VxO Registry Migration Report for the VSPO to NASA Naming Authority Transform

Discussions concerning the naming authorities used for the VxO SPASE metadata registries have occurred frequently over the last few years in our telecoms involving the SPASE Metadata Working Team, SMWT. At the telecom that took place on January 30, 2020, a consensus was reached to apply a transform to the VxO SPASE metadata in order to eliminate the VSPO naming authorities and migrate any resource description from this VxO registry to a new registry with the naming authority NASA. In addition, all SPASE Person resource description appearing in the SMWG registry were also to be moved under the NASA naming authority. The primary motivation for this effort was to remove ambiguity by registering resource descriptor in registries that had accurately and precisely titled naming authorities. The VSPO, for Virtual Space Physics Observatory, and SMWG, for SPASE Metadata Working Group, were adopted years ago. And the usage of VSPO and SMWG slowly became out of sync with the list of other more recently added naming authorities. The newer naming authorities/registries have names that are typically the acronym for the institution responsible for the various facets of Heliophysics and related efforts in the respective countries including project funding, data production, data distribution, etc. And, the list of naming authorities is growing as the effort for producing SPASE descriptions for Heliophysics data products and other resources progresses at institutions around the world. As new naming authorities are added in the future, we encourage the use of appropriately chosen acronyms that follow the spirit of this tenet. The revised VxO naming authority/registry list will include twelve naming authorities: ASWS, ESA, CCMC, CNES, CSSDP, GBO, ISWI, JAXA, NASA, NOAA, NSF, SMWG after completion of the VSPO and SMWG Person to NASA migration. The VSPO Git registry will disappear operationally although a copy will be maintained offline. We also present a scenario where all SMWG SPASE resource description are migrated to the new NASA Git registry. If this plan is approved, the old SMWG registry will not be exposed to the public but a copy of the old metadata collection will be kept offline.

Overview of VxO Registry Contents

Before starting any work to migrate resources, a snapshot of the VxO registry content was obtained by downloading zip files from https://github.com/hpde and then performing simple counts of the SPASE resource descriptions in each of the VxO registries. There were twelve VxO registries that existed when this survey was performed on February 2, 2020. The results of the survey are shown in Table 1. The VxO registry names are listed in the top row of the table while the SPASE resource description types are listed in the leftmost column. Resource type counts are given by VxO in successive columns with systemwide counts listed in the rightmost column. The last row of the table shows total resource counts by VxO. Finally, cells with filled with red highlight zero counts.

Now, let us consider the gross distribution of SPASE resource description counts that are evident in Table 1. Clearly, the table is sparsely filled. And, there are four VxO registries that lack any SPASE resource descriptions: CNES, CSSDP, ISWI. and NSF. These registries are simply acting as place holders for the future. We will ignore these registries for most of the remainder of this document.

SPASE Resource Type	ASWS	CCMC	CNES	CSSDP	ESA	GBO	ISWI	JAXA	NOAA	NSF	SMWG	VSPO	Total
Catalog	0	0	0	0	0	0	0	0	0	0	0	14	14
DisplayData	5	0	0	0	14	46	0	11	3	0	0	145	224
NumericalData	54	7	0	0	171	1175	0	45	300	0	0	2447	4199
Observatory	33	0	0	0	0	0	0	0	0	0	1935	0	1968
Instrument	57	0	0	0	0	0	0	0	0	0	2518	0	2575
Person	1	0	0	0	0	0	0	0	0	0	4920	0	4921
Registry	0	0	0	0	0	0	0	0	0	0	0	0	0
Repository	1	0	0	0	0	0	0	0	0	0	148	0	149
Annotation	0	0	0	0	0	0	0	0	0	0	0	9	9
Document	0	0	0	0	0	0	0	0	0	0	9	0	9
Service	0	0	0	0	0	0	0	0	0	0	25	0	25
	151	7	0	0	185	1221	0	56	303	0	9555	2615	14093

Table 1. A count of SPASE resource descriptions by VxO registry with full VxO registry resource descriptions listed in the rightmost column for the survey performed on February 2, 2020.

Note that the table is divided into five groups based on SPASE resource type. The top group of three resource types includes counts for SPASE Catalog, Display Data, and Numerical Data descriptors. Some number of these resource descriptions are present in all eight relevant VxO registries except the SMWG. This is expected, as by design, only non-data SPASE resource types are supposed to be stored in the SMWG Git registry.

Observatory and Instrument resource descriptions only appear only under ASWS and SMWG. For ASWS, this is again by design as we wish to allow the personnel of the Australian government's Department of Meteorology, Space Weather Service, SWS, to have complete autonomy with regard to their SPASE metadata content. In the spirit of full disclosure, I have added one SPASE Repository description, that for SWS, under the ASWS naming authority so that this registry does indeed stand alone. The old SMWG version of this Repository resource description will need to be deleted. Again, by shifting this single resource description, the ASWS registry becomes completely detangled from the switch of the naming authorities from VSPO and SMWG to NASA. Thus, the ASWS registry can be ignored in much of the remainder of this document.

The VSPO to NASA Naming Authority Migration Task

Migration Task Requirements

The overall impact of the migrating resource descriptions from VSPO and SMWG to NASA is dependent on the five different resource type groupings shown in Table 1. Let us consider the what changes are required to perform the migration on a group by group basis:

- Catalog, Display Data, and Numerical Data: All VSPO data product descriptions need to have their SPASE Resource IDs changed to begin with the NASA naming authority. Also, Prior IDs must be added to store the old VSPO versions of the IDs. The release dates listed in each resource descriptions must be updated as well. Furthermore, any Resource ID that occurs in the SPASE ontology association maps that refer to independent resource descriptions that are now registered under the VSPO or SMWG naming authorities must be updated accordingly. Currently, there are seven types of ontological associations that are mediated by using SPASE resource IDs. See Table 2 for a summary of the association types and number of times that they are exercised within the systemwide set of SPASE resource description stored in the VxO Gits. Even though Resource IDs and Prior IDs instances are independent of association mapping, their counts are listed in Table 2 so that a total count of VxO resource IDs can be compiled. Note that we will use the bolded acronym **AM** to highlight whether association map ID updates are required as a function of SPASE resource type in the remainder of this list.

- Observatory and Instrument: These are mentioned again to emphasize that there are close to 4500 of these SPASE descriptors in the old SMWG registry and that all of these resource descriptor release dates and resource IDs must be updated and prior IDs must be added if they are to be migrated to the NASA naming authority. **AM** changes required.
- Person: 4920 descriptors, only the Release dates and Person Resource IDs need to be updated to swap NASA for SMWG. In fact, the SPASE Person resource is the only SPASE resource type besides SPASE Granule that lacks a SPASE resource header, which contains the Prior ID tag that, if it were to be present, would typically need to be added to capture the old resource ID. We note that the SPASE Granules schema also does not include a resource header but does include prior IDs. The SPASE Metadata Working Team, SMWT, may want to consider adding prior IDs to the person resource in future version of the SPASE data model. **AM** changes are not required for SPASE Person descriptions.
- Repository: Only 148 descriptors are present under the current SMWG registry. They need to updated in much the same way as the Observatory and Instrument descriptors with changes related to Release Date, Resource ID, and Prior ID. **AM** changes are required. Registry descriptions would also be treated the same way but, in fact, no Registry description has ever been committed for registration in the VxO metadata registries. Sometime in the near future a SPASE Registry description must be written for each and every VxO metadata registry.
- Annotation, Document, and Service: Not many of these SPASE resource types exist numbering only 9, 9, and 25, respectively. Release Date, Resource ID, Prior ID, and **AM** changes required.

