



WELCOME TO ARCHIVING

at the PDS Atmospheres Node

Guidelines for archiving under the PDS4 Standard at the PDS Atmospheres Node

Version 1.4

June 2023

DAP EDITION

Nancy Chanover, Lynn Neakrase, Lyle Huber
Department of Astronomy, New Mexico State University

Table of Contents

Table of Contents	1
INTRODUCTION	1
STARTING WITH PDS4	3
PDS4 Hierarchy.....	3
Products – ‘Everything is a Product’.....	3
Bundles.....	3
Collections.....	4
Collection Label & Inventory.....	4
Document Collection.....	5
Data Collections.....	5
Context Collection.....	6
Other Collections.....	6
Basic Products.....	6
Logical Identifiers (LIDs) and Uniform Resource Names (URNs).....	7
Timeline for Success	8
Step 1 - COMMUNICATION.....	8
Step 2 – ARCHIVE SCHEDULE & BUNDLE DESIGN.....	9
Step 3 – ARCHIVE CREATION & VALIDATION.....	9
Step 4 – DOCUMENTATION & PEER REVIEW.....	10
Step 5 – LIEN RESOLUTION & CERTIFICATION.....	11
NASA PDS Quarterly Reporting.....	11
Editing XML Files	12
ARCHIVING CHECKLIST	13
IMPORTANT NODE CONTACTS	14

INTRODUCTION

Welcome to archiving at the NASA Planetary Data System (PDS) Planetary Atmospheres Node (ATM)! First, congratulations on making the decision and commitment to archive your data. Whether you were funded through a NASA grant or made the decision independently to archive with the PDS, archiving data can be a rewarding final step to completing a research project and making your data available to the planetary science community. It also fulfills an important federal requirement to ensure long-term accessibility of digital scientific data.

The goal of this document is to outline the process of archiving with the Atmospheres Node, guiding you through the responsibilities that will be expected of you while working with the node. We will provide primers on the PDS4 Data Standard as well as a step-by-step outline of the process from first contact with you, *the data provider*, through submission to the archive, the peer-review process, and final acceptance and publishing of the data via the ATM website **(See Checklist at the end of this document)**.

The process of archiving with PDS is more involved than sending data to a repository. **The PDS is not merely a repository.** Archiving is a rigorous process that involves attention to details at every step, which requires a working relationship between you and the ATM node personnel. Good communication, first and foremost, is the most important part of this partnership between the data provider and ATM, and it will lead to efficient planning and successful completion of your portion of the PDS archiving requirements.

Your choice to archive with PDS is a commitment to follow the guidelines provided by your archiving node. Whether your immediate goal is to complete requirements from your grant, provide links to your data for a journal publication, or simply add to the communal archive for specific planetary science disciplines, the PDS also has longer-term goals for preservation and accessibility. Please be aware that *archiving* data is preservation of your data for future scientists, with the long-term goal of accessibility over 50-100 years or longer. Because nearly all planetary data collected is funded through federal or state grants, there exist requirements that your data be made publicly available in a timely fashion. PDS satisfies those requirements by conducting scientific peer reviews of the data for certification and providing long-term stability and preservation through the distributed discipline node system, with multiple backups of accepted and certified archive data.

The Atmospheres Node is one of six geographically distributed discipline nodes of the PDS . ATM is located at New Mexico State University in Las Cruces, New Mexico within the Department of Astronomy. Planetary atmospheres data, including atmospheric observations from the surface boundary layer to the exosphere for any planetary body in the Solar System

(with the exception of the Earth) are the focus of the node's holdings. This may include (but is not limited to) boundary layer winds, aeolian processes (including grain movement, dunes/ripples, dust devils, dust storms) and surface meteorology, density, temperature, and pressure profiles through the atmosphere, composition of the atmosphere, cloud dynamics and storm evolution, interactions with the solar wind and atmospheric loss. ATM archives data not only from missions, but also ground-based observations, laboratory and analog field measurements, and associated model work (where applicable). The PDS does not have the means to archive software and modeling code. Data providers wishing to archive software are recommended to use the NASA Planetary Science GitHub (<https://github.com/NASA-Planetary-Science>), where code can be stored and made available for interested users. ATM is in the process of establishing an Atmospheric Modeling Annex where atmospheric model outputs can be stored and affiliated with the PDS if not formally archived.

