Filling Out the Array Data Structure

The **<Array>** class is the generic base on which all the specific *Array_** classes are built. Use this class only when one of the more specific options cannot be reasonably applied.

In most cases, even a generic *Array* data structure should be accompanied by a *discipline dictionary* class in the *Discipline_Area* of the label, to define the correct format to read in the data. Talk to your node representative for more information on which discipline dictionary would work best with your data.

Following are the attributes and subclasses you will find in *<Array>*, in label order.

Note that in the PDS4 master schema, all classes have capitalized names; attributes never do.

<name>

OPTIONAL

This attribute can be used to give a descriptive name to the array.

<local_identifier>

OPTIONAL

If you need to reference this array class from somewhere else in the label (like in the *Discipline_Area* referencing the Display or Particle dictionaries), use this attribute to define a local identifier to use as a hook. Follow the rules for naming a variable in a typical programming language and you should be OK.

<md5_checksum>

OPTIONAL

Use this attribute to supply an MD5 checksum for the data object *only*. In general, if the data object occupies the entire file, then the checksum should be given as an attribute of the *<File>* class. This checksum is calculated using only the bytes defined by this *Array* data structure.

<offset>

REQUIRED

This is the offset in bytes from the beginning of the file containing the array data to the beginning of the array. You must specify a unit of "byte" for this attribute, thus:

```
<offset unit="byte">0</offset>
```

<axes>

REQUIRED

This attribute is required to be present and must have a value that represents the dimensions of your data. This is the only Array* class that allows for data not 2D or 3D. If your data is 2D or 3D use the Array_2D* or Array_3D* class that best describes your data.

<axis_index_order>

REQUIRED

This attribute is required to be present and must have a value of **Last Index Fastest**. "Last" is with respect to the *<sequence_number>* values in the *<Axis_Array>* classes.

<description>

OPTIONAL

This is a place where additional description can be included, if desired.

<Element_Array>

REQUIRED

This class defines the attributes of the array element.

<data_type>

REQUIRED

The value here must be one of the binary numeric types from the list in the [Standard Values Quick Reference](#).

<unit>

OPTIONAL

If there is a unit of measure associated with the array element values, use this attribute to specify it.

If the data are unitless, do not include this attribute.

<scaling_factor>

OPTIONAL

If the data have been scaled (divided by a constant), put the scaling factor in this attribute.

When reading the data, the value is multiplied by the *<scaling_factor>* value before adding the *<value_offset>*.

<value_offset>

OPTIONAL

If an offset has been subtracted from the data, put the offset value in this attribute. Offsets may be positive or negative.

When reading the data, the value is first multiplied by *<scaling_factor>*, then the *<value_offset>* is added.

<Axis_Array>

REQUIRED

This class describes one dimension of the array. The number of instances of this class should match the dimensions of your array.

\<axis_name\>

REQUIRED

This is the name of the array axis being described. The *axis_name* is typically something like "Wavelength" or "Distance". The value should be useful for labeling the axis in a display. For some data structures derived from the \<Array\> structure, the names of the axes may be fixed.

Note: *Axis names should be unique (within the array object), but this is not currently enforced. Cut and paste carefully.*

\<local_identifier\>

OPTIONAL

This is a unique (within the label) name for the axis. So if your label contains three different arrays, all three can have an axis with an *axis_name* value of "Line", but they may not have the same values for *local_identifier*. Include this attribute when you will reference specific axes from some other part of the label. Typically, this will happen when you use discipline dictionaries that reference parts of data structures, as the Spectral Dictionary does in defining the spectral characteristics of an array.

\<elements\>

REQUIRED

This attribute must contain the number of elements along this axis of the array. For example, if the *Array* in question has dimensions 112x1, then the *\<elements\>* value in the first *\<Axis_Array\>* would be "112".

\<sequence_number\>

REQUIRED

This number defines an order for the axes so that the *\<axis_index_order\>* value can be interpreted correctly for this *Array*. One of the axes must have a *sequence_number* of "1". It is necessary that the first *Axis_Array* class in the label have *\<sequence_number\>* equal to 1. The remaining axes should be numbered sequentially in separate *Axis_Array* classes.

\<Special_Constants\>

OPTIONAL

Use this class to define any flag values that appear in the data to indicate drop outs, saturation, and other conditions that render a single data point unknown. Every attribute in this class is optional. If you don't need any of the special constants, don't include this class in your *Array*.