Count	SPASE ID Type
1490	Association ID
6	Input Resource ID
5082	Instrument ID
2019	Observatory Group ID
2575	Observatory ID
20579	Person ID
6893	Prior ID
5056	Repository ID
14092	Resource ID
57792	Total SPASE ID Count

Table 2. SPASE ID type counts summed across all resource descriptions stored in the VxO Git metadata registries. Note that all Prior IDs already present in the metadata before the proposed VSPO and SMWG to NASA naming authority are not subject to any updating during the migration.

This completes the overview of the VxO SPASE metadata registry content and what is the minimum that must be done in order to update the VxO metadata themselves for the conversion to the NASA naming authority from VSPO and from at least those resource description stored under SMWG/Person.

VSPO to NASA Migration Task Execution

The transition of the naming authority from VSPO to NASA for all SPASE resource descriptions stored in the VSPO Git registry and from SMWG to NASA for all Person resource descriptions stored in the SMWG Git registry were performed by using IDL code. The IDL software includes a main program named nasa_authority_update.pro and a function named spase_xml_tab.pro. The main program executed all the changes required for Release Date and Resource ID changes and to add Prior IDs that store the old versions of the resource IDs that existed before the VSPO to NASA migration. The main code also performed updates to the Resource IDs of the Display Data and Numerical Data descriptions in instances where the cadence tuple, which is the last tuple in the IDs, were not in alignment with typically used cadence values. A few other Resource IDs were also updated. See a description of the update for irregularly formed resource IDs in the fifth and last subsection concerning text normalization. The spase_xml_tab.pro function executed the white space management, especially indenting, as required for uniform text normalization.

This completes the overview of the VxO SPASE metadata registry content and what is the minimum that must be done in order to update the VxO metadata themselves for the conversion to the NASA naming authority from VSPO and from at least those resource description stored under SMWG/Person.

However in fact, much more was also done in order to improve the overall quality of the VxO metadata by applying standards for the textual format of the resource descriptions. The formatting updates give the resource descriptions a uniform appearance. And, they also permit reliable and proper rendering of paragraphs, lists, and tables that are imbedded in the resource descriptions. In the next section of this report, we introduce the topic of text normalization. We also explain why it was important and also timely to perform these additional changes in parallel with the execution of the switch of naming authority from VSPO and SMWG to NASA.

Resource Description Text Normalization

Last year, Reine Chimiak, Todd King, and I worked together to ensure that the SPASE metadata describing Heliophysics data resources met the minimum requirements to search for, extract, and archive data. This effort entailed many tasks to guarantee that all SPASE resource descriptions could be validated, had complete resource ID referential integrity, and that all information and access URLs were accurate and precise. Even though we met these goals, there are still many issues that still affect the VxO metadata. Many violate the SPASE conventions and standards as outlined in documents released through the SMWT including: "A Space and Solar Physics Data Model from the SPASE Consortium", Release Version: 2.3.1, Release Date: November 14, 2019, "Guidelines for Resource ID Formation", Updated: September 24, 2009, and "Report on SPASE Text Normalization and Mark-up", Released: March 11, 2009, Version: 1.0. The URL links for SPASE convention related documentation see the reference list located at the end of this report.

Textual errors that exist in our XML resource descriptions impact the overall quality of the back end metadata stored in the VxO registries and can have an influence on the appearance, reliability, and user experience concerning front end Heliophysics related web sites such as the Heliophysics Data Portal, HDP. The different kinds of SPASE document textual errors are given in the list below, which is ordered from the most fundamental of XML document preparation errors to those errors that are more narrowly related to SPASE text normalization conventions or adopted standards for XML metadata text population:

- Unpermitted usage of the XML special symbols (&, <, >, ', ").
- Unpermitted usage of non UTF-8 characters.
- SPASE XML text normalization and mark-up errors.
- SPASE Resource ID issues.
- SPASE 2.3.1 updates.

In the next set of subsections, we will address each of these problems in this order by introducing the underlying issues and explain how the metadata can be properly populated. Then, we explain how the stored metadata can be updated so that the issues are resolved. An overarching goal is ensure that the metadata content is consistent with quality control standards and that the metadata content also closely adheres to a uniform style sheet.

If the suggested changes are enacted, the appearance of the metadata would be consistent across different applications whether they be either something primitive such as the UNIX vi editor or something more advanced such as the version as rendered by the SPASE Metadata Working Group, SMWG, Registry Explorer, found at: http://www.spase-group.org/registry/explorer/. Note that the Registry Explorer can be used to examine VxO resource descriptions located in the registries that provide the backend content that enables the functionality of the public version of the Heliophysics data portal. A second Registry Explorer is also available for examining draft versions of the VxO metadata content. The draft Registry Explorer, located at http://spase-group.org/draft/explorer/, will be used by the mission teams when reviewing their SPASE data product descriptions. Active links to these two VxO registry explorers are listed following the reference section of this report.

XML Special Character Issues

Introduction and Proper Usage

The XML specification includes five predeclared entities to represent so-called special characters. If these special characters are included in an XML document, they must be escaped in order for the XML document to be well formed. Please see Table 3 for information concerning all the XML special characters, their XML context, and widely adopted conventions for their usage. Again in the spirit of best practices, we have chosen to use the predeclared entities throughout SPASE documents with only two exceptions. These exceptions correspond to the first two lines in all the reformatted SPASE XML documents: the XML prolog and the SPASE opening tag. Both of these lines employ the use of quotes, either single or double, to delimit XML attribute values or in a manner that is completely in line with standard XML conventions. See Figure 1, which shows a randomly chosen SPASE Person resource description, shows example instances of these two lines.

Symbol Name	Symbol	Entity	Usage	Entity Escaping Notes
Ampersand	&	&	First character of an entity reference	Escaping mandatory.
Less than	<	<	First character of an XML tag	Escaping mandatory.
Greater than	>	>	Final character of an XML tag	Escaping is always strongly advised though the requirement depends on context.
Apostrophe	'	'	Single quoted XML attribute definitions	Mandatory escaping for XML attribute text within single quote definitions. In all other scenarios, escaping is always strongly advised.
Quote	"	&quot'	Double quoted XML attribute definitions	Mandatory escaping for XML attribute text within double quote definitions. In all other scenarios, escaping is always strongly advised.

Table 3. The XML five special characters including symbol names, symbols, predeclared entities, usages, and entity escaping notes. The conventions we recommend the use of the predeclared entities are listed in the last column.

The ampersand character, &, is the first of the five special characters listed in Table 3. As you will see, XML parsers cannot resolve character references, which allow non-UTF-8 characters to be properly rendered, unless the ampersand character was reserved for what is called entity escaping. The rationale behind ampersand usage is explained in more detail in the next paragraph. As the similar sorts of justifications apply to all five of the predeclared special characters, we will defer explanations concerning the other four special characters unless clarifications are requested.

Ampersand Special Character and Character References

The ampersand character is used as the first letter of a character reference, which is a common mark-up construct, in XML documents. Character references are terminated by a semi-colon character. There are two types of XML character references: numeric character references and character entity references. To date, the SPASE resource descriptions stored in the VxO metadata registries only use numeric character references. Numeric character references denote specific characters or symbols by their Unicode code points by using the format: &#nnnnn; where nnnnn is the code point in decimal form. Note that a completely equivalent set hexadecimal character references also exists the uses the format: &#xhhhhh; where hhhhh represent a string containing characters [0-9A-F]. Indeed, the VxO metadata contain both decimal and hexadecimal character references. Decimal and hexadecimal code points may be composed of any number of characters and code point text strings may include any number of leading zeroes. However, we recommend to adopt a standard that discourages the use of leading zeroes in code points in order to minimize the number of unique character references that appear in SPASE resource descriptions.

