IAEA
International Atomic Energy Agency

INDC(NDS)- 0520
Distr. LP,NE,SK

INDC International Nuclear Data Committee

Summary Report of Consultants' Meeting

XML Schema for Atomic and Molecular Data

IAEA Headquarters, Vienna, Austria
1–2 October 2007

Prepared by

D. Humbert
International Atomic Energy Agency, P.O. Box 100, Vienna, Austria

January 2008

Selected INDC documents may be downloaded in electronic form from
http://www-nds.iaea.org/indc_sel.html or sent as an e-mail attachment.
Requests for hardcopy or e-mail transmittal should be directed to *services@iaeand.iaea.org*
or to:
Nuclear Data Section
International Atomic Energy Agency
PO Box 100
Wagramer Strasse 5
A-1400 Vienna
Austria

Printed by the IAEA in Austria

January 2008

INDC(NDS)- 0520
Distr. LP,NE,SK

Summary Report of Consultants' Meeting

XML Schema for Atomic and Molecular Data

Prepared by

D. Humbert

Abstract

Advanced developments in computer technologies offer exciting opportunities for new distributed tools and applications in various fields of physics. The convenient and reliable exchange of data is clearly an important component of such applications. Therefore, in 2003, the AMD Unit initiated within the collaborative efforts of the DCN (Data Centre Network) a new standard for atomic, molecular and particle surface interaction data exchange (AM/PSI) based on XML (eXtensible Markup Language). A working group composed of staff from the IAEA, NIST, ORNL and Observatoire Paris-Meudon, meets biannually to discuss progress made on the XML schema and to foresee new developments and actions to be taken to promote this standard for AM/PSI data exchange. This meeting is the second such gathering of these specialists in 2007.

January 2008

TABLE OF CONTENTS

Appendices

IAEA Consultants' Meeting: XML Schema for Atomic and Molecular Data

2–3 April 2007, IAEA Headquarters, Vienna, Austria

Present

Mr. D. Humbert, IAEA *(Scientific Secretary)*
Mr. A. Nichols (Opening of the meeting)

Attendees:

Mr. Yu. Ralchenko (United States of America)
Mr. D.R. Schultz (United States of America)
Ms. M-L. Dubernet (France)
Ms. E. Roueff (France)
Mr. R.E.H. Clark (IAEA)

Opening of the Meeting

Alan Nichols, Head of the Nuclear Data Section, opened the meeting. He underlined the interest of the nuclear data developers in the Section for our project, as well as IT specialists within the Agency.

1. NIST Workshop on A+M/PSI Data Exchange, 3–4 May 2007

This first workshop was organized at NIST by Yu. Ralchenko, and had been attended by approximately 15 participants. Yuri reported that most attendees were from the USA, two from Europe and one from Venezuela. Republic of Korea representatives could not attend as they could not organize their visas in sufficient time. The first day covered presentations and the second was open for extensive discussions. Data users and data producers showed a strong interest for an XML format in A+M/PSI data.

The next workshop will be organized within the ICAMDATA meeting in October 2008 in China.

The A+M/PSI data model was also presented to the following:

- IRON Project meeting, August 2007, Yuri Ralchenko,
- Atomic Spectra Workshop, Sweden, Yuri Ralchenko,
- CRAAMD, China, Marie-Lise Dubernet.

2. XML Schema

The schema is split into different "containers" to improve maintenance and development:

- general schema *Yuri Ralchenko*
- biblio *Dave Schultz*
- functions *Denis Humbert*
- methods *Denis Humbert*
- types and attributes *Yuri Ralchenko*
- statesAtoms *Yuri Ralchenko*
- statesMolecules *M-L Dubernet*
- statesSolids *Dave Schultz*
- singleReactanctProcess and multiReactantProcess *Bob, Denis, Yuri*

General Notes
The schema is now very advanced, and it is important to develop applications to ensure and improve the quality of this schema. XML libraries are available for most common programming languages like Perl and Python, and any files generated can be checked through these libraries. A XML data output is available for the NIST ASD database for energy levels, and will be extended to spectral lines. The IAEA ALADDIN program will provide such an output for the collision processes early in 2008.

