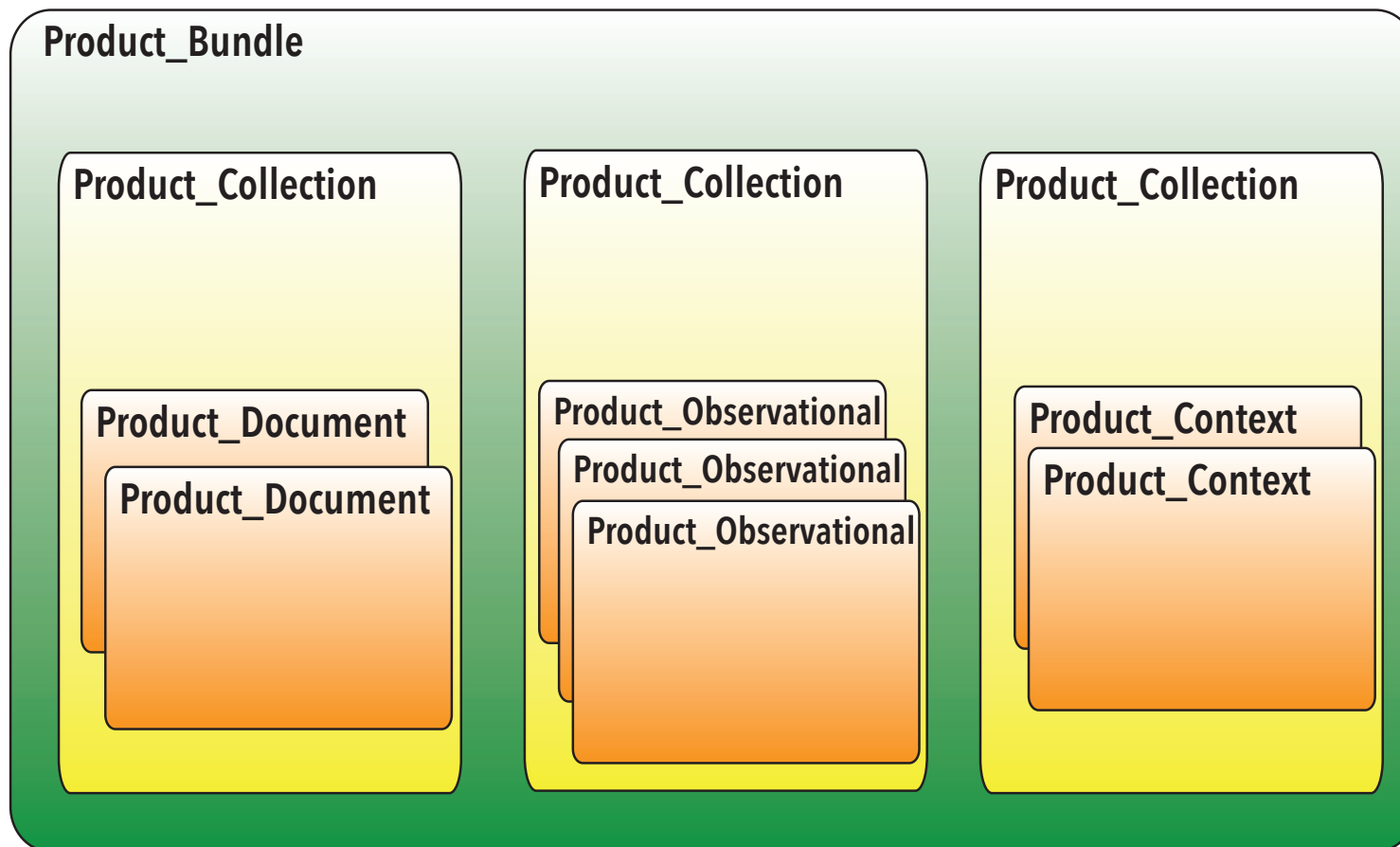
- The PDS4 Information Model (“Core Rules”) expressed as PDS4 Schema and Schematron files
  - PDS4 Schema (“Label Blueprints”) provide a Master File that comprises everything you can do in PDS4.
  - Schematron Files (“Rules Lists”) contain validation rules for the content of most of the fields in a PDS4 label allowing specific ranges and enumerations in values or format restrictions to be required for certain fields.
  - TOGETHER these two files provide the basis for a “Master” PDS4 Label. Nodes can then help to provide templates for specific products such as documents, data files, or the bundles and collections themselves.
- Everything in PDS4 is a product.
  - Throughout this course you will hear everything being referred to as products. All products have labels. Whether this is a Bundle, Collection, or any host of common documents or data products, the PDS4 Label contains metadata used to understand the content of the corresponding product file.
  - All labels in PDS4 are implemented in XML. At first, XML can be confusing because it is a *markup language*, similar to HTML used in webpage design. Typically values for certain fields will be placed between tags : `<tag>Value</tag>`
  - One of the most important fields in any label is the `<logical_identifier>`. **Logical Identifiers (LIDs)** are strings that provide a unique ID for any given product so that it can be searched upon and found within the **PDS Central Registry**.
    - LIDs are constructed as Uniform Resource Names (URNs), which are locaters used to register the products in the Central Registry.

# Basic Structure of PDS4



Everything is a product.  
Products have labels.

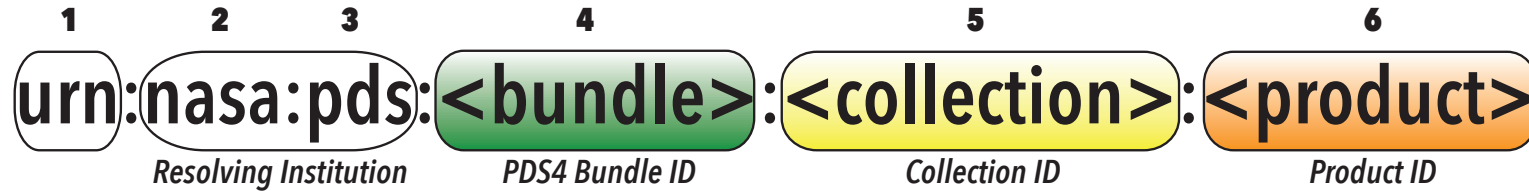
## ***PDS4 organization is arranged in Bundles, Collections, and Basic Products***



*Hypothetical Example Bundle*

Internal Referencing is accomplished via **Logical Identifiers (LIDs)**  
 Logical Identifiers use **Uniform Resource Names (URNs)**

*Uniform Resource Name*



URNs are one type of Uniform Resource Identifier (URI) used in computing.  
 A URI is any string of characters used to identify or name a resource uniquely and persistently.  
 URIs specify the type of access protocol used to retrieve files and may be location dependent or independent.