Example of a Character References

The character reference for the symbol, °, is one of the most prevalent used in our SPASE resource descriptions as it can be used to denote degrees rather they relate to angles or temperatures. We will use the degree symbol as an example for constructing character references. The decimal code point for the degree symbol is 176 and the full character reference is °. The hexadecimal code point and character references for the degree symbol are respectively: B0 and °. Of course, the decimal value of 176 is equal to the hex value of B0.

It may be wise to consider whether SPASE should adopt a convention concerning a preference for using either decimal or hexadecimal character references. The motivation for this suggestion is again to minimize the number of unique character references that appear in the VxO metadata. We note that all modern systems support the use of decimal and hexadecimal character references. And, there are arguments on both sides favoring one preference over the other. For instance, some of decimal character references are easier to remember relative to their hexadecimal equivalents but it appears that most programming languages and character tables use hexadecimal values.

Issue Resolution and Impact

A search on all VxO files was executed via the use of simple UNIX awk, grep, and sed commands to find resource descriptions that contained illegal use cases for all five of the XML special characters. Ultimately, the offensive use cases were edited by using a combination of one-off style UNIX shell scripts or by hand edits. Careful review of each affected files were performed to ensure none of the files were corrupted. Affected file counts were not collected while performing edits to correct for un allowed special character instances but XML rendering would have been degraded.

Non UTF-8 Character Issues

Introduction and Details

The second sentence appearing in Section 3.5.1 of "A Space and Solar Physics Data Model from the SPASE Consortium", Release Version: 2.3.1 stipulates the standard for which set of characters can be utilized within our SPASE resource descriptions. The second sentence states as follows:

To ensure system portability text values in SPASE are sequences of alphanumeric one byte UTF-8 (US_ASCII) characters with white space preserved.

This sentence clearly states that only UTF-8 (US_ASCII) characters may appear in the text of our SPASE documents. However, there are about three dozen non UTF-8 found interspersed within registered VxO resource descriptions. See each of the of non UTF-8 characters, their descriptions, and replacement numerical character entities in Table 4. A few Unicode lookup web sites that we have used over the course of the migration effort appear just after the reference list below.

The control M or cntl-M, ^M, carriage return also appears in Table 4. This is one of the so-called hidden characters that can be found in a generalized ASCII data file. While cntl-M may seem to be harmless but it is indeed a nefarious character. The presence of cntl-M in an ASCII file means that a file is no longer a simple ASCII file at least in the context of the UNIX working environment. Instead, the file is a DOS file and suggests it was created or modified by some application running under a DOS operating system. There are three reasons we should prohibit cntl-M from appearing our SPASE resource descriptions: cntl-M is a shortcut key that is application dependent, it is a carriage return without a line feed that can alter the appearance of something as simple as the listing of file contents, and it is simply an extraneous and unneeded character in the context of XML documents. In fact, SPASE text normalization rules, see the next subsection, requires that every line in a SPASE resource description must end in a new line character. In fact, effort has been wasted while for instance trying to compare two seemingly identical SPASE documents that contained cntl-M characters.

Char.	Non UTF-8 Character Description	Replacement
¸	Cedilla	¸
Å	Angstrom symbol	Å
Å	Latin capital letter A with ring above	Å
Ö	Latin capital letter O with diaeresis	Ö
Ø	Latin capital letter O with stroke	Ø
é	Latin small letter e with acute	é
ê	Latin small letter e with circumflex	ê
ñ	Latin small letter n with tilde	ñ
ø	Latin small letter o with stroke	ø
ü	Latin small letter u with diaeresis	ü
Δ	Greek capital letter delta	Δ
Σ	Greek capital letter sigma	Σ
Φ	Greek capital letter phi	ϕ
β	Greek small letter beta	β
θ	Greek small letter theta	θ
π	Greek small letter pi	π
σ	Greek small letter sigma	σ
μ	Greek small letter mu	μ
µ	Micro symbol	µ

Char.	Non UTF-8 Character Description	Replacement
^M	Cntl-M, carriage return, no line feed	null string
	No Break Space	space char.
°	Degree symbol	°
±	Plus-minus symbol	±
×	Multiplication symbol	×
⨯	Vector or cross product symbol	⨯
∼	Tilde operator	∼
≤	Less than or equal to symbol	≤
≥	Greater than or equal to symbol	≥
–	En dash	- hyphen
—	Em dash	- hyphen
−	Minus sign	- hyphen
‘	Left single quotation mark	'
’	Right single quotation mark	'
′	Prime symbol	'
“	Left double quotation mark	"
”	Right double quotation mark	"
″	Double prime symbol	"

Table 4. Non UTF-8 and non-printable characters that were present in the resource descriptions stored in the VxO Git registries on February 1, 2020. The table is ordered by listing non UTF-8 Latin characters followed by Greek characters in the left hand set of columns and the non-printable characters, ctrl-M and no break space, followed by various mathematical symbols, hyphen related characters, and quotation mark related characters in right hand set of columns.

Issue Resolution and Impact

The non UTF-8 characters have been replaced automatically on a file by file basis via the spawning of a UNIX sed command within the IDL main program called nasa_authority_update.pro. The spawned command utilizes a stream editing sed file named fix_non_utf8.sed that provided the basis for constructing Table 4. Note that the non UTF-8 Latin characters appearing in Table 4 are could be replaced by using their simplified Latin character equivalents. However, we have chosen to use the set of Latin characters extended that includes the variety of accented letters that are used in non-English languages. There were two orthographic ligatures for "fi" and "fl" where two letters were found joined as a single glyph. The occurrences of ligatures were edited out by hand early in the course of the overall VxO naming authority migration effort.

The impact of non UTF-8 characters and of cntl-M instances on the SPASE resource descriptions registered in the VxO Gits is summarized in Table 5. The leftmost column in the table lists the top two tuples present in the directory tree under which the affected files are located. The first tuple is, of course, the naming authority, of which only ASWS, SMWG, and VSPO are present. The second tuple corresponds to the SPASE data model resource type. One can see that the "data" resource types Display Data and Numerical Data have been affected along with the "ancillary" resource types Document, Instrument, Observatory, Person, Repository, and Service. The number of resource descriptions by VxO Git directory are listed for non UTF-8 and cntl-M instances in the two middle columns of the table. The rightmost column lists the number of resource descriptions affected by the union of instances for both non UTF-8 and cntl-M issues. There are 402 SPASE resource descriptions affected and these represented under 3% of the total number of 14,092 files registered in the VxO Gits. Still, the method for resolving these issues is simple to perform and perhaps more importantly easy to implement if we decide to construct an enhanced quality control and validation pipeline to inspect metadata prior to registration of SPASE resource descriptions. Any submitted metadata files that fail to meet certain criteria could be returned metadata contact with suggested updates produced during the pipeline testing.

VxO Git Directory	Non UTF-8 Count	Cntl-M Count	Union Count
ASWS/DisplayData	0	2	2
ASWS/NumericalData	24	38	43
ASWS/Observatory	1	30	30
SMWG/Document	0	2	2
SMWG/Instrument	50	88	136
SMWG/Observatory	30	16	41
SMWG/Person	9	45	49
SMWG/Repository	1	7	8
SMWG/Service	0	4	4
VSPO/DisplayData	1	15	16
VSPO/NumericalData	58	21	71
Total Count	174	268	402

Table 5. Counts of descriptions by VxO registry SPASE resource type that contained either non UTF-8 characters or cntl-M instances listed in columns two and three, respectively. The last column shows the counts for the union of the files affected by either of these issues or both: non UTF-8 ∪ cntl-M. The counts are for the VxO registry content found on Feb. 1, 2020.