Labelling and structure:
- *label* changed to *description* for the whole schema,
- comments tags to be labelled the same way: *comments* → *comment*

With respect to size of the XML file for the atomic states, a compression average of 45 has been achieved.

Altova XMLSpy provides a tool to generate UML from XML data model. More legibility is achieved by defining "types" when possible. More remarks on the model consistency are noted and summarised at the end of this document. A formal document for the data model structure will be elaborated.

Action: draft document (*Denis*)
Action: check the whole data model for consistency (*everyone*)

Some data representations do not comply with the schema as isoelectronic series, isonuclear sequences, quantum defects, average atom.

Atom
The AtomState type has no major changes. Some lists may be switched from ordered (sequence) to random (all), e.g. isotope type.

Molecule
The molecular excited state is very complex. Developing a detailed model which includes all possible descriptions is impossible as the physics is still evolving. Marie-Lise presented the draft document "Molecular State Description". A general description using an ASCII or latex string is available, while a detailed description is still in progress. The following remarks were made:

- Include *nuclearSpin* element in the *OrdinaryStructureFormula* type, e.g. Sections 3.2.1 and 3.2.8.
- Rename *molecularStateType.name* as *molecularStateType.description*
- Molecular state: change the presentation for excited, vibrational and rotational from the linear form to series. Question is where best to place *molecularStateID.*
- Estimate the feasibility of modelling the symmetry group – if too complex, a latex description is also possible.
- Turn *parity* element into an enumeration.
- *RotationQN* type: electronic coupling is properly identified with the intermediate couplings. Also look at *Origin1*, *Origin2* and *QN*.
- The vibrational part is not yet complete.
- Mixing coefficient could be signed or squared. Reference to the method?

Action: update schema and "Molecular State Description" document *(M-L)*

Surface/Solid
The SurfaceState type is renamed solid State to include dust, ices and fulfil the needs in fusion energy research and astrophysics, as well as the needs of industrial plasmas. Regarding the complexity of solids, only a simple schema is proposed to give:

- description of single- and multi-layer materials,
- description of mixed and single materials, with composition and properties,
- temperature (optional).

The schema needs some rearrangement, as some parts are redundant.

Action:
- rearrangement of the schema *(Dave),*
- check the compatibility with collisions *(Denis),*
- IAEA CRP on surface erosion: first RCM (17-19 October 2007) should include a short presentation of AMDML to solicit comments from the PSI community *(Denis).*

Bibliography
The web only provides projects in development with a level of complexity that is too high for our needs. Our present simple approach is sufficient for our needs. Attribute *bibMethRefs* is split into two attributes: *bibRef* and *methodRef.*

Action: update the model *(everyone).*

Processes
This section has been reorganized according to the numerical data type provided. Three data types are identified:

- transition probabilities for radiative processes (*Data type*),
- transition probabilities for autoionization (*Data type*),
- cross sections, rate coefficients and reaction rates for collisions (*DataXY type*).

Radiative processes and autoionization are separate to reflect their own specificity.

An agreement for the *DataXY* type was reached. Errors can be reported: positive or negative dx and dy. The x units are given in the *xType* (former *arguments*); the y units in the yType (former *collParamType*). Type of errors are systematic, computing, experimental, to be compared with IVOA.

Additional changes to be considered:

- *collParamtype* → *yType* and *arguments* → *xType*,
- process choice: arbitrary naming and/or list of codes from the "Classification of Processes" document.

  Actions:
  - send document "classification of processes" *(Denis)*,
  - add enumeration of the processes codes *(Denis)*,
  - verify the molecular radiative transition part and check for possible additions *(Evelyne)*.

3. Namespace, Registry, DAL and Web site

These topics should be developed through a recognized international group, although the creation and organization of such a body remains problematic one recommended action was to create our own web site. Yuri confirmed NIST can host the web site. A web page on XML is also posted on the IAEA A+M data unit web site (http://www-amdis.iaea.org/xml). Denis presented a web page developed for NIST, which based on the NIST hosted web page "International Committee for Radionuclide Metrology (ICRM)". Discussions resulted in the following suggested actions:

- Find a logo and a proper name that also refers to PSI. Two names were proposed without great enthusiasm: AMSML and AMSDML *(everyone)*.
- Develop a web page to be posted at NIST *(Yuri)*.
- Modify the web page at IAEA *(Denis)*.
  - bullets instead of numbers
  - remove reference to KAERI and JAEA
  - contact line: inquiries and comments

4. Collaboration

Yuri reported strong interest from Korea and Russia to participate in this project. Korea (NFRC and KISTI) is developing databases for A+M/PSI with special focus on low temperature plasmas. Russia is interested through the SPECTR-W3 project.

