VAMDC-XSAMS Change Log Documentation

Document Information

Editors:	M. Doronin, M.L. Dubernet
Authors:	C. Hill, M. Doronin, T. Marquart
Contributors:	VAMDC WP6 working group
Type of document:	standards documentation
Status:	draft
Distribution:	public
Work package:	WP6
Version:	0.2
Date:	08/06/2011
Document code:	
Document URL:	http://www.vamdc.org/documents/vamdc-xsams-changelog_v0.2.pdf

Abstract: This document outlines the modifications made to the International Atomic Energy Agency's XML Schema for Atomic, Molecular and Solid Data (XSAMS) version 0.1.1 for use by the VAMDC. The modified schema is the working data model for the VAMDC consortium. The adopted name for the modified schema is VAMDC-XSAMS.

Version History

Version	Date	Modified By	Description of Change
V0.2-r1	22/04/2011	M.Doronin	first draft
V0.2-r2	04/05/2011	M.Doronin	Accuracy description added
V0.2-r3	05/05/2011	M.Doronin	Fixes in Broadenings
V0.2-r4	19/05/2011	M.Doronin	Fixes in CollisionalTransition
V0.2-r5	08/06/2011	M.Doronin	Updated references

Disclaimer

License

All rights reserved

Acknowledgements

VAMDC is funded under the "Combination of Collaborative Projects and Coordination and Support Actions" Funding Scheme of The Seventh Framework Program. Call topic: INFRA-2008-1.2.2 Scientific Data Infrastructure. Grant Agreement number: 239108.

CONTENTS

CHAPTER

ONE

INTRODUCTION

This document outlines the modifications made to the International Atomic Energy Agency's XML Schema for Atomic, Molecular and Solid Data (XSAMS) version 0.1.1 for use by the VAMDC. The modified schema is the working data model for the VAMDC consortium. The adopted name for the modified schema is XSAMS-VAMDC.

XSAMS version 0.1.1: http://www-amdis.iaea.org/xsams/docu/xsams-0.1.1.pdf

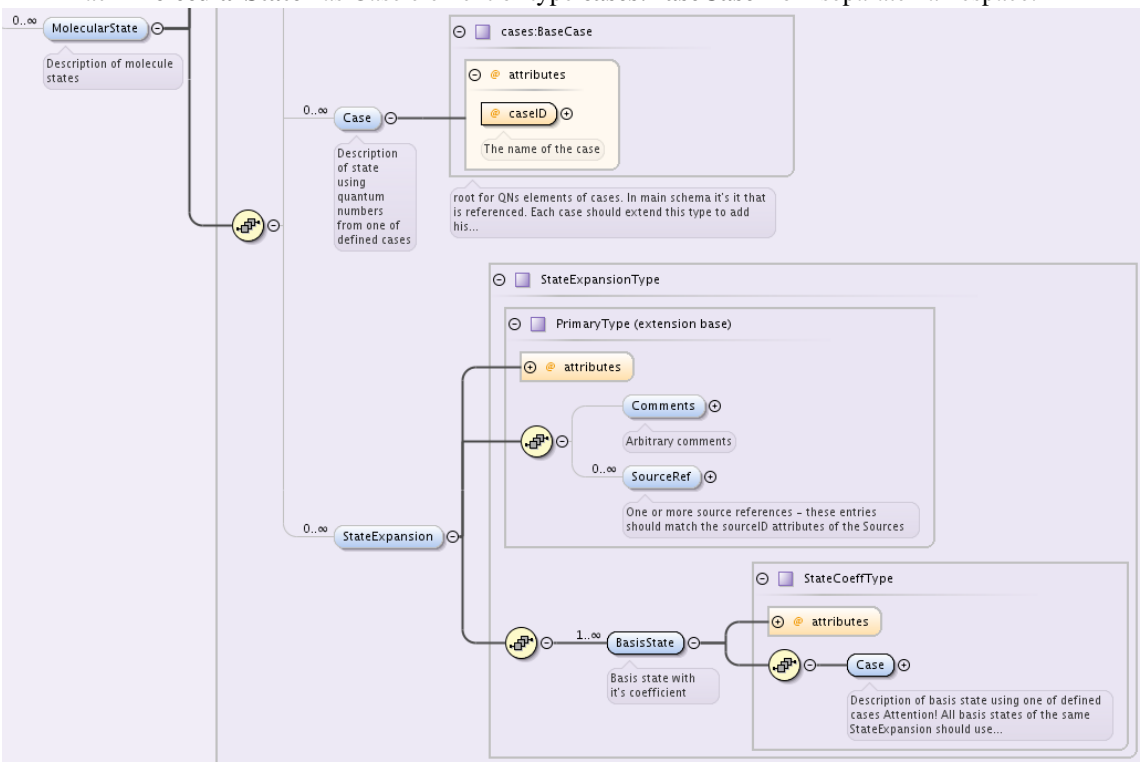
XSAMS-VAMDC version 0.2 xsd: http://www.vamdc.org/downloads/vamdc-xsams-v0.2.tar.bz2

1.1 Changes versus official schema v.0.1.1

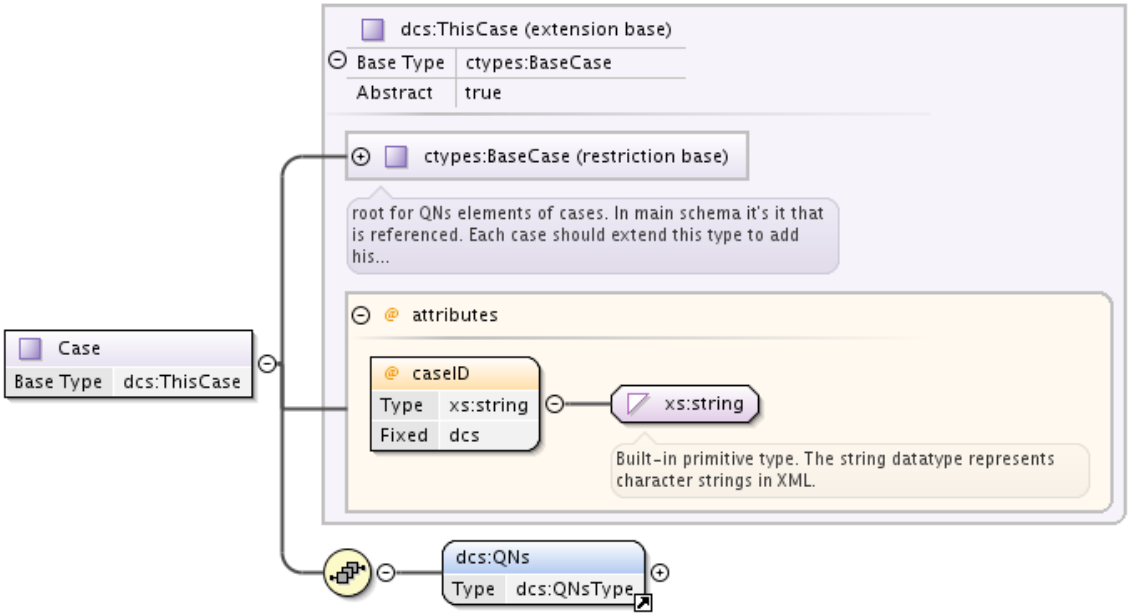
1. **MolecularState** structure has been replaced with case-by-case representation, initially proposed by C.Hill.

   The case-by-case XML description of molecular states within XSAMS is designed to provide a straightforward and flat data structure for representing the quantum numbers and symmetries that denote a molecular state. For full documentation on cases see http://www.vamdc.org/documents/cbc_v0.2.pdf.

   Each **MolecularState** has Case element of type **cases:BaseCase** from separate namespace.



   BaseCase type defines the single attribute, *caseID*, that denotes the case used.



Each specific case, in turn, extends that **BaseCase** type, defining specific *caseID* attribute value and adding **QNs** element that contains a sequence of quantum numbers and symmetries.

1.2 Changes introduced before Cambridge meeting

As things are defined in the current version of schema.

For real changelog see SVN logs and diffs.

Most of the things contributed by Christian Hill.