Before getting started with the preparation of your data, we recommend that you become familiar with PDS4 terminology and the PDS Archiving Process, beginning with this document. This guide will provide a basic tutorial for interacting with ATM and provide necessary links to tutorials and tools to aid in your work. Remember, communication is key! If you run into difficulty or have questions, feel free to contact the ATM node personnel listed within this document.

For more in-depth discussion of PDS4 terminology and archiving process, see the PDS4 Data Standards (https://pds.nasa.gov/datastandards/documents/sr/current/StdRef_1.20.0.pdf) and the PDS4 Data Providers Handbook (https://pds.nasa.gov/datastandards/documents/dph/current/PDS4_DataProvidersHandbook_1.20.0.pdf).

STARTING WITH PDS4

The PDS recognizes that archiving can seem like a daunting process, disconnected from your science and required as some mythical last step in your award period. We hope that by breaking this archiving process down into manageable steps, to be done throughout your grant period, archiving can be a natural consequence of all the work you do.

In order to successfully complete an archive with ATM, it is necessary to understand how the PDS archive works. Working knowledge of the archive itself – how data are organized, what kinds of metadata are useful, etc. – will allow you, the data provider, to more efficiently organize your efforts in producing viable archive products. Within this task, it is important to understand the workings of the PDS4 Archiving Standard. This will include all aspects of the structure of your archive – the basic hierarchy of the archive, creating the system of unique identifiers, utilizing context products to ‘build’ your reference system within your archive.

Communication will be key to your success. Early and frequent communication with the relevant ATM personnel can be the difference in understanding your responsibilities as a data provider and successful completion of the archiving process. The Atmospheres Node is always here to help in this process by facilitating your archiving effort and providing answers to any of your archiving questions.

PDS4 Hierarchy

Products – ‘Everything is a Product’

All pieces of the archive under PDS4 are considered products. Products, in their most basic sense, include PDS4 labels, which contain unique identifiers and metadata. Unique identifiers allow data to be cataloged in a PDS registry and metadata provide details about the individual products, which can be used for search and retrieval, once the archive is validated and registered with the PDS. In order to create a viable archive, this concept is expanded to include the hierarchy that provides the heart of the PDS4 system.

The term ‘product’ will refer to anything within the archive that has a label containing a referenceable unique identifier. Labels contain the metadata and identifiers and references. Under PDS4, labels are expressed as detached XML files that usually have the same root name as the file they describe.

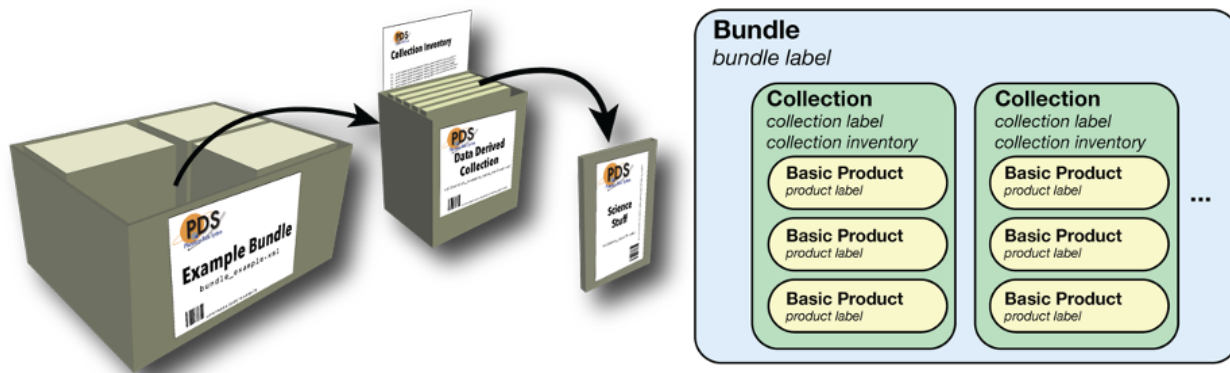


Figure 1. Schematic diagrams of the PDS4 bundle hierarchy. Cartoon view showing the bundle structure as nested boxes with ‘packing’ information (left) with corresponding diagram showing the same structures and labels.

