# Data Representation

The embedded software controlling the ASEC detectors classifies the incident particles by several properties. The number of events in each class is summed over the considered amount of time and transferred to the ADAS server upon the request. Therefore, the ASEC data is represented as a sequence of vectors identified by the readout time and duration of the integration interval (time series). The vector components are representing a number of events registered in each class. The information from the environmental sensors (pressure, temperature, etc…) are obtained at the end of integration time slice and stored in the additional vector components.

The data storage subsystem of the old ASEC data acquisition system has been based on the unstructured ASCII files. The rows were representing vectors and the columns were representing vector components. As it was stated before a lack of structure caused misinterpretations in the data meaning and introduced a lot of obstacles in the data analysis automation. Even more problems of such kind are expected after the international network is set. Therefore, the new XML based data format was developed.

**1. Basic Data Format**

The XML is excellently fit system demands on meta-information describing meaning of the enclosed data. However, the completely new approach will obsolete all existing data analysis software. Further, providing a full range of meta-data with each element of the time series will drastically increase the size of the stored data. This will result in the increased requirements to the network and computational resources and, as a consequence, in a higher installation costs. Finally, the comparison of the XML-based query languages and SQL has shown a tenfold advantage of the SQL approach . Therefore, to compensate the described drawbacks the interim solution is introduced. In the way similar to the ancient data acquisition system the data is represented as vectors within the ASCII strings. These ASCII strings are enclosed in the XML structure providing basic information about the enclosed data and referencing an external document with the thorough detector description.

The data depicted in Fig. 1 illustrates the representation of a single vector by means of the new format. The “installation” attribute references external entity with the detector description. However, the detector structure may be changed several times during detector operation. Therefore, the specific layout used to obtain the vector should be referenced as well. The “layout” attribute is used for that purpose. The “Time” and “Duration” elements indicate the end and duration of the integration time slice. The timestamp and duration are represented following the encoding rules defined by the ISO-8601specification . Special conditions encountered during the data acquisition are described using “Quality” element. Usually, this element indicates hardware failures resulting in partly or completely inaccurate data. The “Value” element holds a data vector in the ASCII representation.

```
<Data installation=”NAMMM” layout=”layoutid”>
   <Time>2006-02-25T16:50:00.0000000+04:00</Time>
   <Duration>P30.0000000</Duration>
   <Quality>Good</Quality>
   <Value>1846 2760 1956 1848 1763 … </Value>
</Data>
```

*Fig. 1. A sample data vector represented using new XML based format is shown on the figure.*

```
<Installation id="NAMMM">
    <Title>Nor-Amberd Multidirectional Muon Monitor</Title>
    <Type>Multidirectional Muon Monitor</Title>
    <Collaboration>
        <Title>Aragats Space Environmental Center</Title>
        <URL>http://crdlx5.yerphi.am/DVIN/</URL>
        <Email>asec@crdlx5.yerphi.am</Email>
    </Collaboration>
    <Maintainer>
        <Title>Cosmic Ray Division of Yerevan Physics Institute</Title>
        <URL>http://crdlx5.yerphi.am/</URL>
        <Email>crd@crdlx5.yerphi.am</Email>
        <PostalAddress>…</PostalAddress>
        <Phone>…</Phone>
        <FAX>…</FAX>
    </Maintainer>
    <Location>
        <Country>Armenia</Country>
        <Region>Aragatsotn</Region>
        <Coordinates>
                <Latitude>40.5</Latitude>
                <Longitude>44.167</Longitude>
                <Altitude>2000</Altitude>
        </Coordinates>
    </Location>
    <CutoffRigidity>7.6</CutoffRigidity>

    <Geometry id="geometryid">
        XML description of the detector components, see below. The multiple geometries may
         be used during the experiment operation and, therefore, described here. The
         "geometryid" is used to reference correspondent geometry from other parts of the
        detector description. The "CurrentGeometry" element  is used to reference currently
        used detector geometry.
    </Geometry>
    <Layout id="layoutid">
        XML description of the logical data layout, see below. The data layout may be altered
        during the experiment operation. The "layoutid" is used to reference correspondent
        data layout from other parts of the detector description. The "CurrentLayout" element
        is used to reference currently used layout.
    </Layout>
    <CurrentGeometry>geometryid</CurrentGeometry>
    <CurrentLayout>layoutid</CurrentLayout>

    <Data installation="NAMMM" layout="layoutid">
        The optional data vectors embedded into the descriptions, for format see Fig. 1
    </Data>
    <Data installation="NAMMM" layout="layoutid">
        The optional data vectors embedded into the descriptions, for format see Fig. 1
     </Data>
</Installation>
```