1. **States** element renamed into **Species**
2. **SourceRef**'s in PrimaryType are now elements of type IDREF, not attribute of type IDREFS
3. **AtomicNumericalData**: new elements: **HyperfineConstantA** and **HyperfineConstantB**, describing hyperfine structure caused by interaction with magnetic dipole and electric quadrupole nuclear moments.
4. Environments: added schema part for line shifting/broadening environment. For more details see *Environments*.
5. Broadening: added schema part for *Broadening*.
6. Shiftings: added schema part for *Shifting*.
7. Cross-sections added schema part for *Absorption cross-sections*.
8. NuclearSpinSymmetry in MolecularStateCharacterisation is now NuclearSpinIsomer. Expected values are "ortho", "para", "meta","A","E"
9. Added more units:

   ```
   atm
   1/m3
   1/cm3
   ```

10. Renamed **IonStateType** into **AtomicIonType** Renamed IonState element into Ion in Atoms.
11. Added **speciesID** attribute to MoleculeType (Molecule element).

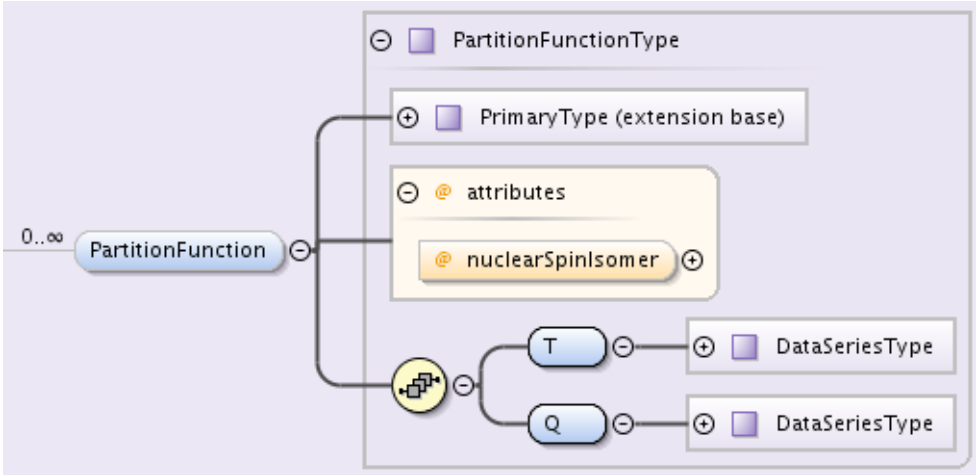
    Added **speciesID** attribute to AtomicIonType in atoms (Atom/Isotope/Ion element).

    In collisions, added optional element **speciesRef** to **Reactants** and **Products**.

    In radiative and nonradiative transitions added optional **speciesRef** element.

12. Added **InChI** and **InChIKey** elements to atomic ions.
13. **InChIKey** is mandatory both for molecules and atomic ions, since it is used for species identification within VAMDC.
14. Removed **Comments** element from **Methods**, since **Comment** element is inherited from **PrimaryType**.
15. **ParticleType** is now extending **PrimaryType**.
16. **AtomicState and MoleculeState are now optional elements.** Since we do have **speciesID/speciesRef**, it is not always necessary to define exact state of molecule/atom to refer to it.
17. CML **atomArray** and **bondArray** elements within **MoleculeStructure** element added to describe molecule structure. They replace **AtomNType**, **BondsType**, **AtomsType** and **MoleculeNuclearSpins** elements and types.
18. New vector and matrix data types
19. Introduced schema-wide namespace, with URL.

    All attributes are defined by type now, not by reference.
20. Removed **WavelengthWavenumberType** in Processes/Radiative/EnergyWavelengthType. All elements within EnergyWavelengthType are now of type DataType and may occur more than once. Type should be determined by methodref: experiment, theory, ritz
21. Added element EffectiveLandeFactor of type DataType to RadiativeTransitionProbabilityType.
22. Added PartitionFunction to MolecularChemicalSpecies element.



23. Added some test examples (broadening test, azulene information test).

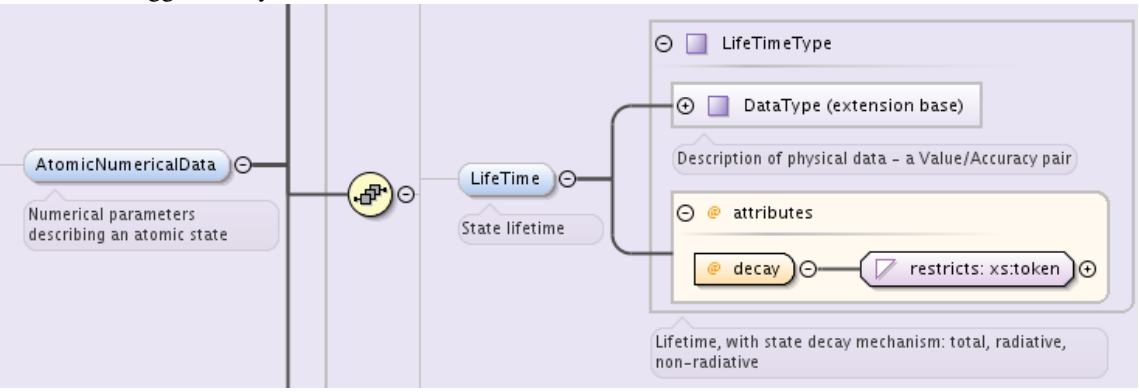
1.3 Changes introduced after Cambridge meeting (after 31.03.2011)

1. Organization of ID/IDREF pairs, all IDREF types are now defined and documented in typesAttributes.xsd
2. Two more units:

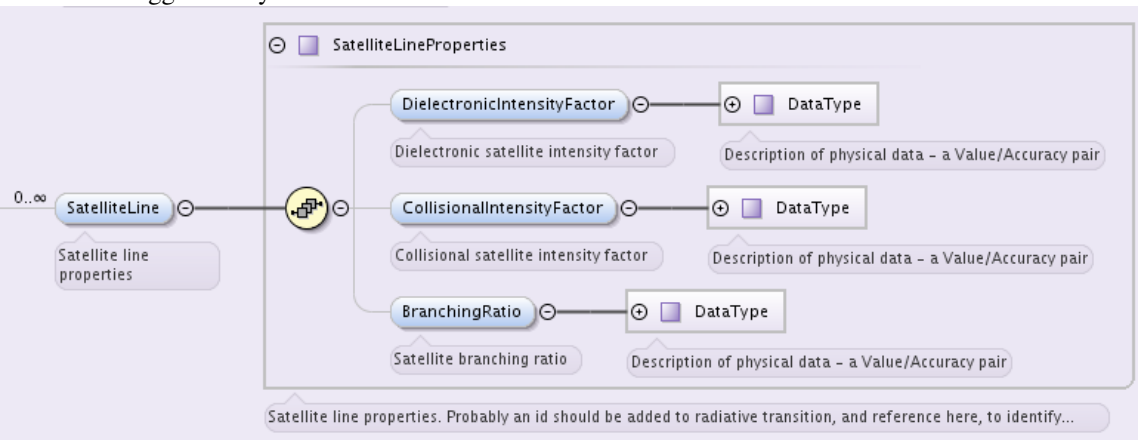
   ```
   1/cm2/atm
   1/cm/atm
   ```

   - Suggestion by Aya Awa Ba
3. Cases are properly linked to the main schema now. Main schema has **Case** element of type commonTypes:**BaseCase**, each case extends it.
4. State expansion is redone as linear combination of Case elements inside BasisState elements

5. In functions, parameter and argument names are attributes now, not elements.
6. **StateRef** type is now called **StateRefType** for consistency
7. Removed *PrimaryType* extension from **MoleculesType**, **AtomsType** and **SolidsType**. Now they are just the containters for respective **Molecule**, **Atom** and **Solid** elements. Removed *PrimaryType* extension from **RadiativeType**, **NonRadiativeType**, **CollisionsType**. The intention is to prevent mixing and confusion **SourceRef**'s in case of merged documents.
8. Removed unused types in statesMolecules that were overridden by CML description of molecular structure.
9. **Probability** in **NonRadiativeTransition** is optional now
    - suggestion by Peter Loboda
10. State lifetime in atoms and molecules now has attribute "decay" that may take values in (total, totalRadiative, totalNonRadiative). It can be specified no more than three times, once for each state decay mechanism.
    - suggestion by Peter Loboda