Bundles

At the top-most level, the data are organized by *bundles*. A bundle (Figure 1) is analogous to a dataset or a grouping of all parts of a dataset including documentation, data (potentially multiple processing levels), and ancillary files (calibrations, related references, etc.). The *bundle product* is a single label file (.xml) that sits in the top-most directory. All bundle files will have the file naming convention:

bundle_<bundle-id>.xml

[bundle file name]

Bundle file names should always include a <bundle-id> segment that identifies which bundle this file describes.

Bundle labels, or bundle product files, contain information about the bundle at the top-most level. Often there will be context references to the investigation, host(s) (spacecraft or facilities), instruments, and targets (see *Context Products below*). The other information will contain a listing of all member collections in the bottom part of the label. Member collections will include the data collection(s), a calibration (if needed) collection, a document collection, and a context collection, etc.

Collections

Collections (Figure 1) are logical groupings of related files that, together, comprise an archive bundle. Your data will be in one or many collections (depending on the type of data), and you will have a document collection. Other collections may be necessary depending on the needs of your project. ATM typically requires the *document* collection and the *context* collection. A good archive should always contain good documentation, allowing future users to be able to understand how and why the data were generated/collected, archived the way they are, and what types of uses are appropriate for the data. Collections under PDS4 are only officially required to have two files: the collection *label* (.xml) and the collection *inventory* (.csv). All other members of the collection may or may not be present in the directory or subdirectories where the label and inventory files are found.

Collection Label & Inventory

Just as was the case with the bundle-level files, collections are considered products and therefore require labels. The collection label contains a unique identifier (the logical identifier of which is the same as the bundle except it contains an extra part for the <collection-id>). Collection files should have the form:

collection_<bundle-id>_<collection-id>.xml

[collection file name]

Other important information that can be found in a collection label includes context references relevant to the bundle and collection and also a pointer to the inventory file.

The inventory file will be a 2-column comma-separated-value table (.csv) that contains ALL products considered part of the collection. The first column describes the status of the member product and may only be set to one of two accepted values: *P* (Primary) or *S*(Secondary). 'Primary' denotes that the associated product is new to the PDS4 system, or rather it is being

registered with PDS for the first time with its inclusion in this bundle. 'Secondary' refers to a product that is already registered at PDS and is being referenced by this bundle. The second column of the inventory is often referred to as the LID::VID reference or Logical Identifier (LID) :: Version Identifier (VID). Essentially this value will be the combined text string of the logical identifier concatenated with its version identifier. This will be of the form of the Uniform Resource Name (URN) (see *next section*), double colon (::), and the version string (e.g., '1.0'). Inventory file names will have the form:

collection_<bundle-id>_<collection-id>_inventory.xml [collection inventory file name]

At ATM, the bundle and collection files have these specific naming schemes to set these important files apart from the rest of the files you may have in your bundle. This allows these files to be quickly found every time within all ATM PDS4 archive bundles.

Document Collection

Under PDS4, ATM will require the presence of a document collection. This provides a common location to place all related documents for the bundle. ATM will require at least one document in this collection. This required document should be a sort of Users' Guide to the bundle. This is analogous to the mission data's Software Interface Specification (SIS) document. The contents of this guide should be necessary background information, data collection methodology, assumptions and calibration information, and anything else that would be needed to understand and use the data in your bundle.

Data Collections

Likely the rest of the collections in the bundle will be some sort of data collection. Whether these are raw, calibrated, partially processed, or derived data, they will all have the same form. Data collections will contain the bulk of your bundle in most cases. Just as in the other cases above, the collection files will include the collection label and the inventory file.

Context Collection

The context collection will be provided by ATM once you are nearing completion of your bundle. Context collections are a simple listing of all context products used in your project. This provides a quick-look guide for ATM to see what's involved in your bundle and can help us facilitate search and retrieval across our holdings. Context products --- official PDS-maintained references to investigations, observing systems, and targets --- are managed by the PDS Engineering Node (EN). In order to fully validate your complete bundle, all used context products should be present in the Context Bundle at the EN, and this includes new investigations and host/instrument pairings as well. ATM will provide context information based on your communication with them and will iterate to make sure your context inventory accurately represents your work.