SPASE XML Text Normalization and Mark-up Issues

Introduction and Details

Now, let us consider the whole first paragraph from Section 3.5.1 of "A Space and Solar Physics Data Model from the SPASE Consortium", Release Version: 2.3.1. This paragraph introduces text normalization in the context of the SPASE data model:

While descriptive text may be brief, some formatting of the text may be necessary to convey the necessary information, for example, multiple paragraphs or nested lists. To ensure system portability text values in SPASE are sequences of alphanumeric one byte UTF-8 (US_ASCII) characters with white space preserved. When text is displayed in some applications (a web browser is the best example) a strict preservation of white space may not result in a desirable presentation. Also, to make the metadata more human readable (for example in XML) additional white space may be introduced in the form of indentation. If strictly preserved, this could result in an undesirable presentation. To allow an author to express a preferred layout for the text, a special set of text "mark-up" rules are defined. The layout can then be determined by normalizing the text and applying a simple set of interpretation rules.

To understand the significance of text normalization, we need to introduce terms that describe the distribution of white space in XML documents. First of all, white space refers to the collection of spaces (as in space bar), tabs, and new line characters. XML documents typically contain both "significant" and "insignificant" white space. Consider the following examples for a person's name within a SPASE Person resource description:

<Name>Lawrence Gordon Tesler</Name>

<Name>LawrenceGordonTesler</Name>

Any IDL processing routine is obligated to recognize and render the white space that is present in the former example, which means that the two examples are fundamental different text elements. The two space characters, which could just as well be tabs, correspond to significant white space content. Next consider an example that demonstrates insignificant white space:

......<Name>...Lawrence..Gordon...Tesler.....</Name>...

Dots are used to are used to represent all instances of white space whether they be either spaces or tabs. All of the dots correspond to insignificant white space because their presence has no bearing on the final rendering of the text element as the two white spaces highlighted in yellow are sufficient to separate the first, middle, and last names shown in the example above. For this reason, all superfluous white can be deleted without affecting the text element content. We propose that insignificant white space should be compressed out prior to registering SPASE resource description in the VxO Gits with two exceptions described in the next two paragraphs. But before going forward, we would like to propose that the usage of tab characters should be prohibited or at least discouraged in SPASE documents because tab characters are not rendered consistently across applications such as when listing or editing files and perhaps even when files are rendered by XML processors. Now let us return to the topic of our suggested white space usage exceptions.

1. First, we have chosen to include leading white space to indent the remastered version of the SPASE descriptors. The rule we have adopted for leading white space uses multiples of three space characters for the purpose of indenting the descriptors and to reveal the hierarchical level relationship amongst all of the XML text elements. Figure 1 shows a randomly chosen example of a SPASE Person description organized by using this indenting scheme. The SPASE opening tags appear at the zeroth level and thus are not indented. The Version text element and Person opening and closing tags appear correspond to first level SPASE schema terms and thus are indented by three space characters. All other text elements appear are second level terms and are thus indented by six spaces., We believe that all of the organization and structure of the XML content contained in resource descriptions is easier to comprehend once raw SPASE XML are indented by using such a set pattern for indenting. The overall state of white space presence in the VxO registries captured in the February 1, 2020 snapshot is best described as haphazard at best and overall appearance of the raw metadata is frankly quite unprofessional. There is great room for improvement and we have chosen a style sheet that is consistent with SPASE output generated by ADAPT software. We recognize that there are many examples of "XML beatify" applications and other that SPASE team members may wish to provide input on what style sheet rules that are eventually adopted.

2. Second, we have also chosen to allow multiple white space text strings in text elements that include mark-up for tables in order to maintain the alignment of columns within the table, see Figure 2. The figure shows the description text element contained in the resource header for

the SPASE Numerical Data document that describes Voyager 1 magnetometer CDF data product sampled at a cadence of 1.92 s from the March 1979 Jupiter encounter. The metadata shown in this example comply with the text normalization rules and the text interpretation rules as outlined in Section 3.5.1.1 and Section 3.5.1.2, respectively, of "A Space and Solar Physics Data Model from the SPASE Consortium", Release Version: 2.3.1, and in the document named "Report on SPASE Text Normalization and Mark-up". Paraphrasing the second rule from Section 3.5.1.1 on text normalization, we can state that all text not preceded by a SPASE XML tag must be left justified with no leading white space. One can see that the first line appearing in Figure 2, which contains the opening tag of the SPASE Description element, is the sole line that contains any leading white space. And, this line is indent by six white spaces, not using any tabs, as it should be for a second level SPASE XML text element per the rules described in the paragraph just above.

We will not repeat all of the rules listed in Section 3.5.1.2 of the SPASE data model description but suffice to say that the example shown in Figure 2 complies with all of the conventions set for both text normalization and text mark-up. Note that we have chosen to employ one additional rule that we could adopt as a SPASE standard or best practice. The rule requires the all of the text in a paragraph should appear in a single line of text within marked up metadata. We do this so that raw display of the metadata, such as the metadata as seen in a text editor, would mimic that presented by rendered displays of the metadata as closely as possible.

Issue Resolution and Impact

Text normalization and text mark-up compliance issues were the most difficult to overcome. There are some resource descriptions that are currently registered in that VxO Git registries where it was clear that the author of the resource description had at least some intent to structure the raw data to comply with SPASE text mark-rules. However as they are currently registered, it was not at all clear that the author intent would become realized. In other instances, it seemed that the metadata was copied from other sources without regard for their impacts on the appearance of the metadata following XML render processing. In either case, we attempted to interpret the author's intent and to effect changes to realize on that intent. UNIX shell scripts were improvised and used whenever possible to execute the changes required for text mark-up compliance. However, we still found it necessary to hand edit over two thousand files as it appeared to be too difficult or too time consuming to design and write automated software to perform the needed edits. Once the updates were completed, UNIX shell scripts were employed whenever possible to ensure the accuracy of the edits whether they were done via automation or performed by hand.

The impact of text normalization and text mark-up issues on SPASE XML content is rather pervasive. Table 6 lists the number and percentages of resource descriptions that were hand edited for each of twenty four VxO registry directories, which are defined as any combination of naming authority and SPASE resource type. Almost one sixth of all of the VxO resource descriptions were updated at least to some extent via hand editing. Only, the ASWS/Person, SMWG/Person, ESA/DisplayData, and GBO/NumericalData registry directories were left untouched or nearly so. The SPASE Person resource description schema is quite simply constructed so the statistics are not too surprising. Also, a super majority of the GBO Numerical Data were generated by ADAPT software that are built to adhere to SPASE text normalization and mark-up conventions, so again not surprising. The stats listed in Table 6 do not reflect the counts of automated editing updates that were made to 680 SMWG/Instrument and 608 SMWG/Observatory resource descriptions. If these counts were added into the table, the calculated file edit percentages would exceed just over 50% for both of these SMWG registry directories.

The vast majority of the edits were relatively simple but some were rather extensive. Figures 2 and 3 show an example of the benefits of updating the metadata for one of the most impacted files. Both figures show the SPASE Description text element from the SPASE Resource Header of the Numerical Data description for Voyager 1 Planetary Radio Astronomy, PRA, low band receiver 6 s cadence data from the 1979 Jupiter encounter. We want to give the good news first so the text normalized and marked up resource description is displayed in Figure 2. The description text found prior to hand editing is displayed in Figure 3. Clearly, the update has the desired positive effect!