11. Radiative transition has now satellite line properties element
    - suggestion by Peter Loboda



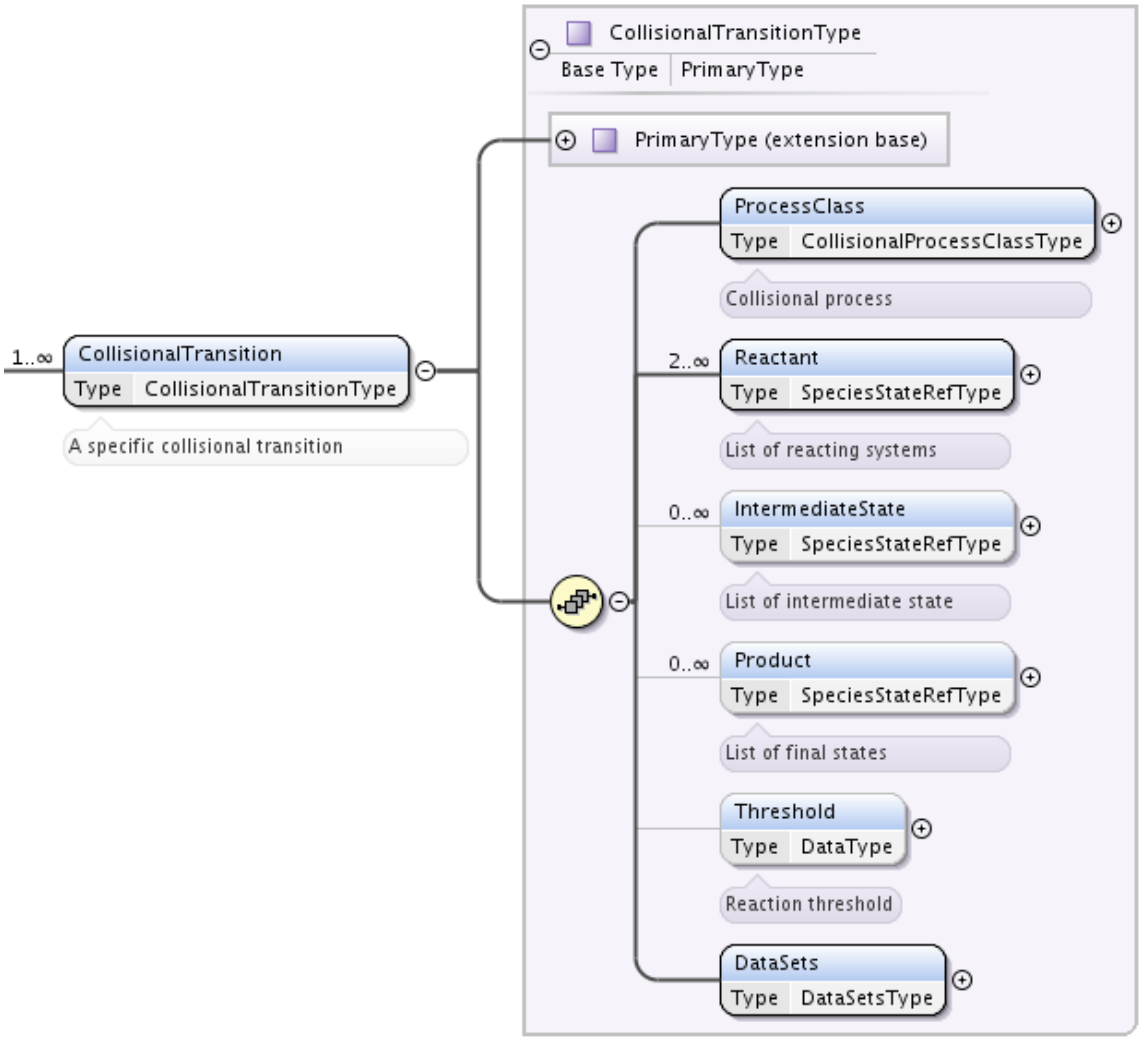
12. Added several test XML files to illustrate changes, moved examples into tests/valid
13. Introduced draft *AccuracyType* description for **DataType**
14. Changes in Collisions:

    In **CollisionalTransition.DataSets.DataSet.FitData** remove redundant **FunctionRef** and **FitValidityLimits** elements, cause they are now kept inside the FitParameters element.

    In **CollisionalTransition** rename

    - **Reactants** into **Reactant**,
    - **Products** into **Product**,
    - **IntermediateStates** into **IntermediateState**,

all of them now may appear more then once and are of type **SpeciesStateRefType**, which has two child elements: **SpeciesRef** and **StateRef**. At least one of them must be specified, both also may be specified. It would be kind to always specify **SpeciesRef**.