### Location Dependent URI Examples

*Uniform Resource Locators (URLs)*

Hypertext Transfer Protocol	(http:)
File Transfer Protocol	(ftp:)

### Location Independent URI Examples

Uniform Resource Name	(urn:)
Tag URI Scheme	(tag:)

For URNs the physical location of the file is not identified in the string.  
 Segments 2 & 3 of the URN designate the entity responsible for resolving the location.

**Examples:**      **ESA(PSA) urn:esa:psa:**

**JAXA(DARTS) urn:jaxa:darts:**

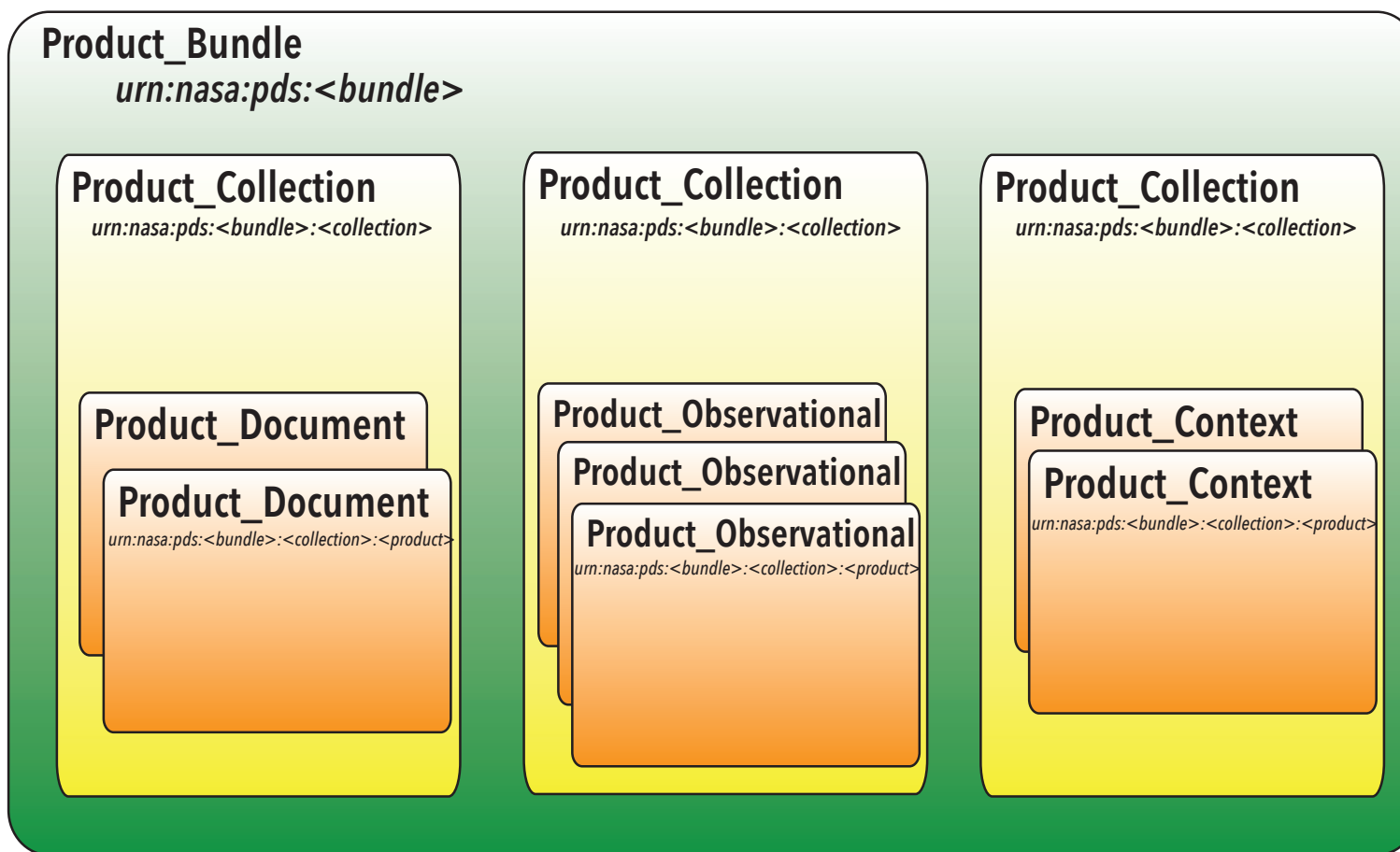


Everything is a product.  
Products have labels.

*PDS4 organization is arranged in  
Bundles, Collections, and Basic Products*



Logical Identifiers (URNs) designate components of the Bundle according to the segments (4,5, & 6).



*Hypothetical Example Bundle*

# PDS4 Structures (cont.)

---

- Bundle files contain officially one file – its XML label.
  - The **Bundle Label** contains a list of its member collections (LIDs) describing what kind of collection each is (relational information).
  - The Bundle will also usually contain the member collections as subdirectories and may also contain an optional readme.txt file.
  - Bundle files must be named in the form: *bundle\_<bundle\_id>.xml*
- Collections must contain at least 2 files – its XML label and an inventory file.
  - Similar in structure (as we will see) the **Collection Label** will contain the metadata pertaining to the contents of the collection.
  - The **Inventory File** is a list of ALL members of the collection. This file must be a 2-column, comma separated table (.csv or .txt or .tab) *See next slide.*
  - Collection files are recommended to be named in the form: *collection\_<bundle\_id>\_<collection\_id>.xml*
    - With the corresponding inventory file : *collection\_<bundle\_id>\_<collection\_id>\_inventory.csv (.txt or .tab)*

# Inventory Files

One of two required files for every collection.  
 The inventory file lists all products considered part of the collection.  
 Inventories consist of a 2-column, comma-separated table.