VxO Registry Directory	Total Count	Edit Count	Edit %
ASWS/DisplayData	5	5	100.0%
ASWS/Instrument	57	21	36.8%
ASWS/NumericalData	54	54	100.0%
ASWS/Observatory	33	33	100.0%
ASWS/Person	1	0	0.0%
ASWS/Repository	1	1	100.0%
CCMC/NumericalData	7	3	42.9%
ESA/DisplayData	14	0	0.0%
ESA/NumericalData	171	62	36.3%
GBO/DisplayData	46	6	13.0%
GBO/NumericalData	1175	16	1.4%
JAXA/DisplayData	11	6	54.5%
JAXA/NumericalData	45	26	57.8%
NOAA/DisplayData	3	0	0.0%
NOAA/NumericalData	300	62	20.7%
SMWG/Document	9	9	100.0%
SMWG/Instrument	2518	621	24.7%
SMWG/Observatory	1935	309	16.0%
SMWG/Person	4920	142	2.9%
SMWG/Repository	147	48	32.4%
SMWG/Service	25	14	56.0%
VSPO/Annotation	9	3	33.3%
VSPO/Catalog	14	7	50.0%
VSPO/DisplayData	145	54	37.2%
VSPO/NumericalData	2447	843	34.5%
VxO Total Count	14092	2345	16.6%

Table 6. The table lists VxO registry directories (NamingAuthority/ResourceType) in the leftmost column. The next three columns list the total number of files, the number of hand edited files, and the percentage of hand edited files by registry directory. The edit percentages are thus equal to the number of files listed in column three divided by the files listed in column two.

Resource ID Formation Conformity Issues

Introduction and Details

The formation rules for SPASE resource IDs are summarized in the document "Guidelines for Resource ID Formation". The most generalized form of a SPASE resource ID is:

spase://NamingAuthority/path

The document listed above gives guidance for defining the "path" portion of the resource IDs for only SPASE Person resources, SPASE Granule resources, and two of the data related SPASE resources: Display Data, and Numerical Data. The ID formation guidelines for the remaining SPASE resource types are not clearly defined or not addressed at al. We suggest that the SPASE Resource ID formation rules need to be extended and the this topic should be addressed by the SMWT in the near future.

In the current context though we wish to focus on the set of Display Data and Numerical Data resource IDs of the files stored in the VxO Git registries. The guideline document offers that these resource IDs should conform with the following pattern:

spase://NameAuthority/ResourceType/Project/Observatory/InstrumentType/Cadence

However, there are many cases where the cadence values are missing entirely. In other Resource IDs, the cadences do appear but not as the last and separate tuple of the ID path. That is, there is no slash character that separates the cadence from the preceding portion of the resource ID path.

We have also noticed that there is no guideline for assigning cadence values except that they must conform with ISO 8601 standards for representing time durations. The ISO 8601 standard is presented in detail in Wikipedia web page: https://en.wikipedia.org/wiki/ISO_8601. The issue which we wish to resolve in this paragraph relates to promotion of uniform cadence assignment. To understand why this is necessary, consider that SPASE resource descriptions of data product that have one hour time resolution has been expressed by using four different cadence test strings that are all consistent with the ISO 8601 standard. These are PT3600.00S, PT3600S, PT60M, and PT1H. We suggest that a standard should be set. And, we argue that common sense choices can be made. First, we suggest that the precision of a cadence has no overriding need to be expressed in the resource ID; so eliminate PT3600.00S from the list. Next, we suggest that the most simple form of the cadence should be favored over more verbose options, thus strike PT3600S and PT60M. So, this leads one to choose PT1H to define the cadence in SPASE Resource IDs for data with one hour cadence.

In order to understand this issue more completely, please refer to Table 7. This look up table shows how one can achieve uniformity of cadence assignment by replacing sets of old ID tuple values a single replacement cadence value. Let us consider one more example. Table 7 lists seven different ways that cadence has been designated in SPASE Resource IDs when describing data with daily time resolution: 24HR, Daily, P1D, PT1D, PT24H, PT86400.00S, and PT86400S. We suggest that there is absolutely no reason to permit such diversity. The most reasonable cadence value is P1D.

Old ID Tuples	New ID Tuples
PT0000028S	PT0.000028S
PT7.8125MS	PT0.0078125S
PT1.0S	PT1S
PT1s	PT1S
PT2.50S	PT2.5S
PT03S	PT3S
PT3.00S	PT3S
EDB.PT3S	EditorB/PT3S
full_PT3S	Full/PT3S
thumb_PT3S	Thumb/PT3S
PT4.0S	PT4S
PT5.00S	PT5S
PT06S	PT6S
PT10.00S	PT10S
L2_PT12S	L2/PT12S
PT15.00S	PT15S
PT24.00S	PT24S
PT30.00S	PT30S
PT01M	PT1M
PT60.00S	PT1M
PT60s	PT1M
PT60S	PT1M
AE	AE/PT1M
PT120S	PT2M
PT1M45S	PT1.75M
124S	PT124S
PT2M30S	PT2.5M
PT168.00S	PT1.8M
PT180.00S	PT3M
PT3M12S	PT3.2M
PT240S	PT4M
PT05M	PT5M
PT300S	PT5M
PT6M24S	PT6.4M
PT600.00S	PT10M
PT600S	PT10M
PT12M48S	PT12.8M
PT900.00S	PT15M
PT900S	PT15M
PT1200S	PT20M
PT1800S	PT30M

Old ID Tuples	New ID Tuples
PT3600.00S	PT1H
PT3600S	PT1H
PT60M	PT1H
PT1HR	PT1H
DST	Dst/PT1H
Electrons_PT1H	Electrons/PT1H
Heavies_PT1H	Heavies/PT1H
Helium_PT1H	Helium/PT1H
Protons_PT1H	Protons/PT1H
PT120M	PT2H
QL.PNG.PT120M	Quicklook/PNG/PT2H
Kp.Ap	KpAp/PT3H
PT3H30M	PT3.5H
24HR	P1D
Daily	P1D
PT1D	P1D
PT24H	P1D
PT86400.00S	P1D
PT86400S	P1D
DailyBrowseImages	BrowseImages/P1D
DailyBrowsePlots	BrowsePlots/P1D
PT72H	P3D
CarringtonPlots	CarringtonPlots/P14D
27-Day	P27D
27DAY	P27D
DATA	Data
diagonal	Diagonal
H-ALPHA	H-Alpha
H-alpha	H-Alpha
IMAGES	Images
Level-zero	Level0
Maps-PNG	Maps/PNG
Maps-TXT	Maps/TXT
MOVIES	Movies
PLOTS	Plots
Solar.Spectra	SolarSpectra
Electron_Density_Profiles_PS	ElectronDensityProfiles/Postscript
L3_Thermospheric_O-N2	Level3/Thermospheric/O-N2/IDL_Save
O-N2_Summary_GIF	Level3/Thermospheric/O_N2/TEC/GIF
Thermospheric_O-N2_PS	Level3/Thermospheric/O_N2/Postscript

Table 7. SPASE Resource ID tuple look up table. The left hand set of two columns show pairs of old and replacement tuples for resource IDs with a cadence of less than one hour. The right had set of two columns show pairs of old and replacement tuples for resource IDs with a cadence between one hour and 27 days at the top. The second and third boxes shown at right list ID tuples we suggest need updated. See the text for more detail.

Other resource IDs issues we want to address are slightly more diverse in nature. Consult the parts of the lookup table enclosed within the second and third boxes appearing on the right hand side of Table 7.

First consider the ID tuples in the second box. For these, we wish to discourage the use of all caps in resource IDs unless the capitalized text string corresponds to a known acronym. We suggest that regular words that appear in IDs should only have their first letter capitalized. Thus use Data not DATA, H-Alpha not H-ALPHA, Images not IMAGES, Movies not MOVIES, and Plots not PLOTS. Also, we have typically expressed a sequence of words in SPASE Resource ID tuples by capitalizing the first letter of all the words and then compressing out the white space so that a word phrase such as electron density profiles would then be expressed as ElectronDensityProfiles. We also recommend that dots should not be used between words, i.e. Solar.Spectra should be replaced by SolarSpectra. The word phrase format described here is already commonly used for the SPASE Resource IDs of many data product descriptions that are stored in the VxO metadata registries. For completeness, we point out that the formation convention for SPASE Person resource IDs does employ the use of dots. Dots are used to delineate the separate parts of a name as seen in Figure 1.