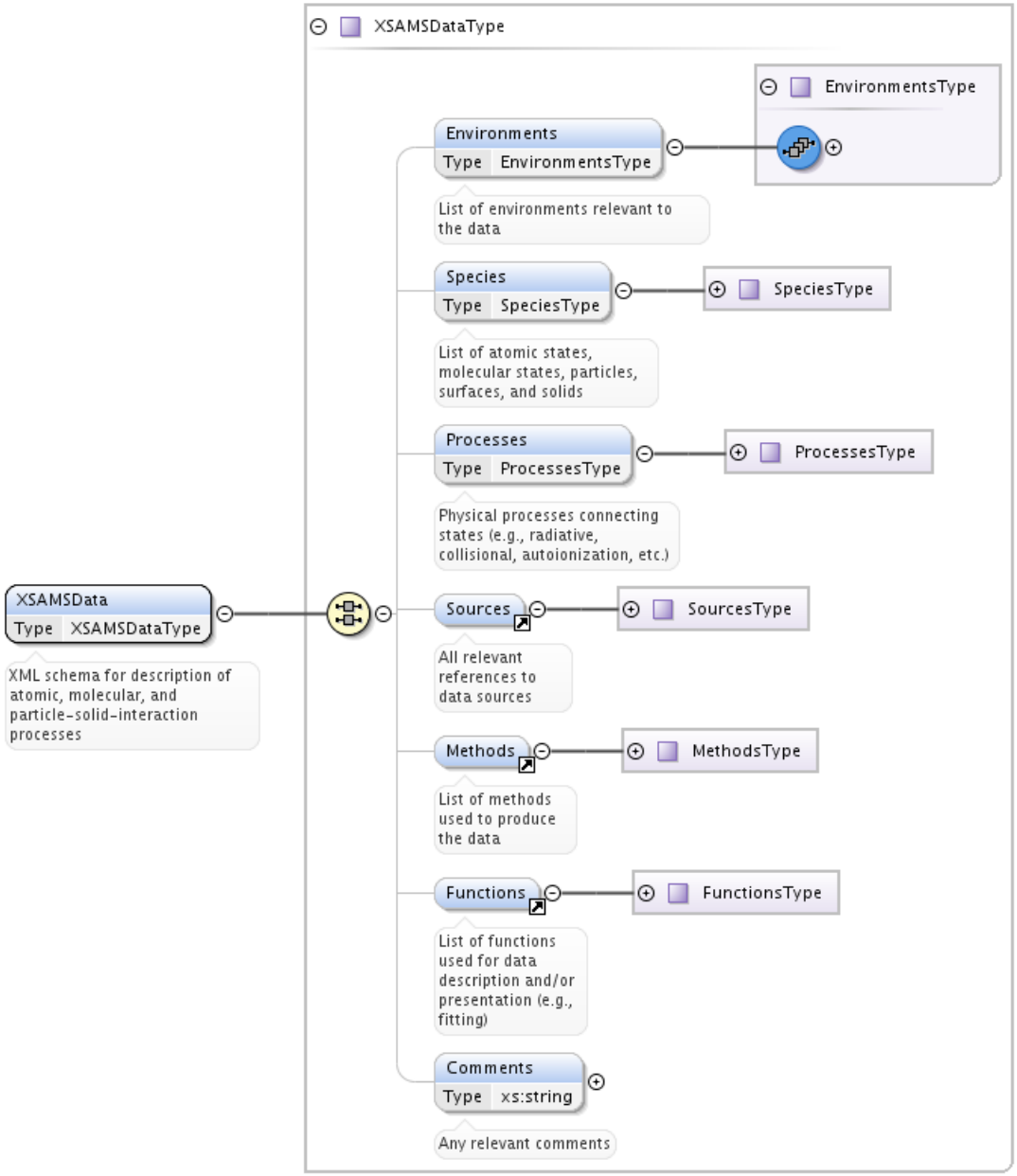


CHAPTER

TWO

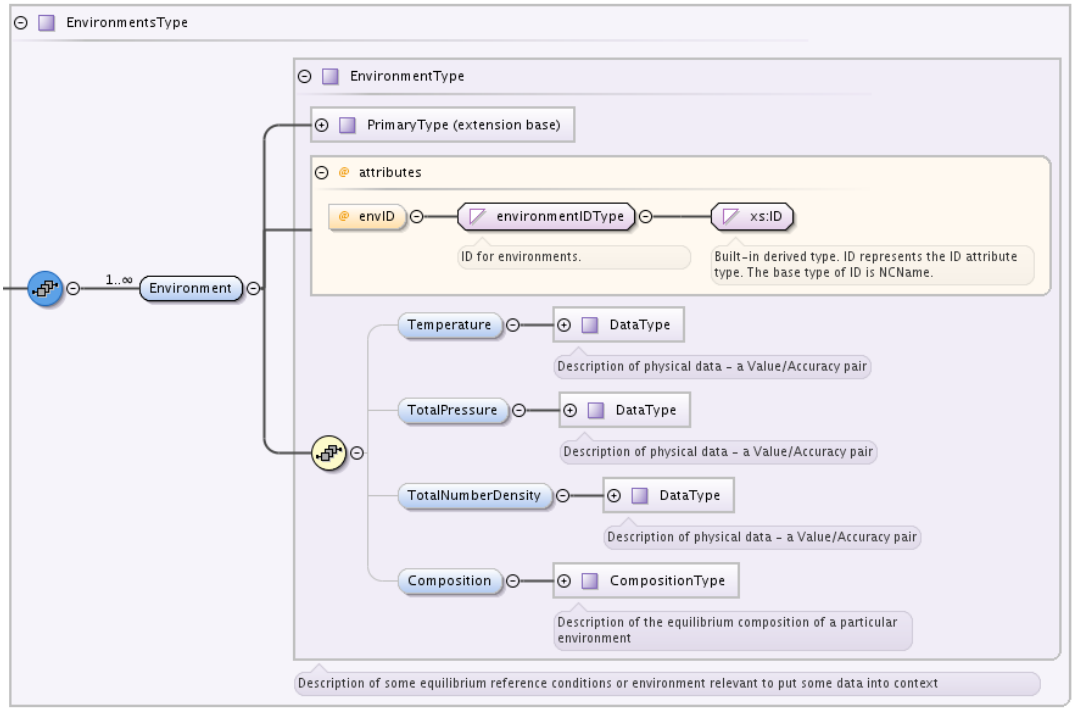
ENVIRONMENTS

XSAMS environments section was first introduced to VAMDC-XSAMS, along with radiative transition line shiftings and broadenings. Currently environments are only used to describe broadening and shifting environment conditions.



2.1 Environment

Each environment is fully described by a single **Environment** element of type **EnvironmentType**



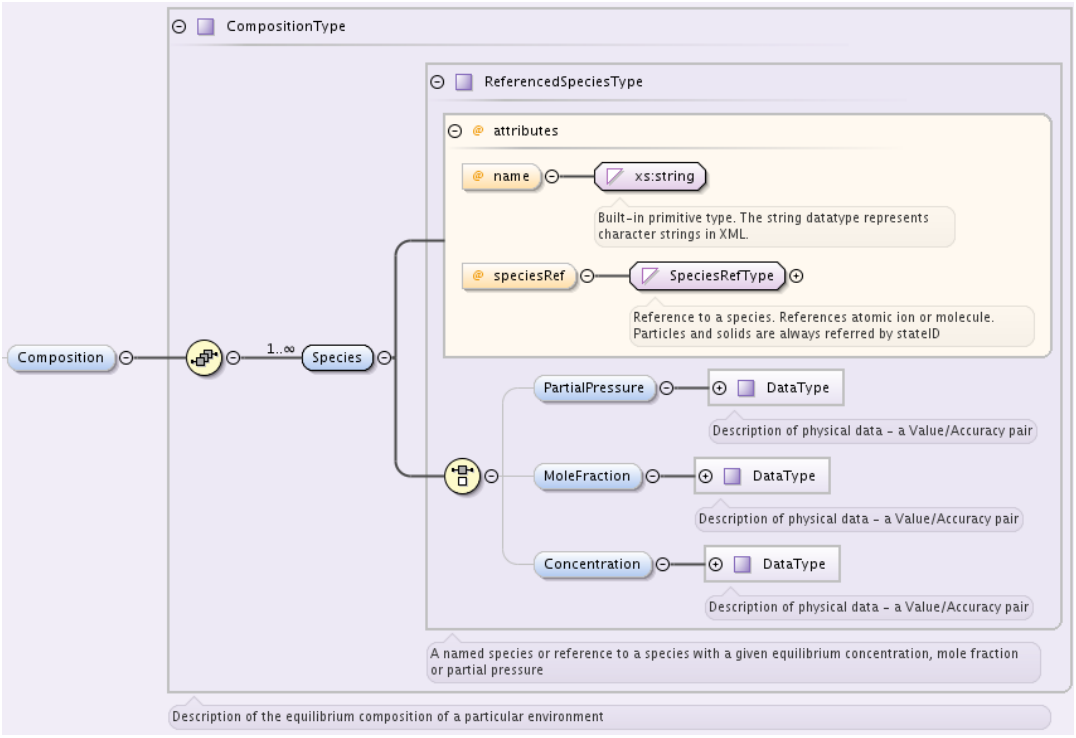
Environment element extends primary type, so reference and method data may be specified.

Environment branch has the following elements and attributes:

- mandatory **envID** attribute. The string identifier starting from letter E, to be referenced from radiative transitions.
- optional **Comments** element, derived from PrimaryType. It is useful to put verbose environment description there.
- optional **Temperature** element of type DataType.
- optional **TotalPressure** element of type DataType.
- optional **TotalNumberDensity** element of type DataType.
- optional **Composition** element. This element allows to description of environment components, e.g. perturber gases.

2.2 Composition

Composition element may describe various kinds of compositions: gas mixture, liquid solution, solid composition, etc.



Composition has one or more **Species** elements, each of them of **ReferencedSpecieType** type.

ReferencedSpecieType allows to define either species **name**, or **speciesRef** reference in corresponding optional attributes. **SpeciesRef** may point through **speciesID** attribute either to atomic ion record or to molecule record in **Species** branch of XSAMS. Both attributes may be defined, in this case name should correspond to atom or molecule pointed by reference.

Component presence measure can be defined by one of the optional elements:

- **PartialPressure**
- **MoleFraction**
- **Concentration**

each of type **DataType**. Only one of them may be present in a component record. To avoid confusion, all components of composition should define the abundance in the same type of elements.

2.3 Examples

Example XML instances of **Environment**:

```
<Environment envID="EHITRAN_refT">
        <Comments>the HITRAN reference temperature, 296 K</Comments>
        <Temperature>
                <Value units="K">296.</Value>
        </Temperature>
</Environment>

<Environment envID="EHITRAN_refpT">
        <Comments>the HITRAN reference pressure and temperature, 1 atm and 296 K</Comments>
        <Temperature>
                <Value units="K">296.</Value>
        </Temperature>
        <TotalPressure>
                <Value units="atm">1.</Value>
```