**Column 1** is the membership status of the product\*\*,  
**Column 2** is the product’s Logical Identifier (LID – URN) and Version Identifier (VID) or together LID::VID

## Example from MAVEN, NGIMS:

P	urn:nasa:pds:maven_ngims:l1a:mvn_ngi_l1a_raw-hk-014000_20131204t100845	::1.2
P	urn:nasa:pds:maven_ngims:l1a:mvn_ngi_l1a_raw-hk-014001_20131204t104746	::1.2
P	urn:nasa:pds:maven_ngims:l1a:mvn_ngi_l1a_raw-hk-014002_20131204t114147	::1.2
P	urn:nasa:pds:maven_ngims:l1a:mvn_ngi_l1a_raw-hk-014003_20140221t075109	::1.2
P	urn:nasa:pds:maven_ngims:l1a:mvn_ngi_l1a_raw-hk-014004_20140221t091744	::1.2
P	urn:nasa:pds:maven_ngims:l1a:mvn_ngi_l1a_raw-hk-014005_20140709t101427	::1.2
P	urn:nasa:pds:maven_ngims:l1a:mvn_ngi_l1a_raw-hk-014006_20140716t102146	::1.2
P	urn:nasa:pds:maven_ngims:l1a:mvn_ngi_l1a_raw-hk-014007_20140716t131058	::1.2
P	urn:nasa:pds:maven_ngims:l1a:mvn_ngi_l1a_raw-hk-014008_20141007t135414	::1.2
P	urn:nasa:pds:maven_ngims:l1a:mvn_ngi_l1a_raw-hk-014009_20141007t172834	::1.2

\*\*For Column 1, Acceptable values are “P” (Primary Member) or “S” (Secondary Member).

“Primary” means the products are present in the current collection and are being registered with the PDS for the first time.

“Secondary” means the products may or may not be present in the current collection but have been already registered in another bundle or collection.

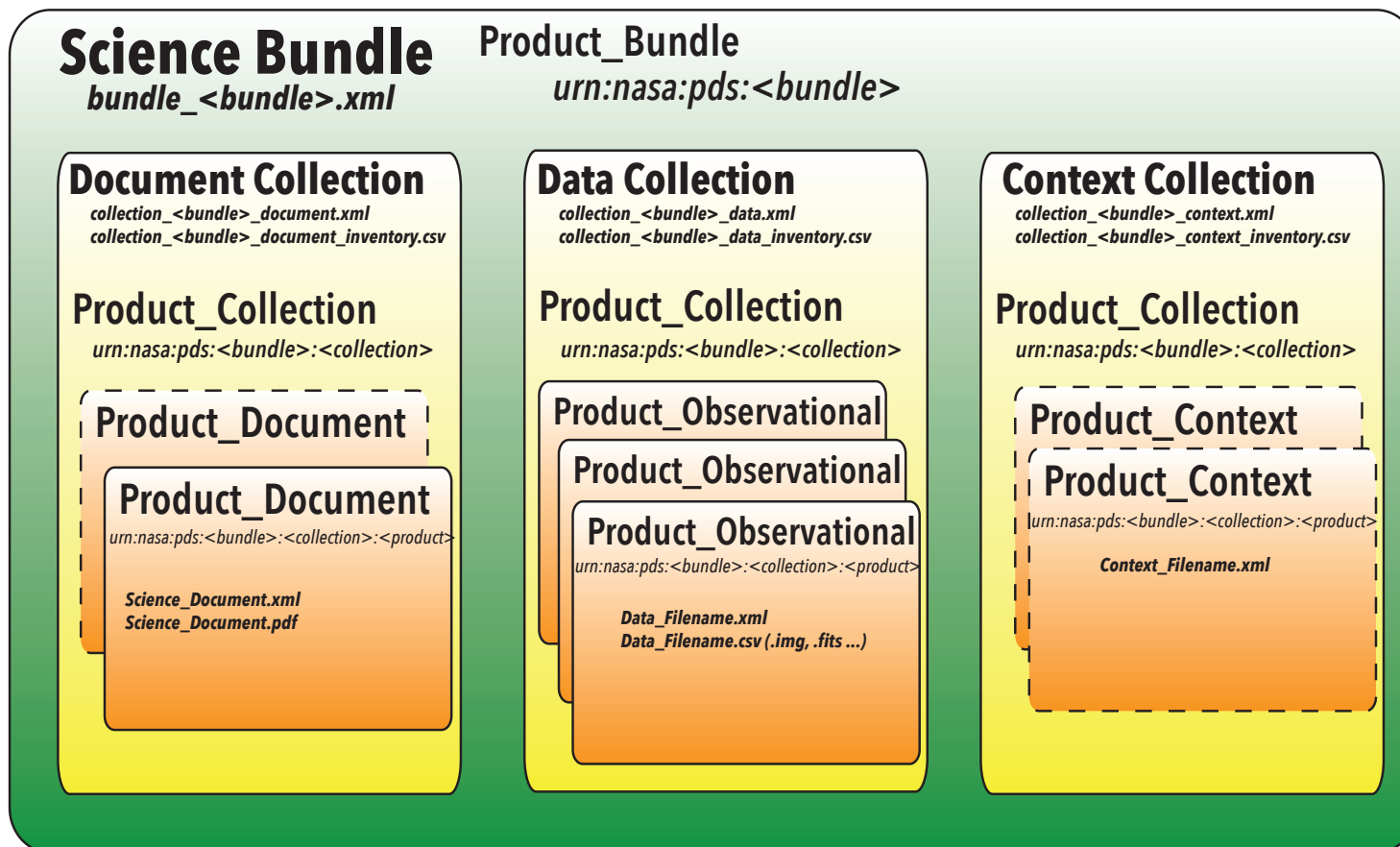
*Context products are managed by the PDS Engineering Node (EN) and will likely always be Secondary products listed in the optional Context Collection*

Everything is a product.

Products have labels.