Finally, we suggest that it makes sense to encourage that Resource IDs maintain conformity for data products that are similar in nature and especially for those from the same mission and/or instrument. For instance, consider the examples of old and new Resource ID tuples listed in the third box on the right hand side of Table 7. All of these examples are suggested updates for SPASE Resource IDs that have been assigned to TIMED GUVI data product resource descriptions.

Issue Resolution and Impact

The overall goal of addressing this issue is to promote uniform SPASE resource ID construction across all of the VxO registries and to do this prior to the assignment of DOIs for the landing pages of the VxO data product resource descriptions. And, we have updated 397 SPASE Resource IDs by substituting what we consider the poorly formed portions of the ID paths with replacement text strings found via use of the Table 7 look up. Prior IDs have been added to the SPASE descriptors in order to capture the old SPASE Resource IDs in accordance with or standard practice. We cannot state that issues concerning standardization of SPASE Resource IDs is complete even after the actions that we have taken. However, the steps we have accomplished does get us closer to resolving this important issue.

SPASE Version Updating Issues

Introduction and Details

Issue Resolution and Impact

SPASE Version	Count Before	Count After	Resource Type After
1.2.0	62	62	Repository
1.2.2	3	3	Repository
1.3.0	4	4	Repository
1.3.4	1	1	Repository
2.0.0	522	0	-
2.0.1	375	0	-
2.0.2	14	0	-
2.0.3	14	0	-
2.1.0	37	0	-
2.2.0	5163	566	Observatory
2.2.1	836	75	Observatory
2.2.2	1241	42	Observatory
2.2.3	217	0	-
2.2.4	25	15	Observatory
2.2.6	220	1	Observatory
2.2.8	772	9	Observatory
2.2.9	1282	0	-
2.3.0	3291	0	-
2.3.1	14	13315	All other 2.3.1 Files

The overall goal of addressing this issue is to promote uniform SPASE resource ID construction across all of the VxO registries and to do this prior to the assignment of DOIs for the landing pages of the VxO data product resource descriptions. We have updated the SPASE resource IDs values of 397 resource IDs by replacing the poorly formed portions of the ID paths via use of a look up table containing pairs of old tuples and new or replacement tuples, see Table 4 below.

```xml
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>

<Spase xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.spase-group.org/data/schema" xsi:schemaLocation="http://www.spase-group.org/data/schema http://www.spase-group.org/data/schema/spase-2_3_1.xsd" lang="en">

   <Version>2.3.1</Version>

   <Person>

      <ResourceID>spase://NASA/Person/Lee.Frost.Bargatze</ResourceID>

      <ReleaseDate>2020-02-18T12:34:56.000Z</ReleaseDate>

      <PersonName>Lee Frost Bargatze</PersonName>

      <OrganizationName>Institute of Geophysics and Planetary Physics and the Department of Earth and Space Sciences, University of California, Los Angeles</OrganizationName>

      <Address>Institute of Geophysics and Planetary Physics, University of California Los Angeles, Box 951567, Los Angeles, CA 90095-1567</Address>

      <Email>lfb@igpp.ucla.edu</Email>

      <PhoneNumber>+1 310 206 6073</PhoneNumber>

   </Person>

</Spase>
```


Figure 1. A random example of a SPASE Person resource description. The description exhibit indenting performed by using 3 white spaces × N, where N is the hierarchal level of the XML text element defined line by line. The first line of this resource description is the XML type declaration or prolog. The second is the SPASE opening tag line with all of the text listed with no intermediate character returns. The quotes that appear within the tag boundaries, that is between the less than and the greater than symbols, are permitted and do not require entity escaping and thus the predeclared entity " does not appear. The text "2_3_1" shown near the end of SPASE opening tag line changes to match the SPASE version number, which is listed as 2.3.1 in the example. Otherwise, the format of the text appearing in the SPASE opening tag line is uniform across the full complement of reformatted VxO resource descriptions.

```
<Description>(Description based on material from VG1_PRA_JUP_HRES_DS.CAT) Voyager 1 Radio Astronomy (PRA) data from the Jupiter encounter (1979-01-06 to 1979-04-13). The data set provides 6 second high resolution lowband radio mean power data. The data are provided for 70 instrument channels, covering 1.2 to 1326.0 kHz. This data set (VG1-J-PRA-3-RDR-LOWBAND-6SEC-V1.0) contains data acquired by the Voyager-1 Planetary Radio Astronomy (PRA) instrument during the Jupiter encounter. The bounding time interval set for most Voyager 1 Jupiter PDS data sets is the Voyager project defined 'far encounter' mission phase boundary (1979-02-28 to 1979-03-22). Since, however, the PRA instrument is able to observe planetary phenomenon at much larger ranges than other fields and particles experiments, this boundary is artificial with respect to PRA. Hence, PRA lowband data provided here cover the entire Jupiter Encounter Phase (1979-01-06 to 1979-04-13). Data from beyond the far encounter interval is contained in the cruise data archive which is available from the NSSDC. VG1-J-PRA-3-RDR-LOWBAND-6SEC-V1.0 contains data at the highest time resolution possible during normal operations. The normal mode of PRA operations during the planetary encounters was to sweep through the two radio receiver bands, high band (40.5 to 1.5 MHz in 128 channels spaced 0.3072 MHz apart) and low band (1326.0 to 1.2 kHz in 70 channels spaced 19.2 kHz apart) in a period of 6 seconds. The receivers measured, on alternate samples, the left hand circular and right hand circular (radio definition) power.

* Measured Parameters
* ==================

The data here are from the low frequency receiver band and are 'packaged' into spacecraft major frame records. Each major frame is 48 seconds long or eight sweeps through the PRA

* Bit Value(s)
- 0: 15 dB Attenuator in use when equal to 1
- 1: 30 dB Attenuator in use when equal to 1
- 2: 45 dB Attenuator in use when equal to 1
- 9 and 10 together: Polarization of first Channel sampled (1326.0 kHz) according to the Scheme:

+-------------------------------------------------------+
|                  | Value Bit 10 = 0 | Value Bit 10 = 1 |
|-------------------------------------------------------|
| Value Bit 9 = 0 | R                 | L                |
| Value Bit 9 = 1 | L                 | R                |
+-------------------------------------------------------+

Polarization at successively lower frequencies is opposite to the frequency above it, i.e. either a LRLR or an RLRL pattern. Successive 6-second sweeps start on the opposite

* Data Coverage
* ============

The data are stored as 4 ASCII tables (.TAB), each accompanied with a PDS label file (.LBL) which describes properties of the data file. Data cover the following time intervals: Volume ID: VGPR_1201

+---------------------------------------------------------------------------+
| Filename     | Records | Start                    | Stop                     |
|---------------------------------------------------------------------------|
| PRA I.TAB    | 35569   | 1979-01-06T00:00:34.000Z | 1979-01-30T23:59:47.000Z |
| PRA II.TAB   | 39493   | 1979-01-31T00:00:35.000Z | 1979-02-25T23:59:47.000Z |
| PRA III.TAB  | 41371   | 1979-02-26T00:00:35.000Z | 1979-03-22T23:59:56.000Z |
| PRA IV.TAB   | 24587   | 1979-03-23T00:00:44.000Z | 1979-04-13T23:59:08.000Z |
+---------------------------------------------------------------------------+

* Confidence Level Overview
* ========================

The accuracy of calibration in the PRA low band is approximately 2 dB, except at frequencies below 100 kHz where it is somewhat worse. Interference from the Voyager power subsystem is a major problem to the PRA instrument, affecting many of the 70 low band channels. This interference manifests itself by abrupt changes in background levels. Some channels, notably 136 and 193 kHz, are almost always affected, whereas, others are only affected for short intervals. Usually, this interference is only a problem when the natural signals are weak.

Additional information associated with this data set is available in the following files:

+------------------------------------------------------------------------------------------------------------------------------------------------------------------+
| File                                                                         | Contents                                                     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| http://ppi.pds.nasa.gov/ditdos/download?id=pds://PPI/VGPR 1201/CATALOG/VG1 PRA1 INST.CAT       | VG1 PRA instrument description                               |
| http://ppi.pds.nasa.gov/ditdos/download?id=pds://PPI/VGPR 1201/CATALOG/VG1 PRA JUP HRES DS.CAT | Data set description                                         |
| http://ppi.pds.nasa.gov/ditdos/download?id=pds://PPI/VGPR 1201/CATALOG/PERSON.CAT              | Personnel information                                        |
| http://ppi.pds.nasa.gov/ditdos/download?id=pds://PPI/VGPR 1201/CATALOG/REF.CAT                 | Key reference description                                    |
| http://ppi.pds.nasa.gov/ditdos/download?id=pds://PPI/VGPR 1201/DOCUMENT/INSTRUMENT             | ASCII and HTML versions of the PRA investigation description paper |
+------------------------------------------------------------------------------------------------------------------------------------------------------------------+</Description>
```