Other Collections

There are other collections that may be necessary for your project. These may include (but not limited to) browse, calibration, geometry, and xml-schema collections. Browse collections are typically used for thumbnails or quick-look products that represent data files but are not data. Calibration and geometry collections can be separate collections for organizing calibration, location, and pointing information, not handled in the metadata of data collections. If geometry or calibrations are more involved than simple standard values or if they change from product to product it may make more sense to have separate collections of these products.

XML-schema collections are PDS collections like the context collection (above) that ATM will provide if deemed necessary. These include a secondary inventory of the different versions of the PDS4 Information Model Schema used to generate your labels. In general, most bundles will be written to a single version of the PDS Information Model. In some circumstances such as long-term accumulating datasets, improved functionality of later versions of PDS4 may be useful. In such a case, an xml-schema collection may be used to keep track of which versions of the IM have been used in a bundle.

Basic Products

The bulk of your archive will be comprised of *basic products*. These include your documents, data, and ancillary files (i.e., calibration, browse files, or geometry information, etc.). Just as was the case with bundle and collection products, basic products require PDS4 XML labels to contain all the vital metadata for future use of the data. Basic product labels are, in general, more detailed than the previous types of labels because much more information has to be stored in these labels to provide all the relevant information for the data. Some of this information will be a detailed breakdown of the observing system component context (as in the other labels) – this context information will likely be a subset of the entire system provided in the bundle/collection levels. Observations will include start/stop times, information about exposure times, linkages to calibration data, display instructions, table column headers, among a few of the options. Document files will provide publishing information such as dates and digital object identifiers (DOIs), and also file format information and associated other versions of the document (e.g., different languages or formats for convenience).

Different from bundles and collections, there are no required formats for the filenames for basic products. Typically, the filenames of the individual files will suffice. When prompted, the ATM will suggest that you follow rules for logical identifier formation (see next section) where all filenames are in lower case (no capital letters) and the only allowed separators are ‘_’, ‘-’, or ‘.’. There is no requirement to do this, however for simplicity, when filling out labels, following these suggestions will allow you to quickly complete parts of the label.

Basic products come in such a variety of types that describing each in detail here is not appropriate. ATM prefers to supply data providers with templates for appropriate products after discussion of the project. This allows the ATM to tailor some of what is provided in an attempt to lessen the work for you, the data provider. Iteration of this process is almost always necessary to ensure the appropriate fit to your needs. This means that starting your conversation with ATM early is the most efficient use of everyone's time, and the early preparation and interface with ATM will pay off when you are ready to submit your bundle(s) to the archive.

Logical Identifiers (LIDs) and Uniform Resource Names (URNs)

Throughout this guide, Logical Identifiers (LIDs) and Uniform Resource Names (URNs) have been used almost synonymously. In this section we will delve into their construction and use and provide justification for why these text strings are so important to the PDS4 Archiving Standard.

As stated above, nearly everything under PDS4 is a product, and all products are registered into the PDS4 Registry System once submitted and reviewed. In order to do this, every product needs a unique identifier. Under PDS4 we accomplish this with our Logical Identifier (LID) system. Every product is assigned a LID and this text string acts like a barcode on items in a store. The PDS4 system understands the parts of the LID string and knows how to organize these within your bundles and the entire PDS archive. PDS4 utilizes Uniform Resource Names (URNs) to register your products. URNs designate a specific order and form of the text string that serves as your unique identifiers for every product in your bundle. URNs are non-location-specific virtual links that are made up of several *segments* (up to, not exceeding a total of six). Each segment serves a purpose in determining physically where each product resides within the PDS Archive. These work much the same way as Uniform Resource Locators (URLs) for web pages, however, the URLs send you to a specific location to find the product at the end of that link. URNs rely on *resolving institutions* to provide the knowledge of where everything is within the system without the user needing to know the location specifically.

For PDS, the LIDs will have the form:

urn:nasa:pds:<bundle-id>:<collection-id>:<product-id>

In this example, the resolving institution is handled by the first three segments (urn:nasa:pds:). The URN points to NASA and PDS as the responsible institution to know where the files are stored, without the user having to link directly to the file. Several international planetary archives are committed to using the PDS4 Archiving Standard through membership with the International Planetary Data Alliance (IPDA) meaning that their resolving institutions would look

differently (e.g., JAXA – urn:jaxa:darts;; ESA – urn:esa:psa;; etc.). These initial segments essentially ensure that there is a registry at that institution responsible for pointing to all products housed within.