*PDS4 organization is arranged in Bundles, Collections, and Basic Products*

Logical Identifiers (URNs) designate components of the Bundle according to the segments (4,5, & 6).



*Hypothetical Example Bundle*

Labels for each part of the bundle sit with the files they describe.  
All Collections must contain an **inventory file** (.csv or .txt or .tab).

# PDS4 Structures (cont.)

---

- Basic Product files are found in most collections. These include things like document files (.txt, .pdf) or data files (.csv, .img, .fits, etc.) for example.
  - Every product needs a label (.xml) with a corresponding logical identifier (URN).
  - All labels are organized in roughly the same order although content may vary based on the needs of the different types of products.
  - Data providers will be responsible for helping PDS determine what products need labels and how these should be organized into Collections within the Bundle(s).
- Context products are special products managed by the PDS Engineering Node (EN) for use in cross-referencing missions, facilities, instruments, targets, etc.
  - Because these are held in a repository at the EN, new products must be created/submitted by your discipline node representatives for inclusion in the EN repository.
  - All context products within the Context Collection will therefore be Secondary products as they should already be registered in the PDS Central Registry and are maintained by the EN outside of your Bundle.

# Anatomy of a PDS4 Label

Logical, standard organization for each label for each product.

*PDS4 Labels are implemented using XML.*

*The layout is nearly identical for each product.*

In general, every label has 4 main areas.

## XML Declaration Block

*System information, schema/schematron versioning, XML information*

## Identification Area

*Information about how the file was created and how it should be used including reference information.*

## Observation Area/Context Area

*Additional metadata describing the context (mission, instrument host, instrument, target(s) etc.)*

## File Area

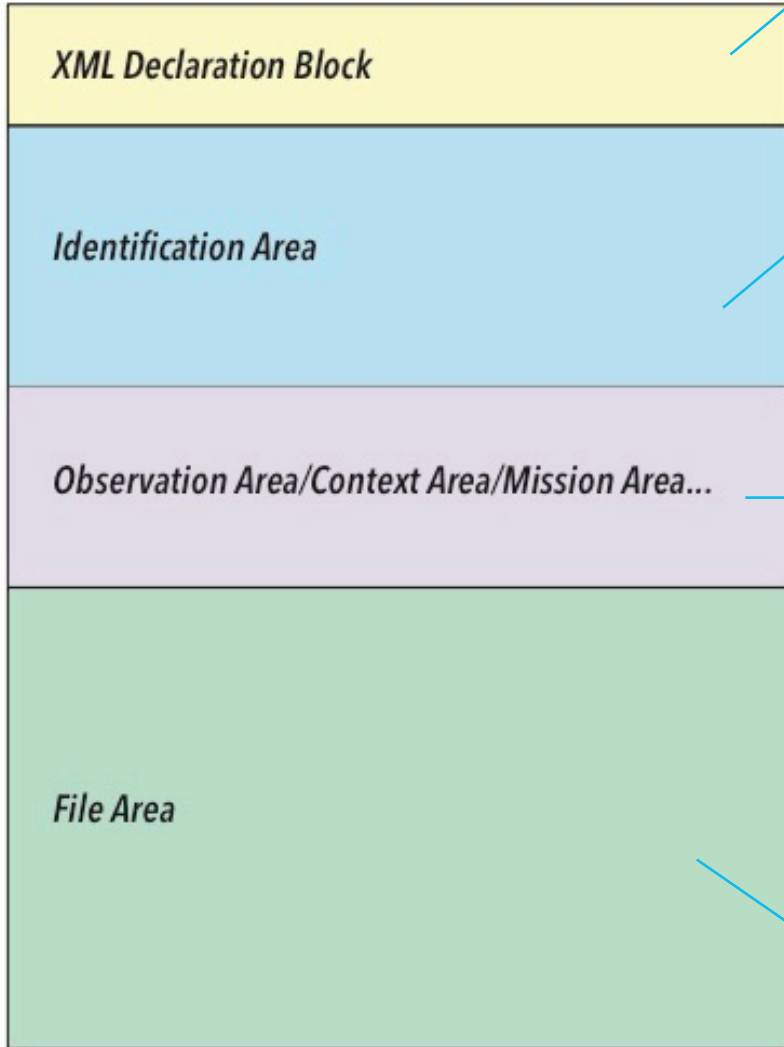
*Information about the file itself, properties of the file, display instructions, and/or pertinent data type.*



*First time users of PDS4 labels may find reading the labels from the bottom easier.*