Figure 2. The SPASE Description text element located within the SPASE Resource Header of the Numerical Data description for Voyager 1 Planetary Radio Astronomy, PRA, low band receiver 6 s cadence data from the 1979 Jupiter encounter. The description is shown after formatting the text to comply with text normalization and text mark-up interpretation rules. Post-migration SPASE Resource ID: NASA/NumericalData/Voyager1/PRA/Jupiter/Low/PT6S.

```
      <Description>(Description based on material from VG1_PRA_JUP_HRES_DS.CAT) Voyager 1 Radio Astronomy (PRA) data from
the Jupiter encounter (1979-01-06 to 1979-04-13). The data set provides 6 second high resolution lowband radio mean power
data. The data are provided for 70 instrument channels, covering 1.2 to 1326.0 kHz. This data set (VG1-J-PRA-3-RDR-LOWBAND-
6SEC-V1.0) contains data acquired by the Voyager-1 Planetary Radio Astronomy (PRA) instrument during the Jupiter encounter.
The bounding time interval set for most Voyager 1 Jupiter PDS data sets is the Voyager project defined 'far encounter'
mission phase boundary (1979-02-28 to 1979-03-22). Since, however, the PRA instrument is able to observe planetary
phenomenon at much larger ranges than other fields and particles experiments, this boundary is artificial with respect to
PRA. Hence, PRA lowband data provided here cover the entire Jupiter Encounter Phase (1979-01-06 to 1979-04-13). Data from
beyond the far encounter interval is contained in the cruise data archive which is available from the NSSDC. VG1-J-PRA-3-RDR-
LOWBAND-6SEC-V1.0 contains data at the highest time resolution possible during normal operations. The normal mode of PRA
operations during the planetary encounters was to sweep through the two radio receiver bands, high band (40.5 to 1.5 MHz in
128 channels spaced 0.3072 MHz apart) and low band (1326.0 to 1.2 kHz in 70 channels spaced 19.2 kHz apart) in a period of 6
seconds. The receivers measured, on alternate samples, the left hand circular and right hand circular (radio definition)
power. Measured Parameters ================== The data here are from the low frequency receiver band and are 'packaged'
into spacecraft major frame records. Each major frame is 48 seconds long or eight sweeps through the PRA receiver. The data
are calibrated and are given in units of 'millibels' which is 1000 times the log of the received power. Zero millbels
corresponds to approximately 1.4 x 10^-21 W m^-2 Hz^-1, however, this value is never seen in practice. The minimum values
detected, which includes receiver internal and spacecraft generated noise, are about 2300 to 2400 millibels, or about 3.5 x
10^-19 W m^-2 Hz^-1; even higher values are seen at the very lowest frequencies. The data format is ASCII and consists of a
time indicator followed by an array containing the eight low band sweeps. Time is spacecraft event time (SCET) which is
basically universal time at the spacecraft. Specifically, time is in the form of YYMMDD and seconds into YYMMDD. Both are
written as I6. Example: July 1, 1979 at 12 hours SCET would be 790701, 43200. The seconds correspond, to the nearest second,
to the start of the sweep (which occurs in PRA high band). The first value in low band (1326.0 kHz) occurs some 3.9 seconds
after this time and samples at successively lower frequencies are spaced 0.03 seconds apart. Only one time is given for the
entire major frame, thus the start of each sweep is the time given plus 6 times the sweep number minus 1 (i.e., 0 through
7). The data array is dimensioned as 71 X 8 and written as I4 format (i.e. 568I4). The '8' corresponds to the eight PRA
sweeps. The lowest 68 of the 70 low band channels (1287.6 to 1.2 kHz) are in positions 2-69. Positions 70-71 should be
ignored. Missing or bad data values are set to zero. In position 1 of each sweep is a status word where the 12 least
significant bits have used, although not all 12 have meaning for PRA low band. Numbering those bits 0 for least significant
to 11 for most significant, the bits that have meaning are as follows: bit 0: 15 dB attenuator in use when equal to 1 1: 30
dB attenuator in use when equal to 1 2: 45 dB attenuator in use when equal to 1 9,10 (together): polarization of first
channel sampled (1326.0 kHz) according to the scheme: +--------------------------+ value bit 10= 0 1 value bit 9= 0 R L 1 L
R +--------------------------+ Polarization at successively lower frequencies is opposite to the frequency above it, i.e.
either a LRLR or an RLRL pattern. Successive 6-second sweeps start on the opposite polarization as the previous sweep as
indicated in the status bits. Note that this polarization is the received polarization, not necessarily the emitted
polarization. Correct interpretation of the received polarization depends on the antenna plane orientation relative to the
radio source. A good description of this concept can be found in Leblanc Y., Aubier M. G., Ortega-Molina A., Lecacheux A.,
1987, J.Geophys. Res. 92, 15125 and in Wang, L. and Carr, T.D., Recalibration of the Voyager PRA antenna for polarization
sense measurement, Astron. Astrophys., 281, 945-954, 1994. and references therein. Missing or bad data values are set to
zero. If the status word is zero, any data in that receiver sweep should be discarded. Data Coverage ============= The data
are stored as 4 ASCII tables (.TAB), each accompanied with a PDS label file (.LBL) which describes properties of the data
file. Data cover the following time intervals: Volume ID: VGPR_1201 +---------------------------------------------------------
----------------+ Filename Records Start Stop ------------------------------------------------------------------------
PRA_I.TAB 35569 1979-01-06T00:00:34.000Z 1979-01-30T23:59:47.000Z PRA_II.TAB 39493 1979-01-31T00:00:35.000Z 1979-02-
25T23:59:47.000Z PRA_III.TAB 41371 1979-02-26T00:00:35.000Z 1979-03-22T23:59:56.000Z PRA_IV.TAB 24587 1979-03-
23T00:00:44.000Z 1979-04-13T23:59:08.000Z +------------------------------------------------------------------------+
Confidence Level Overview ======================== The accuracy of calibration in the PRA low band is approximately 2 dB,
except at frequencies below 100 kHz where it is somewhat worse. Interference from the Voyager power subsystem is a major
problem to the PRA instrument, affecting many of the 70 low band channels. This interference manifests itself by abrupt
changes in background levels. Some channels, notably 136 and 193 kHz, are almost always affected, whereas, others are only
affected for short intervals. Usually, thisinterference is only a problem when the natural signals are weak. Additional
information associated with this data set is available in the following files: +-----------------------------------------
-------------------------------------------------------------------------------------+ file contents
http://ppi.pds.nasa.gov/ditdos/download?id=pds://PPI/VGPR_1201/CATALOG/VG1_PRA1_INST.CAT VG1 PRA instrument description
http://ppi.pds.nasa.gov/ditdos/download?id=pds://PPI/VGPR_1201/CATALOG/VG1_PRA_JUP_HRES_DS.CAT data set description
http://ppi.pds.nasa.gov/ditdos/download?id=pds://PPI/VGPR_1201/CATALOG/PERSON.CAT personnel information
http://ppi.pds.nasa.gov/ditdos/download?id=pds://PPI/VGPR_1201/CATALOG/REF.CAT key reference description
http://ppi.pds.nasa.gov/ditdos/download?id=pds://PPI/VGPR_1201/DOCUMENT/INSTRUMENT ASCII and HTML versions of the PRA
investigation description paper +---------------------------------------------------------------------------------------
---------------------------------------+</Description>
```