The next three segments may or may not all be used depending on the product they describe, and this represents the PDS4 hierarchy described above and have the form:

Bundle	<i>urn:nasa:pds:<bundle-id></i>
Collection	<i>urn:nasa:pds:<bundle-id>:<collection-id></i>
Product	<i>urn:nasa:pds:<bundle-id>:<collection-id>:<product-id></i>

Because of the relationships between each of these identifiers, ATM recommends thinking about your proposed bundle and making decisions on what your archive layout will be in order to set your LID naming scheme very early in the process.

Timeline for Success

The PDS and your NASA program officers hope for an efficient, positive experience with your archiving effort. In general, we suggest you begin as early as possible to ensure you have plenty of time to get through the archiving process. In order to efficiently complete a successful archive, ATM recommends the following to guide your workflow:

Step 1 - COMMUNICATION

Communication is key and begins with your initial contact with ATM. Your communication might begin with your request for a Letter of Support during your proposal writing, before you are funded. After receiving confirmation of your funding, it is important to inform ATM of your success. This allows us to begin the process of getting appropriate templates created for your specific project. This initial communication with ATM informs us that your project exists and that you are going to begin preparing an archive. You may or may not have an idea of what all your products will be at this stage, but it is important to begin the conversation with PDS to allow us to plan for accommodating your needs effectively. The PDS nodes work with a wide variety of data providers, and it is essential that we understand who is expected to provide data so that we can allocate our staffing resources most effectively.

Step 2 – ARCHIVE SCHEDULE & BUNDLE DESIGN

Archiving Schedule should be discussed with ATM in order to accommodate your project effectively. The archiving schedule also provides a guideline for you, the data provider, to work toward managing your project. Changes to this schedule are both expected and respected as long as PDS is informed of the changes. Good communication results in efficient archiving, and letting ATM know when issues arise allows effective adjustment of the busy schedule of the node.

Label Design can begin the moment you have funding and an idea of what your data will look like. In order to help ATM prepare label templates for your bundle, it is helpful to provide sample data products. This should be considered for the entire archive. Providing samples of every different type of expected product will help ATM tailor label templates for your use. If no data exist at that time, it could be helpful to provide information about which data yours might closely resemble. ATM will then send you a template package zip file that should have partially populated label templates for your specific needs. Iteration between you and the ATM node personnel will likely be necessary to fine-tune your label templates.

Step 3 – ARCHIVE CREATION & VALIDATION

Archive Creation is the responsibility of the data provider. The PDS can and will provide assistance every step of the way, however the actual work of generating the completed labels and assembling the bundle is the responsibility of the data provider. For larger projects (though not limited to large datasets) it is often a good idea to fill out some test labels to check with ATM that all needed metadata is complete and formatted correctly. Once both you and ATM agree that the labels are adequate for your needs, full creation of the bundle can progress. The remaining labels can be filled out by hand or scripted to complete as in a batch.

Archive Validation can describe a number of validation processes. While you input values into tags within the XML label templates, XML-editors can provide some syntactical validation. When you and ATM iterate on your labels, this is the type of validation that is usually completed. Once your bundle is completed and submitted to the node, ATM will run a formal validator on the bundle. This step will check the linkages between labels and all content to ensure compliance with the PDS4 Archive Standard. If issues are found at this stage, ATM may request you fix the issues. Any and all fixes that can be completed before the formal review should improve the review.

Step 4 – DOCUMENTATION & PEER REVIEW

Documentation is one of the most important parts of any archive. Documentation should be completed with the bundle submission, but we draw attention to it here. As the bundle goes to the peer review it is important that all documentation is in order. ATM recommends that for derived data sets and/or DAP-style data projects, the minimum documentation should be some sort of Users' Guide to your dataset. A User's Guide should contain enough background information to understand the context of your dataset – how the data were collected, assumptions that were made, potential sources of errors, and some discussion of the organization of the archive. The idea behind good documentation is imagining someone accessing your data fifty years from now – would they be able to use it and reproduce results?

NOTE: If producing your data archive is part of a funded proposal, a rough draft of much of the necessary documentation can be formulated early in your project by cutting and pasting from your proposal.