# Anatomy of a Bundle File

*bundle\_<bundle\_id>.xml*

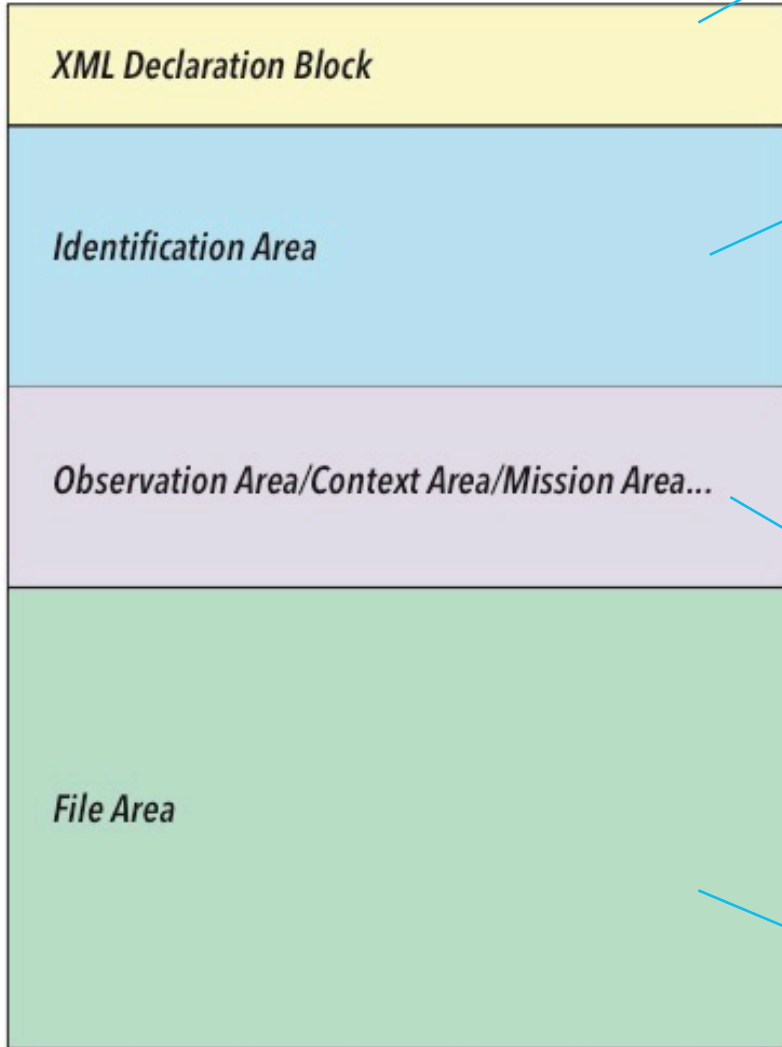


```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="http://pds.nasa.gov/pds4/schema/released/pds/v1/PDS4_PDS_1100.sch" ?>
<Product_Bundle
  xmlns="http://pds.nasa.gov/pds4/pds/v1"
  xmlns:pds="http://pds.nasa.gov/pds4/pds/v1"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://pds.nasa.gov/pds4/pds/v1 http://pds.nasa.gov/pds4/schema/released/pds/v1/PDS4_PDS_1100.xsd">
  <Identification_Area>
    <logical_identifier>urn:nasa:pds:ladee_nms</logical_identifier>
    <version_id>1.0</version_id>
    <title>LADEE Neutral Mass Spectrometer Data</title>
    <information_model_version>1.1.0.0</information_model_version>
    <product_class>Product_Bundle</product_class>
    <Citation_Information>
      <author_list>Benna, Mehdi; Lyness, Eric</author_list>
      <publication_year>2014</publication_year>
      <description>
        This bundle contains the data collected by the Neutral Mass Spectrometer (NMS) instrument aboard the Lunar Atmosphere and Dust Environment Explorer (LADEE) satellite, along with the documents and other information necessary for the interpretation of that data.
      </description>
    </Citation_Information>
  </Identification_Area>
  <Context_Area>
    <Investigation_Area>
      <name>LADEE</name>
      <type>Mission</type>
      <Internal_Reference>
        <lid_reference>urn:nasa:pds:context:investigation:mission.ladee</lid_reference>
        <reference_type>bundle_to_investigation</reference_type>
      </Internal_Reference>
    </Investigation_Area>
    <Observing_System> [18 lines]
    <Target_Identification>
      <name>Moon</name>
      <type>Satellite</type>
      <Internal_Reference>
        <lid_reference>urn:nasa:pds:context:target:satellite.moon</lid_reference>
        <reference_type>data_to_target</reference_type>
      </Internal_Reference>
    </Target_Identification>
  </Context_Area>
  <Bundle>
    <bundle_type>Archive</bundle_type>
  </Bundle>
  <File_Area_Text>
    <File>
      <file_name>readme.txt</file_name>
    </File>
    <Stream_Text> [4 lines]
  </File_Area_Text>
  <Bundle_Member_Entry> [4 lines]
  <Bundle_Member_Entry> [4 lines]
  <Bundle_Member_Entry> [4 lines]
  <Bundle_Member_Entry> [4 lines]
  <Bundle_Member_Entry> [4 lines]
  <Bundle_Member_Entry> [4 lines]
  <Bundle_Member_Entry>
    <lid_reference>urn:nasa:pds:ladee_nms:xml_schema</lid_reference>
    <member_status>Primary</member_status>
    <reference_type>bundle_has_schema_collection</reference_type>
  </Bundle_Member_Entry>
```