Figure 3. The SPASE Description text element located within the SPASE Resource Header of the Numerical Data description for Voyager 1 Planetary Radio Astronomy, PRA, low band receiver 6 s cadence data from the 1979 Jupiter encounter. The description is shown prior to formatting the text to comply with text normalization and text mark-up interpretation rules. Pre-migration SPASE Resource ID: VSPO/NumericalData/Voyager1/PRA/Jupiter/Low/PT6S.

```
      <Description>* Data Set Overview
* ================

* Version 1.1
* ===========

This Version 1.1 Data Set replaces the Version 1.0 Data Set (DATA_SET_ID=VG1-J-MAG-4-1.92SEC) previously archived with PDS.
Changes to this Version include the Addition of Data Columns not included in Version 1.0, the Modification of Time Format
and Flag Values, and Upgrade of associated Labels and Catalog Templates to PDS Version 3.2.

* Data Set Overview
* ================

This Data Set includes calibrated Magnetic Field Data acquired by the Voyager 1 Low Field Magnetometer (LFM) during the
Jupiter Encounter. Coverage begins in the Solar Wind inbound to Jupiter and continues past the last outbound Bowshock
Crossing. The Data are in System III (1965) (SYS3) Coordinates and have been averaged from the 60 ms Instrument Sampling
Rate to a 1.92 s Sample Rate. All Magnetic Field Measurements are given in nT. The Magnetic Field Data are calibrated (see
the Calibration Description included in the Voyager 1 Magnetometer Instrument Catalog File for Details).

* Parameters
* ==========


The full LFM Instrument Sampling Rate is one Sample per 0.06 s. Full Telemetry Resolution "Detail" Data must be obtained
from the Instrument Team. These Data have been resampled at 1.92 s from the Detail Data.

The LFM has eight Dynamic Ranges. The Instrument is designed to switch between Dynamic Ranges automatically depending on the
observed Magnetic Field Magnitude and Fluctuations. Instrument Digitization Uncertainty depends on the Dynamic Range as
indicated in the following Table from (Behannon et al., 1977).

* LFM Dynamic Ranges and Quantization Uncertainty
* ==============================================

+-------------------------------------------------+
| Range Num | Range (nT)        | Quantization (nT) |
|-------------------------------------------------|
| 1         | &#xB1;8.8         | &#xB1;0.0022      |
| 2         | &#xB1;26          | &#xB1;0.0063      |
| 3         | &#xB1;79          | &#xB1;0.019       |
| 4         | &#xB1;240         | &#xB1;0.059       |
| 5         | &#xB1;710         | &#xB1;0.173       |
| 6         | &#xB1;2,100       | &#xB1;0.513       |
| 7         | &#xB1;6,400       | &#xB1;1.56        |
| 8         | &#xB1;50,000      | &#xB1;12.2        |
+-------------------------------------------------+

* Processing
* ==========

Voyager EDRs undergo the following Processing in order to produce these 1.92 s averaged Summary Data:

* Read EDR
* Unpack Header Block (Receiver ID, Spacecraft ID, Telemetry Mode, FDS Counts, Data Flags)
* Convert selected Time Tags to Integer Time (YY/DDD/HH:MM:SS.FFF)
* Unpack Sub-header Block (MAG Status Words, Plasma Data)
* Unpack Science Block (MAG Counts)
* Convert Counts to nT
* Apply Sensor and Boom Alignment Matrices
* Rotate (optional) 1.92 s Averages while averaging Detail Data in nT to create 1.92 s Averages
* Write Summary Record

Counts are measured onboard using 12-bit Words that may represent Values ranging from 0 to 4096. Integer Counts are
converted to Magnetic Field Units (nT) by subtracting a Zero Offset from the measured MAG Value and multiplying this
Difference by the Sensitivity of the Instrument.

* Data
* ====

The Data Files are given in ASCII, Fixed Field Width, Comma-delimited Tables. The Record Structure is described in the
following Table:

* Resampled Magnetic Field Data in System III (1965) Coordinates
* ==============================================================

+----------------------------------------------------------------------------------------------------------+
| Column   | Type | Description                                                                            |
|----------------------------------------------------------------------------------------------------------|
| Time     | a23  | Spacecraft Event Time (UT) of the Sample in the Format: YYYY-MM-DDTHH:MM:SS.SSS          |
| sclk     | a12  | Spacecraft Clock in the Format: MOD65536:MOD60:FDS-LINE                                  |
| mag_id   | i1   | Magnetometer ID (1=LFM, 2=HFM)                                                           |
| Br       | f9.3 | Average of Detail Magnetic Field R Component in nT                                       |
| Btheta   | f9.3 | Average of Detail Magnetic Field Theta Component in nT                                   |
| Bphi     | f9.3 | Average of Detail Magnetic Field Phi Component in nT                                     |
| Bmag     | f    |                                                                                          |
+----------------------------------------------------------------------------------------------------------+</Description>
```

Figure 4. The SPASE Description text element from the SPASE Resource Header from the SPASE Numerical Data description of Voyager 1 magnetometer 1.92 s cadence data from the 1979 Jupiter encounter. SPASE Resource ID: NASA/NumericalData/Voyager1/MAG/Jupiter/PT1.92S. The description contains examples for SPASE list and table mark-up rules.

SPASE Convention Documentation References and URLs

A Space and Solar Physics Data Model from the SPASE Consortium, Version: 2.3.1, Release Date: November 14, 2019, Document Generated: November 21, 2019, http://www.spase-group.org/docs/dictionary/spase-2_3_1.pdf.

Guidelines for Resource ID Formation, Last Update: September 24, 2009, http://www.spase-group.org/docs/conventions/Resource-ID-Formation-Rule-v4.pdf.

Report on SPASE Text Normalization and Mark-up, Document Generated: March 11, 2009, http://www.spase-group.org/docs/conventions/Text-Normalization-and-Mark-up-v1.pdf.

Specification for Heliophysics Event List (Catalogue) Format ("HPEvent" format), Version 1.0, Heliophysics Data and Model Consortium (HDMC), Document Generated: June 12, 2012, http://www.spase-group.org/docs/conventions/HDMC-Event-List-Specification-v1.0.3.pdf.

SPASE Metadata Working Group, SMWG, Registry Explorer URLs

Final VxO Metadata Registry Explorer: http://www.spase-group.org/registry/explorer/

Draft VxO Metadata Registry Explorer: http://www.spase-group.org/draft/explorer/

Promotion of a uniform standard for XML white space usage.

Another good unicode look up web site: http://unicode.scarfboy.com

https://unicodelookup.com

Other Relevant URLs

ISO 8601 Wikipedia Web Site: https://en.wikipedia.org/wiki/ISO_8601