\<saturated_constant\>

OPTIONAL

This value indicates the data value was lost because of detector saturation.

\<missing_constant\>

OPTIONAL

This value indicates the data value is known to be missing for some reason not covered by the other constants available in this class.

<error_constant>
OPTIONAL

This value indicates the data value originally reported was known to be in error for some reason, and was replaced by this flag.

<invalid_constant>
OPTIONAL

This value indicates the data value originally recorded or calculated was outside the valid range for array elements.

<unknown_constant>
OPTIONAL

This value indicates the data value in this file is unknown because it was unknown in the source and cannot be recovered.

<not_applicable_constant>
OPTIONAL

This value indicates that the concept underlying the datum is not applicable in a particular context.

<valid_maximum>
OPTIONAL

This value is the maximum possible observational value that *might* be in the data. This is useful if your flag values are greater than this value and you want to simplify the exclusion logic.

<high_instrument_saturation>
OPTIONAL

This value indicates the original datum was in the high-end saturation range of the instrument.

<high_representation_saturation>
OPTIONAL

This value is used to indicate that, while the original observed value was valid, it is out of range of the numeric format chosen for this *Array* in a way that would be considered "too high" - absolute magnitude too great, positive value too large, or positive exponent too large to be represented.

<valid_minimum>
OPTIONAL

This value is the minimum possible observational value that *might* be in the data. This is useful if your flag values are less than this value and you want to simplify the exclusion logic.

<low_instrument_saturation>
OPTIONAL

This value indicates the original datum was in the low-end saturation range of the instrument.

<low_representation_saturation>
OPTIONAL

This value is used to indicate that, while the original observed value was valid, it is out of range of the numeric format chosen for this *Array* in a way that would be considered "too low" - negative value too large or negative exponent too large to be represented.

<Object_Statistics>

OPTIONAL

This class provides a place for statistical values calculated from the real data values of the pixels in the array. Every attribute in this class is optional. If you don't need any of the statistics, don't include this class in your *Array*.

<local_identifier>

OPTIONAL

If you need to refer to this specific set of *Object_Statistics* from elsewhere in this label, this is the place to attach an identifier to it. If your identifier looks like a variable name in a typical programming language, you should be OK.

<maximum>

OPTIONAL

Maximum real data value found in the array as it exists in its file. That is, after any flag values identified in the corresponding *<Special_Constants>* class are ignored and any relevant bit mask is applied, but before *offset* or *scaling_factor* are applied.

<minimum>

OPTIONAL

Minimum real data value found in the array it exists in its file. That is, after any flag values identified in the corresponding *<Special_Constants>* class are ignored and any relevant bit mask is applied, but before *offset* or *scaling_factor* are applied.

<mean>

OPTIONAL

This is the arithmetic mean of the values in the array, excluding those elements containing flag values defined in the associated *<Special_Constants>* class, in the same units as the element. Any bit mask is applied before the calculation, but offset and scaling factor are not.

<standard_deviation>

OPTIONAL

This is the standard deviation of the *<mean>*, excluding those elements containing flag values defined in the associated *<Special_Constants>* class, in the same units as the element. Bit mask is applied; offset and scaling factor are not.

<median>

OPTIONAL

This attribute contains the median value of the real data values (excluding flag values) in the array, in the same units as the element. Any bit mask is applied prior to determining the median, but offset and scaling factor are not.

<maximum_scaled_value>

OPTIONAL

This is the maximum observational value represented in the array. Flag values are excluded; bit mask, scaling factor, and offset are all applied before determining this value.

\<minimum_scaled_value\>

OPTIONAL

This is the minimum observational value represented in the array. Flag values are excluded; bit mask, scaling factor, and offset are all applied before determining this value.

\<description\>

OPTIONAL

If you need to provide any additional information or caveats about the statistics, this is the place to do it.

\<comment\>

OPTIONAL

This attribute holds free-format text which you can use to, for example, explain what it is you are cross-referencing and why.

\<local_identifier_reference\>

REQUIRED

This attribute *must* have a value that corresponds exactly to the value of a *\<local_identifier\>* attribute someplace else in the same label.

\<local_reference_type\>

REQUIRED

This attribute names the relationship between this array and whatever is pointed to by the preceding *local_identifier_reference*. Values must come from a list of permitted values, which does not currently exist.