*Example from LADEE NMS*

# Anatomy of a Collection File

collection\_<bundle\_id>\_<collection\_id>.xml



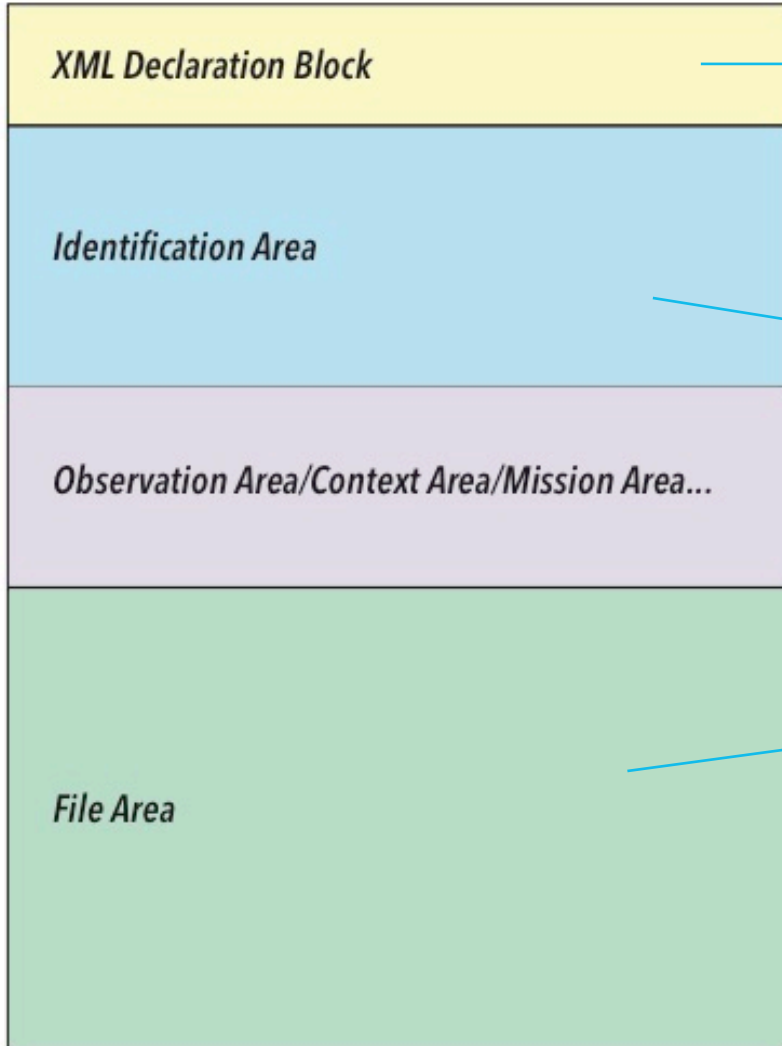
```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="http://pds.nasa.gov/pds4/pds/v1/PDS4_PDS_1100.sch"?>
<Product_Collection
  xmlns="http://pds.nasa.gov/pds4/pds/v1"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://pds.nasa.gov/pds4/pds/v1 http://pds.nasa.gov/pds4/schema/released/pds/v1/PDS4_PDS_1100.sch"
  >
  <Identification_Area>
    <logical_identifier>urn:nasa:pds:ladee_nms:data_raw</logical_identifier>
    <version_id>1.0</version_id>
    <title>NMS Raw Data Collection</title>
    <information_model_version>1.1.0.0</information_model_version>
    <product_class>Product_Collection</product_class>
    <Citation_Information>
      <author_list>Benna, Mehdi; Lyness, Eric</author_list>
      <publication_year>2014</publication_year>
      <description>
        Raw data acquired from the NMS instrument during its
        primary mission. No calibrations, corrections, or
        interpretations are applied.
      </description>
    </Citation_Information>
  </Identification_Area>
  <Context_Area>
    <Investigation_Area>
      <name>LADEE</name>
      <type>Mission</type>
      <Internal_Reference>
        <lid_reference>urn:nasa:pds:context:investigation:mission.ladee</lid_reference>
        <reference_type>collection_to_investigation</reference_type>
      </Internal_Reference>
    </Investigation_Area>
    <Observing_System> [21 lines]
    <Target_Identification>
      <name>Moon</name>
      <type>Satellite</type>
      <Internal_Reference>
        <lid_reference>urn:nasa:pds:context:target:satellite.moon</lid_reference>
        <reference_type>data_to_target</reference_type>
      </Internal_Reference>
    </Target_Identification>
  </Context_Area>
  <Collection>
    <collection_type>Data</collection_type>
    <description>
      Raw data acquired from the NMS instrument during its
      primary mission. No calibrations, corrections, or
      interpretations are applied.
    </description>
  </Collection>
  <File_Area_Inventory>
    <File>
      <file_name>collection_nms_data_raw_inventory.csv</file_name>
    </File>
  </File_Area_Inventory>
</Product_Collection>
```

Example from  
LADEE NMS



# Anatomy of a Document File

<filename>.xml



```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="http://pds.nasa.gov/pds4/pds/v1/PDS4_PDS_1301.sch"
  schematypens="http://purl.oclc.org/dsdl/schematron"?>
<?xml-model href="https://pds.nasa.gov/pds4/disp/v1/PDS4_DISP_1004.sch" schematypens="http://purl.oclc.org/dsdl/schematron"?>
<Product_Document
  xmlns="http://pds.nasa.gov/pds4/pds/v1"
  xmlns:pds="http://pds.nasa.gov/pds4/pds/v1"
  xmlns:disp="http://pds.nasa.gov/pds4/disp/v1"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://pds.nasa.gov/pds4/pds/v1
  http://pds.nasa.gov/pds4/disp/v1 http://pds.nasa.gov/pds4/disp/v1/PDS4_DISP_1004.xsd
  http://pds.nasa.gov/pds4/pds/v1/PDS4_PDS_1301.xsd">
  <Identification_Area>
    <logical_identifier>urn:nasa:pds:maven.iuvs.raw:document:sis</logical_identifier>
    <version_id>7.0</version_id>
    <title>MAVEN/IUVS PDS Archive Software Interface Specification</title>
    <information_model_version>1.3.0.1</information_model_version>
    <product_class>Product_Document</product_class>
    <Citation_Information>
      <author_list>Deighan, Justin; Jeppesen, Chris; Holsclaw, Gregory</author_list>
      <publication_year>2016</publication_year>
      <description>
MAVEN/IUVS PDS Archive Software Interface Specification
      </description>
    </Citation_Information>
  </Identification_Area>
  <Document>
    <document_name>MAVEN/IUVS PDS Archive Software Interface Specification</document_name>
    <publication_date>2016-08-26</publication_date>
    <description>
MAVEN/IUVS PDS Archive Software Interface Specification
    </description>
    <Document_Edition>
      <edition_name>MAVEN/IUVS PDS Archive Software Interface Specification</edition_name>
      <language>English</language>
      <files>1</files>
      <Document_File>
        <file_name>iuvs_pds_sis.pdf</file_name>
        <document_standard_id>PDF/A</document_standard_id>
      </Document_File>
    </Document_Edition>
  </Document>
</Product_Document>
```