The Peer Review will be organized by the PDS once the entire bundle is submitted and validated and can take up to six months to complete. It is important to remember that this process can be long. The ATM personnel will query you for suggested science reviewers, which should include people knowledgeable in your field, but not associated with your archiving effort. ATM will contact these suggested colleagues and augment the review with PDS personnel who are familiar with PDS4 structure. Because of the enlistment of volunteer reviewers, the process can be drawn out depending on the responsiveness of the reviewers. The point of the review is to check your data for scientific viability and validity as well as double-checking that the PDS4 syntax and organization is complete. Reviewers will be asked to download/access your bundle(s) and look at the documentation, the data, and all materials present in the archive to determine if there are any errors or if there are suggestions that could improve the products.

Step 5 – LIEN RESOLUTION & CERTIFICATION

Review Results and Lien Resolution are the likely outcomes of the peer review process. There will be a scheduled telecon to discuss the results of the review. All reviewers and data providers are asked to participate in the telecon so that all responses are clear to the ATM staff and the data providers. If there are suggestions or errors found by the review panel, the telecon will determine who is responsible for fixing the issues. The responsibility of the fixes could be either the data provider (usually to address science issues) or ATM personnel (usually to fix structural/XML/PDS4 issues). The discussion on the telecon will determine the time necessary for completion of the liens and allow PDS and the data provider to decide what is reasonable for completion of the archive. Depending on the complexity of the fixes, it may be necessary to have a follow-up telecon with reviewers and the data providers to determine if all liens have been resolved.

Data Certification & Distribution is the final step in the process. Following lien resolution and subsequent reviews, the PDS4 archive bundle can be accepted by ATM. Fully complete bundles can be assigned a digital object identifier (DOI) and will be considered certified. Certified status is essentially the equivalent to officially publishing your data. As part of that final step, ATM will host your data via web links on science specific websites. PDS is responsible for the creation and maintenance of the websites and may ask you for input or final checks that all information presented on the data access web pages is accurate. At the same time, PDS will harvest the data into the PDS Central Registry effectively creating all of the LID references within the system. From that point on, your newly certified archive bundle will be discoverable in the PDS system.

Reporting Requirements

Quarterly reporting on the status of all archiving efforts is required for each discipline node within the PDS. The PDS Project Office tracks all funded archiving projects every three months and reports this information to the NASA program officers. The PDS uses a 3-tier (“stop light chart”) system of denoting progress on archiving projects:

Green – denotes everything is on track and there are no issues impeding the successful archiving of the project with the PDS. Data providers are in good communication with their node representative, and work is progressing in a positive manner.

Yellow – denotes that there may be some issues with the archiving process, or unforeseen impediments have stalled the production of the archive products. Data providers may be in good contact with their node representative or not. Progress on the archiving effort tends to be slower in this case, and there may be valid reasons for this. Communication with the node representative allows for accurate planning and adjustments to schedules as needed. In general there are no major concerns with the project, but there could be delays.

Red – denotes major issues with the archiving effort. Data providers may be unresponsive to node representatives or major issues with collecting and/or preparing the data could jeopardize the successful completion of the archiving responsibilities within the timeframe of the award. The PDS Project Office can and will step in to help facilitate communication between the nodes and the data providers when a project is flagged as red. NASA Headquarters program officers are also notified. Again, communication is key to the success of your archiving efforts, and it is essential that data providers notify the PDS as well as their program officers when there are significant delays in the archiving process.

In general most projects will be denoted *Yellow* until the majority of the work is done and the project is progressing toward the Peer Review phase. Brand new projects that are just beginning work or have yet to contact the PDS will also tend to be marked as *Yellow* in the ATM tracking system. If a project has gone a significant period of time without contact, the node will turn the status *Red*, and the PDS Project Office will likely contact the data provider and/or the NASA program officer. As a reminder, it is always a good idea to alert your node representative as soon as a problem arises so that appropriate adjustments can be made.