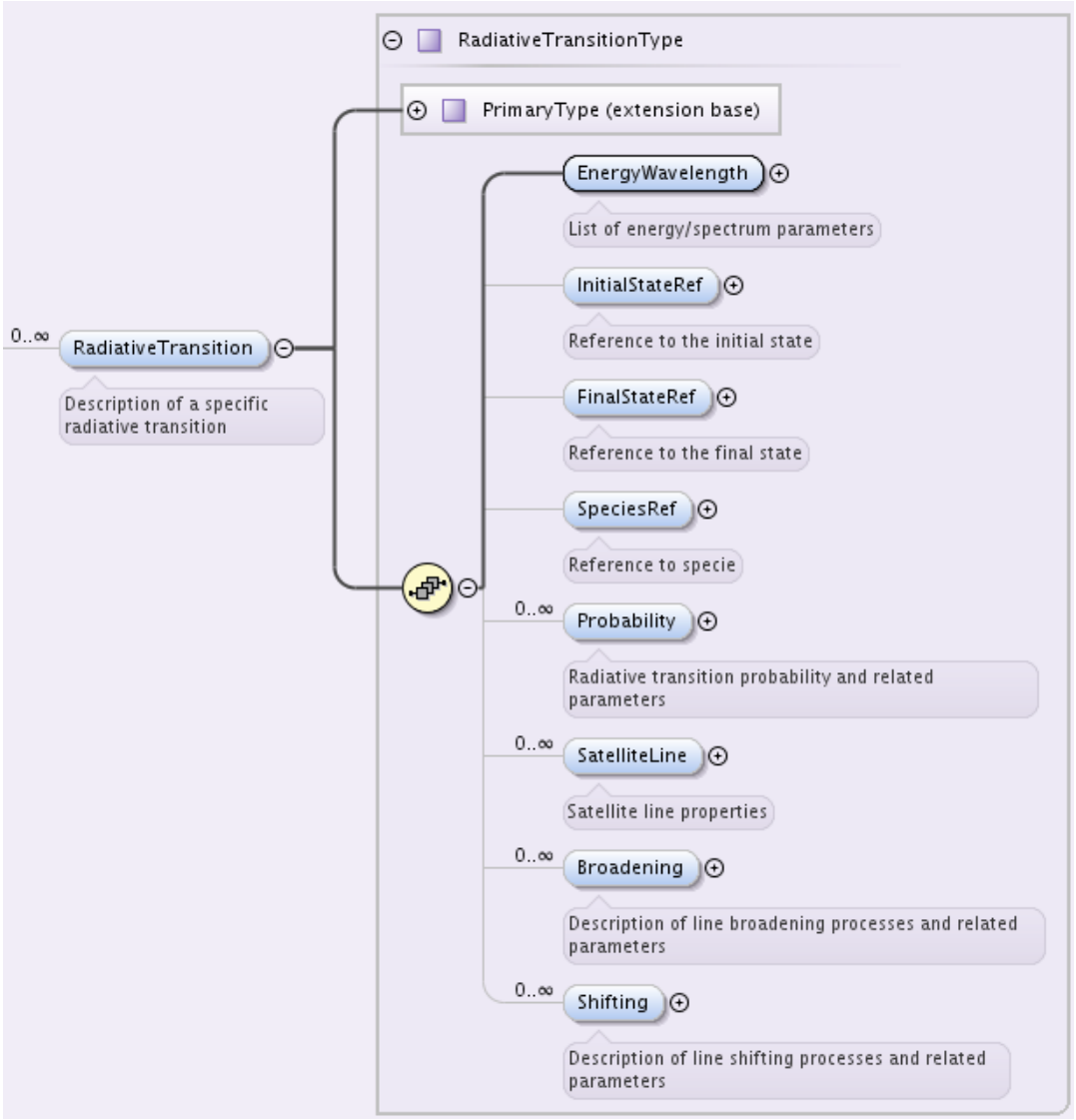
```
        </TotalPressure>
</Environment>

<Environment envID="Eair-broadening-ref-env">
        <Comments>the HITRAN air-broadening reference conditions</Comments>
        <Temperature>
                <Value units="K">296.</Value>
        </Temperature>
        <TotalPressure>
                <Value units="atm">1.</Value>
        </TotalPressure>
        <Composition>
                <Species name="N2">
                        <MoleFraction>
                                <Value units="unitless">0.79</Value>
                        </MoleFraction>
                </Species>
                <Species name="O2">
                        <MoleFraction>
                                <Value units="unitless">0.21</Value>
                        </MoleFraction>
                </Species>
        </Composition>
</Environment>

<Environment envID="Eself-broadening-ref-env">
        <Comments>the HITRAN self-broadening reference conditions</Comments>
        <Temperature>
                <Value units="K">296.</Value>
        </Temperature>
        <TotalPressure>
                <Value units="atm">1.</Value>
        </TotalPressure>
        <Composition>
                <Species name="self">
                        <MoleFraction>
                                <Value units="unitless">1.</Value>
                        </MoleFraction>
                </Species>
        </Composition>
</Environment>
```

CHAPTER

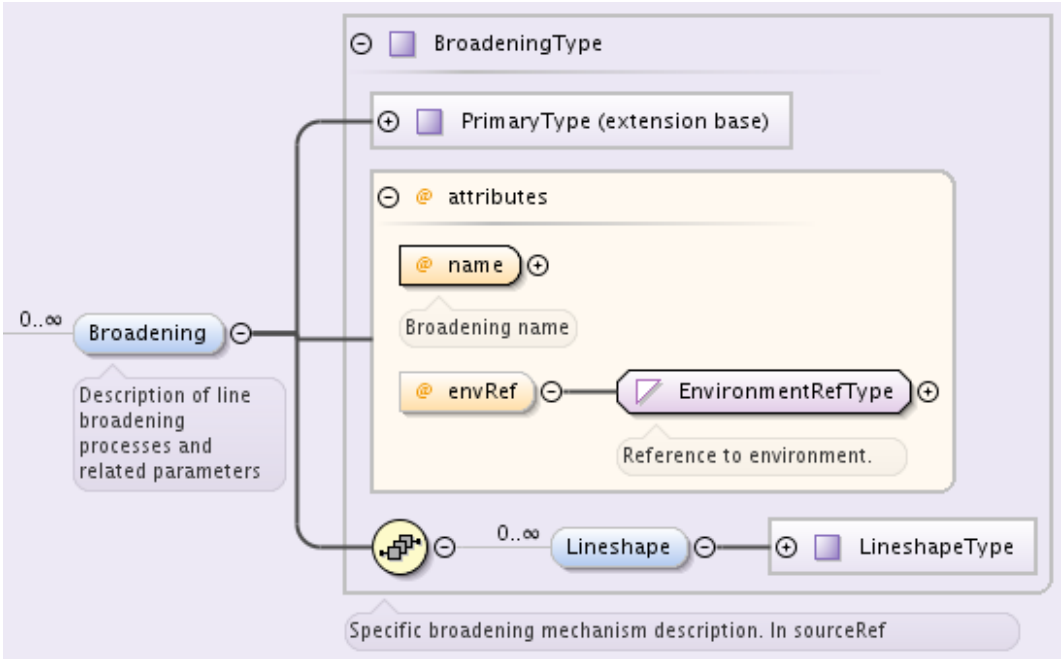
THREE

RADIATIVE LINESHAPES, BROADENING AND SHIFTING

Radiative element now has, among all other, two new elements, called *Broadening* and *Shifting* Those elements may be specified more than once to represent different broadening/shifting processes.



3.1 Broadening

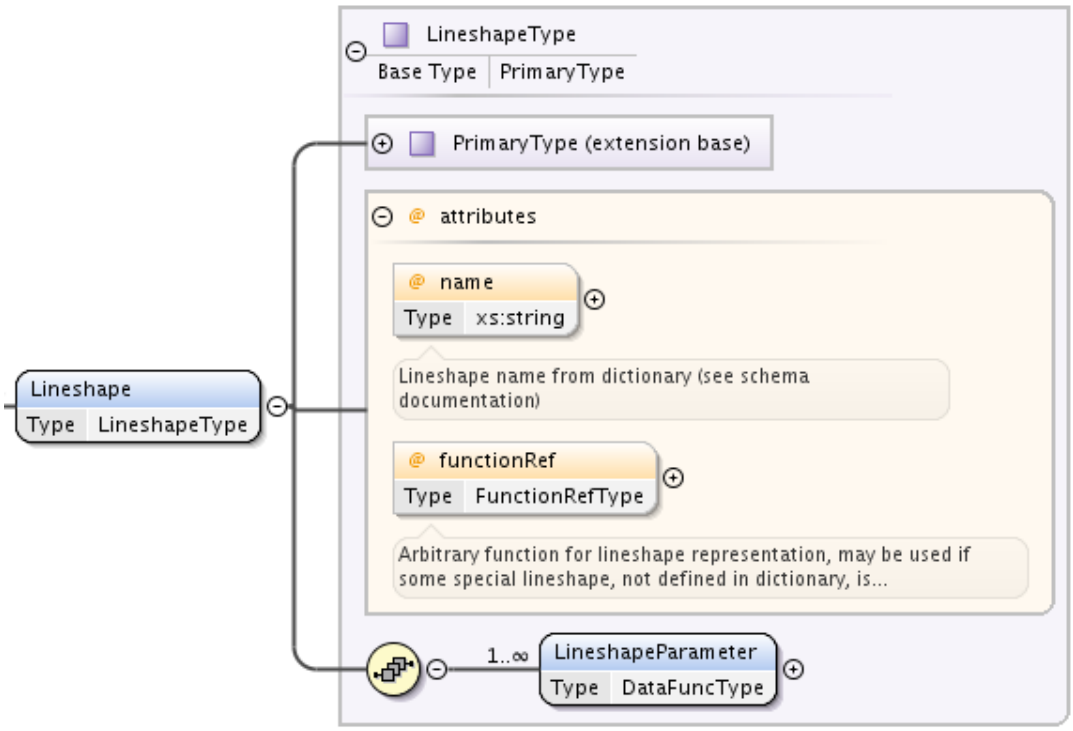


Each broadening element has

- **name** attribute, that must contain one of the names from *Lineshapes dictionary*
- **envRef** attribute, referencing the specific environment conditions, for example, for collisional broadening.
- one or more *Lineshape* elements, with their respective parameters.