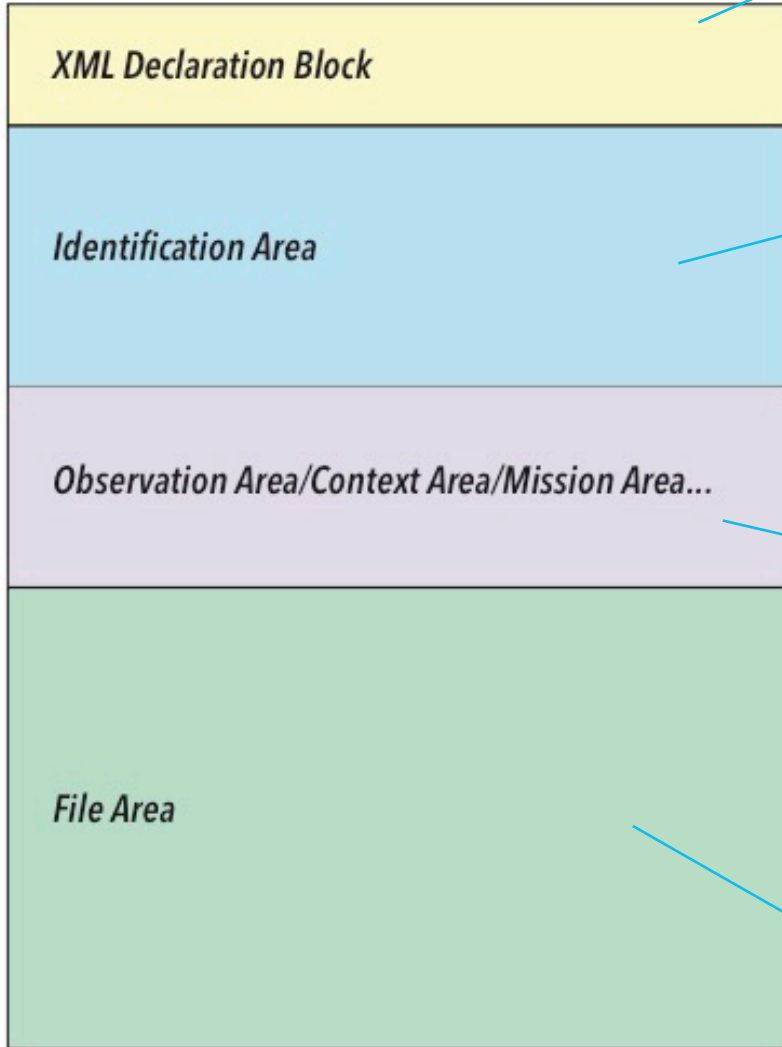
Example from  
MAVEN IUVS

# Anatomy of a Data File

<filename>.xml

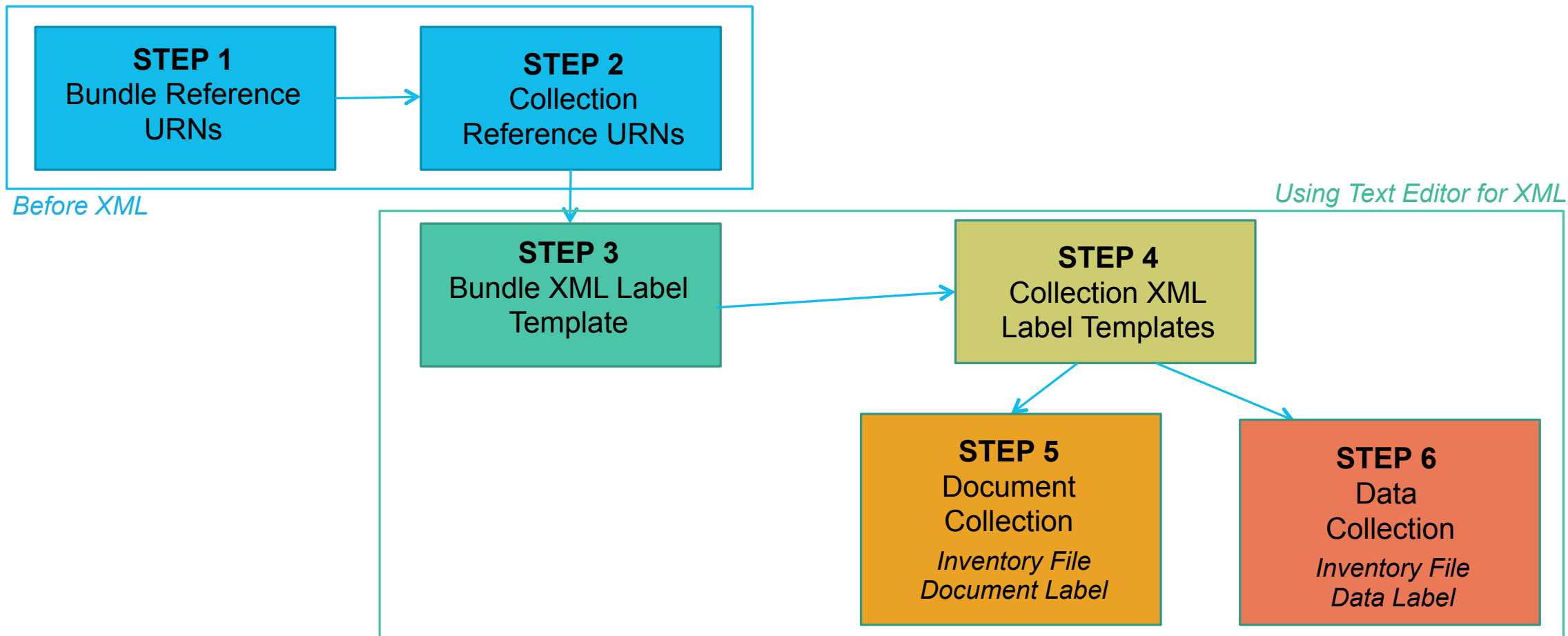


Example from MAVEN ACC



```
<?xml version='1.0' encoding='UTF-8'?>
<?xml-model href="http://pds.nasa.gov/pds4/schema/released/pds/v1/PDS4_PDS_1301.sch"
schematypens="http://purl.oclc.org/dsdl/schematron" ?>
<Product_Observational
xmlns="http://pds.nasa.gov/pds4/pds/v1"
xmlns:pds="http://pds.nasa.gov/pds4/pds/v1"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://pds.nasa.gov/pds4/pds/v1
http://pds.nasa.gov/pds4/pds/v1/PDS4_PDS_1301.xsd">
  <Identification_Area>
    <Logical_Identifier>urn:nasa:pds:maven_acc:profile:pro_acc_p05043</Logical_Identifier>
    <Version_ID>1.0</Version_ID>
    <Title>MAVEN Accelerometer Profile P05043</Title>
    <Information_Model_Version>1.3.0.1</Information_Model_Version>
    <Product_Class>Product_Observational</Product_Class>
    <Modification_History>
      <Modification_Detail>
        <Modification_Date>2014-04-15</Modification_Date>
        <Version_ID>1.0</Version_ID>
        <Description>Version 1 Release 1 uses onboard accel of the COM. Averaging over 99s reduces effects of fuel slosh. Cal.
of the aero DB is incomplete, so 20% biases may exist for science orbits and 10% for deep dip orbits. </Description>
      </Modification_Detail>
      <Modification_Detail>
        <Modification_Date>2016-10-31</Modification_Date>
        <Version_ID>1.1</Version_ID>
        <Description>Version 1 Release 2 adds unsmoothed ACC data and associated uncertainties at 1 second intervals. </Description>
      </Modification_Detail>
    </Modification_History>
  </Identification_Area>
  <Observation_Area>
    <Time_Coordinates> [3 lines]
    <Primary_Result_Summary> [8 lines]
    <Investigation_Area>
      <Name>Accelerometer</Name>
      <Type>Mission</Type>
      <Internal_Reference>
        <LID_Reference>urn:nasa:pds:context:investigation:mission:maven</LID_Reference>
        <Reference_Type>data_to_investigation</Reference_Type>
      </Internal_Reference>
    </Investigation_Area>
    <Observing_System>
      <Name>MAVEN</Name>
      <Observing_System_Component>
        <Name>Accelerometer</Name>
        <Type>Instrument</Type>
        <Internal_Reference>
          <LID_Reference>urn:nasa:pds:context:instrument:acc:maven</LID_Reference>
          <Reference_Type>is_instrument</Reference_Type>
        </Internal_Reference>
      </Observing_System_Component>
      <Observing_System_Component>
        <Name>MAVEN</Name>
        <Type>Spacecraft</Type>
        <Internal_Reference> [3 lines]
      </Observing_System_Component>
    </Observing_System>
    <Target_Identification> [7 lines]
  </Observation_Area>
  <Reference_List> [9 lines]
  <File_Area_Observational>
    <File>
      <File_Name>mvn_acc_l3_pro-acc-p05043_20170506_v01_r02.tab</File_Name>
      <Local_Identifier>PRO_acc_P05043_file</Local_Identifier>
      <Creation_Date_Time>2017-07-29T19:44:52</Creation_Date_Time>
      <File_Size_Unit>"byte">38464</File_Size_Unit>
      <Records>601</Records>
    </File>
    <Table_Character> [90 lines]
  </File_Area_Observational>
</Product_Observational>
```

# PDS4 Label Exercise



# PDS4 Label Exercise

---

- Using the files on the Thumb Drive you were given when you came in today, open the directory titled “2\_Bundle\_Exercise”
  - In this directory you will find copies of today’s agenda, and this presentation.
  - Please open the file: “Training\_Exercise\_Worksheet.docx”
    - We will be working through this simple example of how to create a PDS4 Bundle with the use of templates