Editing XML Files

All labels under the PDS4 Archiving Standard are expressed as XML files. XML files can be edited using any text editor. Text editors are not word processors (so Microsoft Word is not a

good tool for this). There are many types of basic text editors, and most modern versions of many common text editors are *XML-aware*, meaning the editor recognizes that the file is XML and will color-code and preserve the hierarchy present in the XML file. These are good for general editing – editing values within XML tags is straightforward. However, if you are worried about validating values within the XML tags, using an XML editor (e.g, oXygen) may be desirable. A variety of XML editors are available online for cost and/or free. PDS staff will typically use oXygen, which is available through subscription. Academic institutions may have site licenses, so check with your home institution, if necessary. oXygen also has a free 30-day trial period if you like to try before you buy. Eclipse is another common editor that may be downloaded for free, but the validation capability has changed over time.

XML files used by PDS have the following extensions:

<i>Labels (product metadata)</i>	<i>.xml</i>
<i>XML Schema (information model)</i>	<i>.xsd</i>
<i>Schematron (rules sets)</i>	<i>.sch</i>

When the ATM node provides initial files for data providers to use, the files will be label *templates*, meaning these are valid labels when filled out properly. These will be completed (or nearly completed) labels (.xml). A couple of important reference files will be listed in every label. These files ensure valid content in every label. **Schema (.xsd)**, are created from a particular version of the PDS4 Standard and provide structure for every label. Alongside these schema are sets of rules for how content of individual tags/fields within the .xml file are validated. **Schematron (.sch)** files provide specific string construction rules, patterns of fields, enumerated lists, etc. that further limit what is valid in a completed XML label (.xml). The pair of these files are always referenced in the opening lines of every XML label (.xml), and the ATM node will provide you with these references when the templates are provided. Every pair of files (.xsd and .sch) will tell our validation software which rules to check when validating the content of your labels. When editing your label templates, basic text editors may not provide any feedback as to adherence to either the schema or schematron rules, allowing potential mistakes in the editing process. Editors designed for XML development and editing provide feedback by accessing the referenced pairs of schema (.xsd) and schematron (.sch) files allowing real-time content/syntax validation. Depending on the complexity of your data there may be multiple pairs of these files to accommodate local data dictionaries.

ARCHIVING CHECKLIST

- Contact the PDS Atmospheres Node to inform them of funding and begin scheduling your archiving effort.
- Provide information about your data to PDS Atmospheres Node – Types of data, processing levels, supporting products.
- PDS Atmospheres Node begins working on tailoring XML Label Templates for your bundle(s).
- Bundle hierarchy structure decided – Layout of the Bundle-Collection-Product structure allows creation of bundle-specific Logical Identifiers (LIDs).
- Create the bundle – Use label templates together with decided structure to create your complete bundle.
- Complete Documentation – Adequate documentation shall be included in the bundle.
- Iterate with PDS Atmospheres Node until your bundle can pass validation.
- Provide a list of potential reviewers to the Atmospheres Node for your peer review.
- PDS Atmospheres Node solicits reviewers for the Peer Review.
- PDS Atmospheres Node conducts the Peer Review online and receives/collates reviews.
- Concluding telecon presents results of the Peer Review to Data Provider(s) and determines schedule for any Lien Resolution.
- Lien Resolution and Validation – Once lien resolution is complete, reviewers are satisfied and validation is complete the data are accepted into the archive.
- Data Harvest and Registration – PDS Atmospheres Node pushes the newly certified bundle to the PDS Central Registry.
- PDS Atmospheres Node creates a landing webpage for distribution of the bundle.
- PDS Atmospheres Node advertises/announces new bundles in PEN.

IMPORTANT NODE CONTACTS

Dr. Nancy Chanover, PI of the NASA PDS Atmospheres Node.

Department of Astronomy, New Mexico State University,
P.O. Box 30001, MSC 4500, Las Cruces, NM 88003-8001
Office: (575)646-2567 email: nchanove@nmsu.edu

Lyle Huber, Data Manager of the NASA PDS Atmospheres Node.

Department of Astronomy, New Mexico State University,
P.O. Box 30001, MSC 4500, Las Cruces, NM 88003-8001
Office: (575)646-1862 email: lhuber@nmsu.edu

Dr. Lynn Neakrase, Senior Research Scientist at NASA PDS Atmospheres Node.

Department of Astronomy, New Mexico State University,
P.O. Box 30001, MSC 4500, Las Cruces, NM 88003-8001
Office: (575)646-2566 email: lneakras@nmsu.edu