Normally, one broadening record should be created for each broadening mechanism and for each source (data origin). Many Lineshape elements allow to represent, for example, processing of the same experimental data with different lineshapes. Usually, there will be only one Lineshape element.

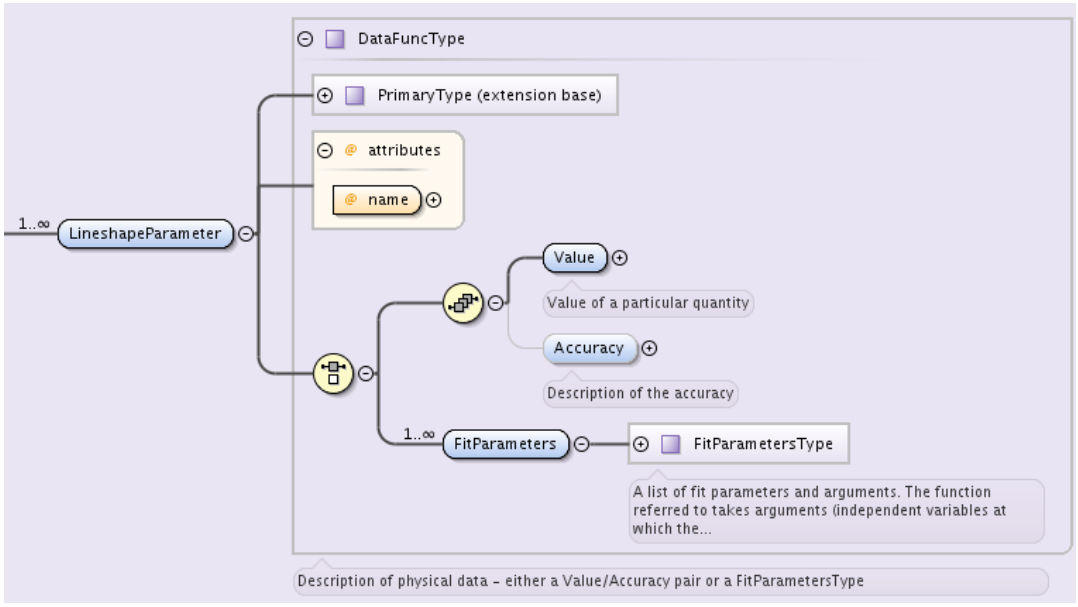
3.1.1 Lineshape



Lineshape in turn also has

- optional **name** attribute from *Lineshapes dictionary*, like Lorentz, Voigt, Doppler, etc.
- optional **functionRef** attribute of type **FunctionRefType**, with a reference to a function describing the lineshape. This attribute should be used only in case when the lineshape used in data fitting is absent in the *Lineshapes dictionary*
- one or more *LineshapeParameter*, each representing specific lineshape parameter as either a constant value or a function of environment parameters

3.1.2 LineshapeParameter



LineshapeParameter either **FitParameters** or **Value/Accuracy** pair must be specified

- **name** attribute corresponds to one defined in *Lineshapes dictionary* for specific lineshape
- **Value**
- **FitParameters** allows representation of this parameter as a function of environment parameters, for example:

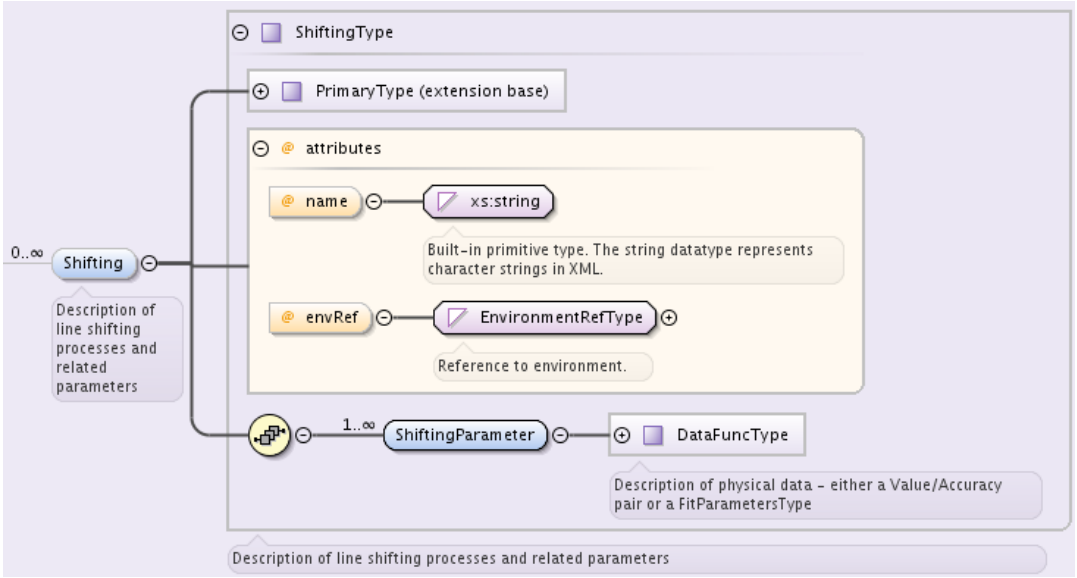
```
<Broadening envRef="Eair-broadening-ref-env" name="pressure">
        <Lineshape name="Lorentzian">
                <Comments>The temperature-dependent pressure broadening
                Lorentzian lineshape</Comments>
                <LineshapeParameter name="gammaL">
                        <FitParameters functionRef="FgammaL">
                                <FitArgument units="K" name="T">
                                        <LowerLimit>240</LowerLimit>
                                        <UpperLimit>350</UpperLimit>
                                </FitArgument>
                                <FitArgument units="atm" name="p">
                                        <LowerLimit>0.</LowerLimit>
                                        <UpperLimit>1.2</UpperLimit>
                                </FitArgument>
                                <FitParameter name="gammaL_ref">
                                        <SourceRef>BHIT-B_HITRAN2008</SourceRef>
                                        <Value units="1/cm">0.0635</Value>
                                        <Accuracy>0.003175</Accuracy>
                                </FitParameter>
                                <FitParameter name="n">
                                        <SourceRef>BHIT-B_HITRAN2008</SourceRef>
                                        <Value units="unitless">0.75</Value>
                                        <Accuracy>0.15</Accuracy>
                                </FitParameter>
                        </FitParameters>
                </LineshapeParameter>
        </Lineshape>
</Broadening>
```

Where function FgammaL is defined as follows:

```
<Function functionID="FgammaL">
        <Comments>This function gives the pressure- and
        temperature-dependence of the Lorentzian component of the
        pressure-broadened line width (HWHM)</Comments>
        <Expression computerLanguage="Fortran">
                gammaL_ref * p * (296./T)**n
        </Expression>
        <Y name="gammaL" units="1/cm"/>
        <Arguments>
                <Argument name="T" units="K">
                        <Description>The absolute temperature, in K</Description>
                </Argument>
                <Argument name="p" units="atm">
                        <Description>The partial pressure of the broadening species,
                                                        in atm</Description>
                </Argument>
        </Arguments>
        <Parameters>
                <Parameter name="gammaL_ref" units="1/cm">
                        <Description>The Lorentzian HWHM of the line, broadened at
                                Tref = 296 K and broadening species partial pressure
                                pref = 1atm</Description>
                </Parameter>
                <Parameter name="n" units="unitless">
                        <Description>
                                The temperature exponent of the gammaL function
```

```
                        </Description>
                </Parameter>
        </Parameters>
</Function>
```

3.2 Shifting



Line shifting is defined by **name** and/or **envRef** attributes. In case of linear collisional shifting, only environment is sufficient.