*Fig. 2. The figure contains an example of the detector description. The information corresponds to Nor-Amberd Multidirectional Muon Monitor.*

In that way, the ASCII strings can be easily extracted from the data and used by the legacy applications. The newly developed applications are considering the XML description in order to extract the appropriate data from the ASCII strings.
So, the detector data consists of the data vectors of the aforesaid type and the detector description. This description is not transported with the data but available upon the request from the URCS server by means of the OPC XML-DA protocol. Therefore, the network utilization of new data acquisition system does not exceed that of the old one. However, the data and description can be reconciled in a single document destined for the data exchange with the collaborating groups of scientists (see example in Fig. 2).

Three main components may be segregated in the detector description: the global detector description, the description of the detector components and description of the logical data layout (scientific meaning of each vector component). The first two components are preliminary filled by the operators and the data layout is automatically generated by the URCS software.

## 2. Detector Description

The information included in general detector description is presented in Table 1. The description of the NAMMM monitor is provided as an example in Fig. 2.

*Table 1. The standard fields of the detector description.*

| Element | Description |
|---|---|
| Title | Detector name |
| Type | Type of the detector |
| Collaboration | Information (Web Server, E-Mail) about network the detector is participates in, if any |
| Maintainer | Information (Web Server, E-Mail, Postal Address, Phone, FAX) about organization maintaining the detector |
| Location | The detector location (country, region and geographical coordinates in WGS84 system) |
| Cutoff Rigidity | The cutoff rigidity at the detector location |
| Geometry<br>Layout<br>CurrentGeometry<br>CurrentLayout | Geometry contains description of the hardware component parts; Layout contains description of the data arrangement within the vector. The configuration may be changed during the detector operation. Therefore, several geometries and layouts may be described. CurrentGeometry and CurrentLayout elements are referencing present-day descriptions. |
| Data | The detector description can be reconciled with one or more data vectors by means of Data elements. |

## 3. Detector Geometry

The detector geometry describes the detector component parts and their mutual disposition. This information is used to help hardware engineers in finding damaged components, to help physicists in finding direction the particles are coming from and to provide the basic detector description for the third parties overlooking the data.
The “Coordinate” branch defines a coordinate system by means of two angles. “Slope” is a slope of the XY-plane and “Angle” is an angle between northern direction and Y-axis. The point specified in the “Location” branch of the general detector description is used as an origin of coordinates. Then, the detector placement is described in this coordinate system.
Most of ASEC detectors have layered architecture. Several layers of the scintillation sensors are segregated by the transition layers filtering part of the particles spectra. Therefore, the sensors are

described layer by layer. For each layer the unique ID number, vertical alignment, short assignment description and a list of sensors are defined. For each sensor the following information is defined: ID number unique within the layer, the coordinates, dimensions, type and short HTML description. contains parts of the Nor-Amberd Multidirectional Muon Monitor description as an example.
Each system component part is uniquely identified by the layer ID and sensor ID. This two ID numbers are used to reference component parts from the logical data layout in order to describe which detector component is used to obtain information contained in the certain data channel.

## 4. Data Layout

The most important part of the detector description is an explanation of the component layout within the data vector. This data layout description indicates the physical meaning and acceptable value range for each vector component. The XML description providing the layout explanation consists of an element list. Each element provides information about a single data component and specifies its location within the data vector (see Fig. 3).