    - The goal of this exercise is to guide you through editing template labels so that you become more familiar with the basics of PDS4 label design. Especially focusing on the LID(URN) referencing and the common structures used throughout the bundle.
    - For the exercise today we will be using your favorite or most convenient text editor.
  - We will be starting with the description of the data we will be using as an example today.
    - Open the ”readme.txt” file that provides the starting description of the data.
      - Example data for our exercise today is:
        - ***Mars Science Laboratory Entry, Descent, and Landing Atmospheric Profiles***

# PDS4 Exercise Wrap-Up

---

- Now that you've completed an example bundle...

# PDS4 Tools and Resources

---

- Today we utilized a basic text editor to edit templates to produce labels
- Other tools exist to make your experience easier.
  - **XML Editors** – are special text editors that allow some basic content validation as you go. They connect to online resources referenced in the declaration block and check values against the rules sets set by the IM, PDS4 Schema and Schematron files.
    - **oXygen** – (subscription/license options for academic institutions) provides good validation and XML hierarchy support.
    - **Eclipse** – (free) Good XML hierarchy support, schematron validation is not as functional in latest version.
  - **Bundle/Label Design Tools** – Tools developed by PDS Nodes to help in creation of labels and designing PDS4 Bundles.
    - **ELSA** – Educational Labeling System for Atmospheres (Atmospheres Node) provides URN generation, bundle/collection template support and basic product template organization and provides interfaces for submission to the Atmospheres Node. *(under development - limited closed beta testing)*
    - **OLAF** – On-Line Archiving Facility (Small Bodies Node) provides an online environment for submission and generation of PDS4 products for the Small Bodies Node
    - **PLAID** – Planetary Label Assistant for Interactive Design (Cartography and Imaging Science Node) – interactive label design tool for basic products
    - **iggp-docgen** – (Planetary Plasma Interactions Node)
    - **MakeLabels** – (Geosciences Node) Windows based tool for label creation.

# PDS4 Tools and Resources

---

- Contact personnel at PDS Nodes
  - **Atmospheres** – Lynn Neakrase ([lineakras@nmsu.edu](mailto:lineakras@nmsu.edu)), Nancy Chanover ([nchanove@nmsu.edu](mailto:nchanove@nmsu.edu))
    - <https://pds-atmospheres.nmsu.edu/>
  - **Cartography and Imaging Sciences** – Lisa Gaddis ([lgaddis@usgs.gov](mailto:lgaddis@usgs.gov))
    - <https://pds-imaging.jpl.nasa.gov/>
  - **Geosciences** – Susan Slavney ([slavney@wunder.wustl.edu](mailto:slavney@wunder.wustl.edu)), Ed Guinness ([guinness@wunder.wustl.edu](mailto:guinness@wunder.wustl.edu))
    - <http://pds-geosciences.wustl.edu/>
  - **Planetary Plasma Interactions** – Joe Mafi ([jmafi@igpp.ucla.edu](mailto:jmafi@igpp.ucla.edu)),
    - <https://pds-ppi.igpp.ucla.edu/>
  - **Ring-Moon Systems** – Mitch Gordon ([mgordon@seti.org](mailto:mgordon@seti.org))
    - <https://pds-rings.seti.org/>
  - **Small Bodies Node** – Anne Raugh ([raugh@astro.umd.edu](mailto:raugh@astro.umd.edu)),
    - <https://pds-smallbodies.astro.umd.edu/>

# Evaluation

---

- If you would like to provide feed back on today's exercise or any part of the training, please use track changes on the Worksheet file or simply contact Lynn Neakrase with your questions or comments.
  - [lnekras@nmsu.edu](mailto:lnekras@nmsu.edu)
  - With feedback from these training sessions, I can redistribute a revised worksheet for future use and put you on a mailing list for future training session opportunities.
- Thanks in advance to anyone willing to take some time to review what we've done today.