ShiftingParameter element is defined the same way as broadening *LineshapeParameter*, it is either value or function of environment parameters.

Example of a Shifting definition:

```
<Shifting envRef="Eair-broadening-ref-env">
        <ShiftingParameter name="delta">
                <FitParameters functionRef="Fdelta">
                        <FitArgument name="p" units="atm">
                                <LowerLimit>0.</LowerLimit>
                                <UpperLimit>1.2</UpperLimit>
                        </FitArgument>
                        <FitParameter name="delta_ref">
                                <SourceRef>BHIT-B_HITRAN2008</SourceRef>
                                <Value units="unitless">-0.001</Value>
                                <Accuracy>0.1</Accuracy>
                        </FitParameter>
                </FitParameters>
        </ShiftingParameter>
</Shifting>
```

CHAPTER

FOUR

LINESHAPES DICTIONARY

Possible broadening and shifting names, lineshapes, parameters are listed below. The list is incomplete, it will be expanded before the release and extended later based on user requests and feedback.

4.1 Broadening names

- **pressure** for collisional broadening processes
- **doppler** for Doppler broadening
- **instrument** for instrument-specific broadening
- **natural** for line broadening caused by finite lifetime of initial and final states. Usually, Lorentzian line profile should be used.

4.2 Shifting names

- **pressure** for linear pressure shifting in hi-res spectroscopy

4.3 Lineshape names and parameters

- Lorentzian
  - gammaL
- Voigt
  - gamma
  - sigma

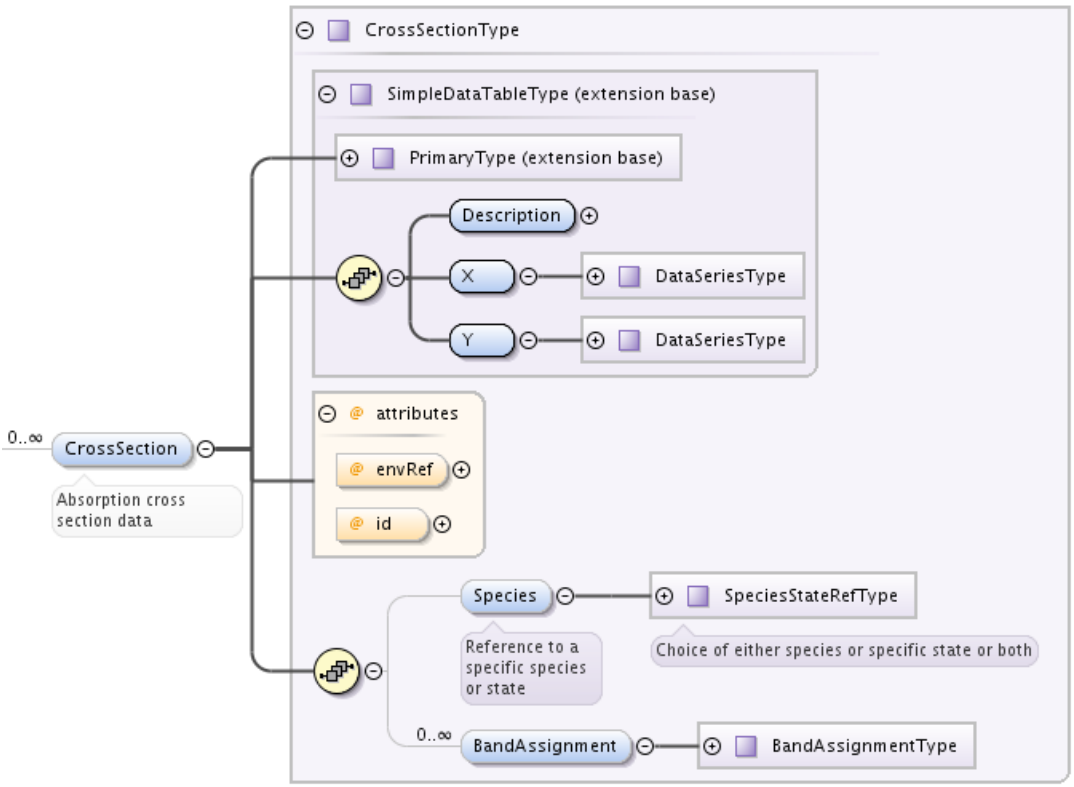
to be continued...

CHAPTER

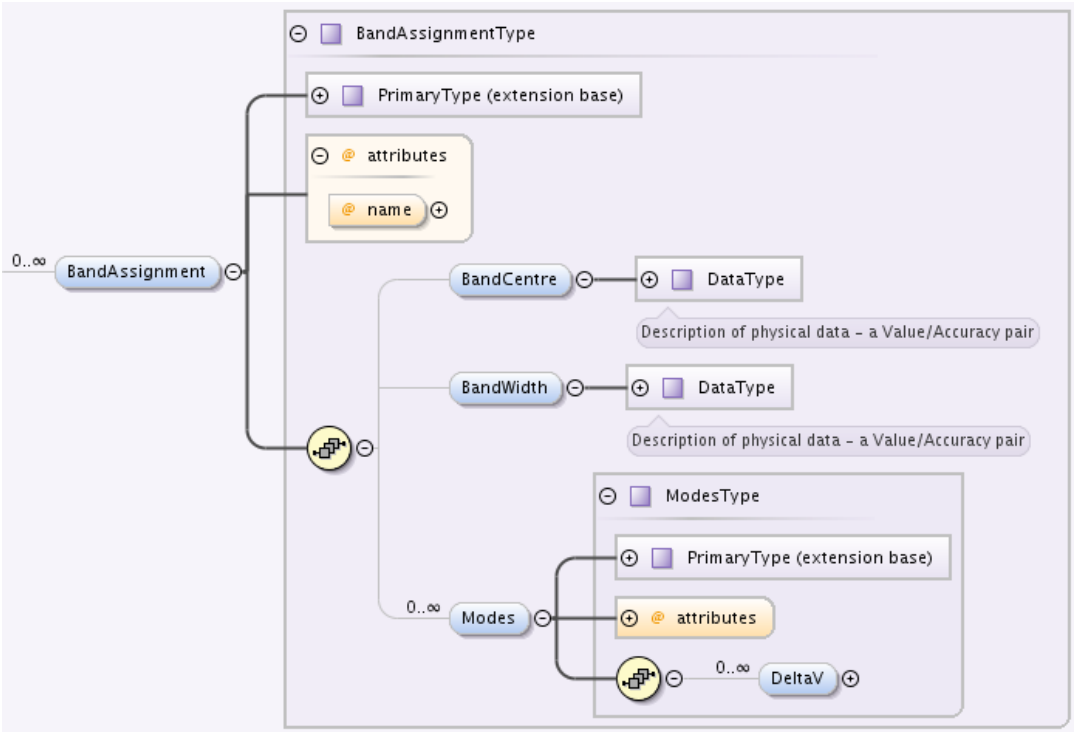
FIVE

ABSORPTION CROSS-SECTIONS

Among with **RadiativeTransition** elements, **Radiative** processes block now has a **CrossSection** element, which allows description of absorption cross-section data and vibrational bands assignment in case of complex molecules.



- **Description**, **X** and **Y** elements describe cross-section data in tabular form, where X can be absorbed radiation frequency, wavelength or wavenumber in a form of a list of values or sequence. Y then represents a sequence of sigma values.
- **Species** element may have **StateRef** and/or **SpeciesRef** child elements, indicating species or specific states, to which crossection data applies.
- **BandAssignment** allows to indicate specific vibrational modes in cross-section data.