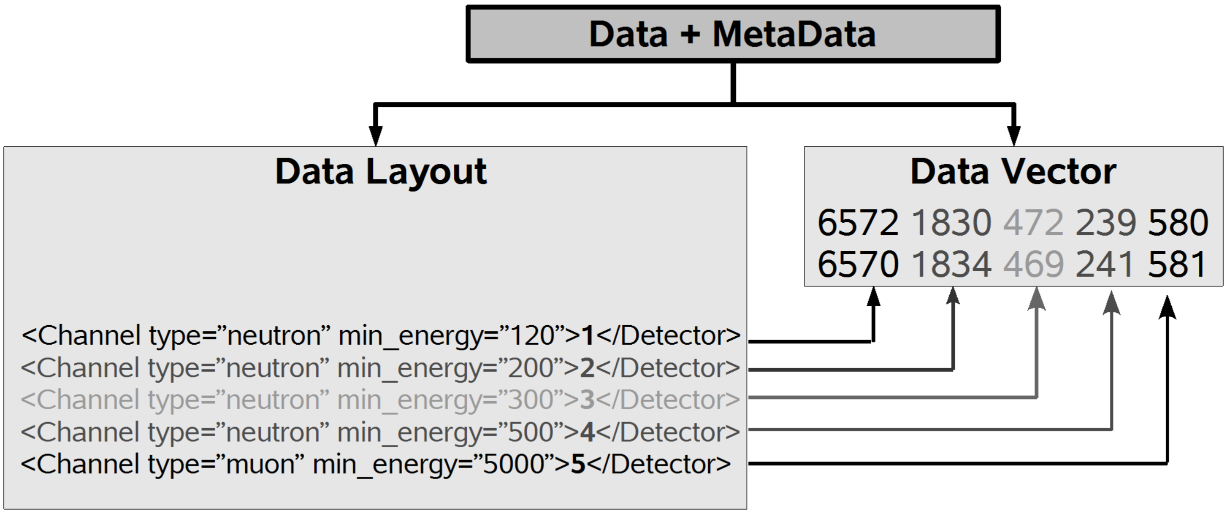


*Fig. 3. The figure illustrates how the data layout description is used to provide information on the physical meaning of the data components.*

Utilizing the proposed approach the application may execute XPath query on the layout description and select the nodes of interest. For, example the query "*//Layout/Channel[@type = 'moun']*" will extract all elements providing the information on the muon flux intensity. Then, the application finds positions in the data vector containing the interesting data and extract it. Thus, the data acquisition system can automatically find the required information and in the case if the some channels are added to the detector, removed from the detector or rearranged the software will automatically handle new data layout without code adjustment.

Currently, there are defined three types of channel description elements: "Channel", "Variance" and "Correlation". The "Channel" describes a value representing measurements of a certain sensor. As it was described in introduction, the ASEC detectors are measuring the number of incident particles registered during the certain amount of time. At the moment the minute intervals are used. However, the embedded software is able to operate with intervals having a precision up to the hundreds of milliseconds. This information is not stored, but used to calculate variances of the stored minute data and correlations between the data channels. The "Variance" and "Correlation" elements are used to describe values containing these variance and correlation information, accordingly.

The following attributes are used to describe the value meaning and acceptable range:
*id*: Unique ID number, used to reference value.
*detector*, *layer*: The attributes contains detector and layer ID numbers indicating the detector part used to obtain value. The attributes are not included for the data obtained by means of several parts, like coincidences (a count rate of a particle flux registered by the sensors in the upper and lower layers simultaneously).
*type*: The "type" is an enumeration describing the physical meaning of the value. Currently defined types are listed in Table 2.
*units*: The attribute specifies engineering units in which the value is represented.
*low*, *high*: These attributes are pointing to the minimal and maximal values likely to be obtained in normal operation. If the values are exceeding these limits the alert is sent to the operator.
*energy_min*, *energy_max*: For the sensors counting a number of incident particles, these attributes are specifying minimal and maximal energy thresholds. Only the particles with energy within the specified thresholds are counted. If only "energy_min" attribute is specified, then the data value represents a number of incident particles with energy above the specified threshold. The energy is specified in megaelectron-volts.
*direction*: In order to estimate the intensity of the particle flux coming from the certain direction, the ASEC electronics separately counting number of particles simultaneously registered by the several layers of the scintillation sensors. The "direction" attribute is used to describe such type of the coincidence data.
To indicate the direction, the ID numbers of the sensors, which are registered the particles, are presented top-down. The numbers are divided by the '-' sign. For example, "*1-5*" means that the counted particles are passed through the first detector in the upper layer and the fifth detector in the lower one.
*description*: This attribute contains a short description of a free form.

For the "Variance" and "Correlation" elements only ID number and information about the primary data channels (for which the variance and correlation are calculated) is included. For the "Variance" element it is the "targetid" attribute and for the "Correlation" the "targetid1" and "targetid2" attributes are used.  illustrates the data layout description.

*Table 2. Basic types of the ASEC sensors*

| Type | Description |
|---|---|
| temperature | Temperature sensor |
| pressure | Pressure sensor |
| humidity | Humidity sensor |
| magnx, magny, magnz | Sensor measuring components of the geomagnetic field vector |
| intensity_charged | Sensor measuring intensity of the incident charged particle flux |
| intensity_muon | Sensor measuring intensity of the incident muon flux |
| intensity_neutron | Sensor measuring intensity of the incident neutron flux |