5.1 Example cross-sections record

Warning: **DataList** is truncated for clarity, originally it contains 880 space-separated numbers. You may see the original element in schema examples (*tests/valid/azulene-working.xml*).

```
<CrossSection>
        <SourceRef>B_NIST1</SourceRef>

        <Description>The IR transmittance cross section of azulene from the NIST
        Standard Reference Data Program Collection</Description>

        <X parameter="wavenumber" units="1/cm">
                <LinearSequence n="880" units="1/cm" a0="450." a1="4"/>
        </X>
        <Y parameter="sigma" units="arbitrary">
                <DataList n="880">
                        0 85 94 .. 102
                </DataList>
        </Y>

        <Species>
                <SpeciesRef>X-CUFNKYGDVFVPHO-UHFFFAOYAT</SpeciesRef>
                <StateRef>SX_Azulene-1</StateRef>
        </Species>

        <BandAssignment name="2v1+v2">
                <BandCentre>
                        <Value units="1/cm">410</Value>
                        <Accuracy>2</Accuracy>
                </BandCentre>
                <BandWidth>
                        <Value units="1/cm">40</Value>
                        <Accuracy>5</Accuracy>
                </BandWidth>
                <Modes>
```

```
                        <DeltaV modeID="V1">2</DeltaV>
                        <DeltaV modeID="V2">1</DeltaV>
                </Modes>
        </BandAssignment>
        <BandAssignment name="3v4+2v5">
                <BandCentre>
                        <Value units="1/cm">1657</Value>
                        <Accuracy>10</Accuracy>
                </BandCentre>
                <BandWidth>
                        <Value units="1/cm">120</Value>
                        <Accuracy>15.5</Accuracy>
                </BandWidth>
                <Modes>
                        <DeltaV modeID="V2">3</DeltaV>
                        <DeltaV modeID="V3">2</DeltaV>
                </Modes>
        </BandAssignment>

</CrossSection>
```

CHAPTER

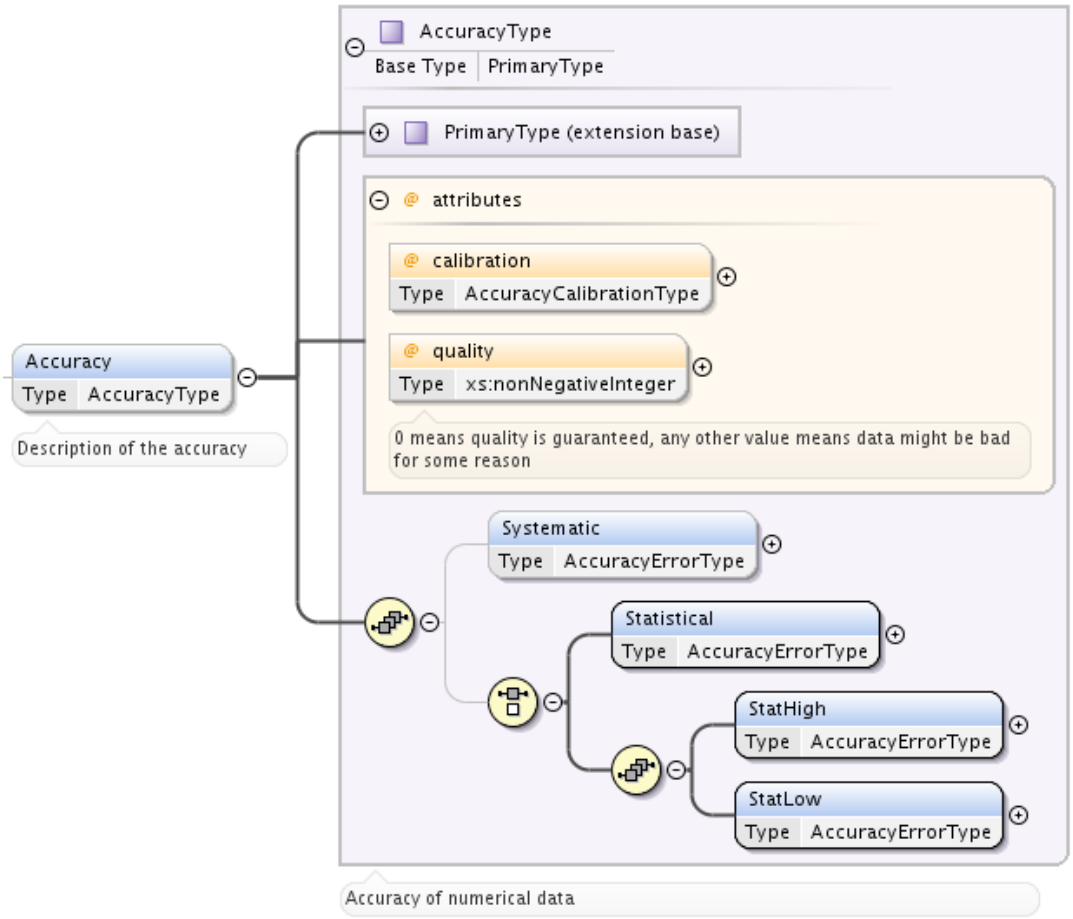
SIX

ACCURACY OF NUMERICAL DATA

For the sake of properly defining numerical data accuracy, new AccuracyType was been introduced, that is used now in the **Accuracy** element of **DataType** and **DataFuncType**.

Accuracy description was been inspired by the one in IVOA Spectral Data Model v. 1.03

6.1 AccuracyType



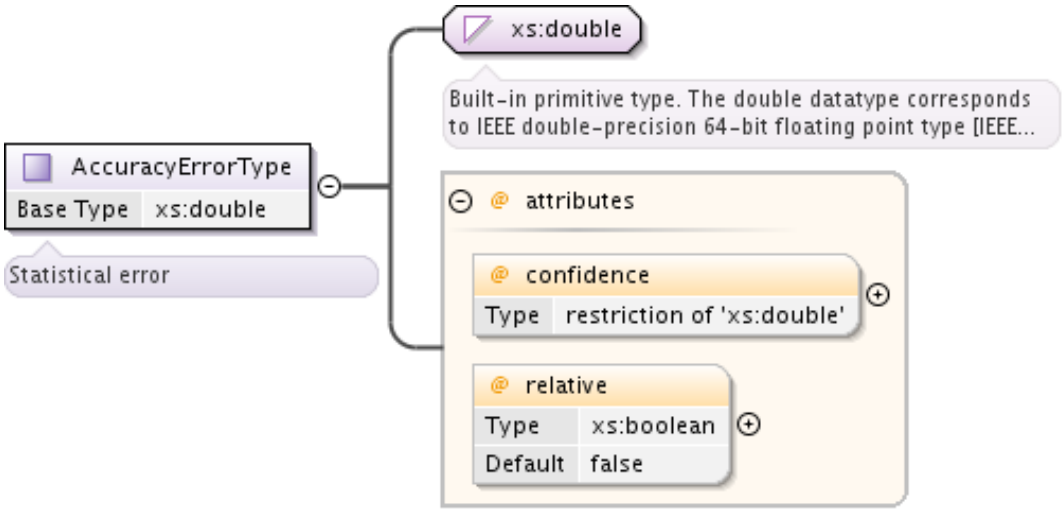
AccuracyType, extending **PrimaryType** allows to describe quality of data and has the following attributes and elements:

- Optional **calibration** attribute that describes the kind of reference frame for data. Possible values are:
  - **absolute** indicates that the values in the data are expected to be correct within the given uncertainty

  - **relative** indicates that although an unknown systematic error is present, the ratio and difference of any two values originating from the same source will be correct.
  - **normalized** indicates that the values, originating from this source, have been divided by a certain reference quantity. In this case units field of ValueType should be 'unitless'
  - **uncalibrated** indicates that not only an unknown systematic error is present in data, originating from that source, but also some unspecified value-dependant error. Thus, for example, for transitions frequencies only the order of transitions is guaranteed, neither frequencies, nor their difference/ratio are accurate.

- Optional **quality** attribute of integer type that may be used for distinguishing quality-assessed data. Zero value means data, accurate within their errors, other values means that there were some problems with data.
- Optional **Systematic** element of type **AccuracyErrorType** for systematic errors
- Optional **Statistical** element of type **AccuracyErrorType** for total statistical error, i.e. upper/lower range are equal.
- Optional **StatHigh** and **StatLow** elements group may be specified instead of single **Statistical** element to indicate statistical errors in case when upper and lower error ranges are not equal.

6.2 AccuracyErrorType



AccuracyErrorType is an extension of **xs:double** type, adding two attributes:

- **confidence** of type **xs:double**, with valid ranges from 0 to 1, indicating confidence level for which this accuracy is calculated. Ususal values would be like **0.95** or **0.99**.
- **relative** of type **xs:boolean**, indicating whether this accuracy value is absolute(**false**) or relative(**true**). By default, accuracy should be treated as absolute.