Bob stated that the IAEA is unable to join any European Union FP7 collaborating projects as a full member.

Question is how and where to develop the proposed XML project in terms of manpower and funding. IAEA, Observatoire de Paris and ORNL do not possess the capabilities.

5. Draft Document

A draft document is recommended for each container, with the UML and the element description. The IOVA document presentation can be used (e.g. M-L document).

6. Work Plan and Time Schedule

The first release of the XML schema is planned for June 2008. Application of the schema to existing databases is essential for its validation. The NIST Atomic Spectra Database (ASD) has already applied the schema to energy levels and is extending the application to spectral lines. A+M Data Unit will also upgrade the ALADDIN web interface to incorporate the schema for collision processes.

The ICAMDATA gathers together data centre, data users and data producers in AM/PSI research. The meeting to be held in October 2008 in China is a further opportunity to present our work. Jun Yan (local organizer of the conference) and most members of the Scientific Board will be present in Vienna for the DCN meeting (3-5 October 2007) where progress on the schema will be presented. A request to have a presentation and/or panel discussion during the ICAMDATA will be formulated.

Following reviews of individual topics, actions were proposed and milestones formulated:

- Denis to send the list of processes “Recommended Classification of Processes”.
- Yuri to circulate the XML schema.
- DCN Meeting, 3-5 October 2007.
- Presentation of the XML schema at first RCM of the IAEA CRP on “Surface Erosion”, 17-19 October 2007.
- XML meeting, 6-7 December 2007, Paris
    - home page proposition to be posted at NIST *(Yuri)*,
    - review of each container.
- XML meeting, 6-7 May 2008, Vienna.
- Version 1.0 of the XML schema, June 2008.
- XML meeting, 23-24 October 2008, China.
- ICAMDATA, 28-31 October 2008, China.

7. Next Meetings

- 6-7 December 2007, Observatoire de Paris, France,
- 6-7 May 2008, IAEA Vienna,
- 23-24 October 2008, immediately before the ICAMDATA, China.

8. Data Consistency

During the meeting, the following statements were made regarding data consistency, and will assist in elaborating the structure of the data model.

Types:

- Types may be labelled with *extension* when referring to a global type. Consistency must be checked. TransitionProbabilities: data type can be defined with the extension label (oscillatorStrength) and without (lineStrength).
  Action for each container.

Names:

- Name syntax: first letter to be lower case, uppercase for each new word (*atomicState*). Types begin with an uppercase in a text document.
- Full name or abbreviation: *oscStrength* vs *oscillatorStrength.*

Structure:

- All groups must be defined, with begin and end tags:
  particles → *particle* →….
- Unbounded and bounded tags appear in the order: unbounded and bounded

Appendix A

IAEA Consultants' Meeting: XML Schema for Atomic and Molecular Data

1–2 October 2007, building-C, floor-7, room-51 (C0751), IAEA Headquarters, Vienna, Austria

---

LIST OF PARTICIPANTS

Dr. Marie-Lise Dubernet
Observatoire de Paris
5, Place Janssen
F-92195 Meudon Cedex
FRANCE
Tel.: +33-1-4507-7446
Fax: +33-1-4507-7100
E-mails: marie-lise.dubernet@obspm.fr

Dr. Evelyne Roueff
LUTH & UMR 8102 du CNRS
Observatoire de Paris
5, Place J. Janssen
F-92190 Meudon
FRANCE
Tel.: +33-1-4507-7435
Fax: +33-1-4507-7971
E-mail: evelyne.roueff@obspm.fr

Dr. Yuri Ralchenko
Atomic Spectroscopy Group
Stop 8422
National Institute of Standards and Technology
Gaithersburg MD 20899-8422
U.S.A.
Tel.: +1-301-975-3210
Fax: +1-301-975-1350
E-mail: yuri.ralchenko@nist.gov

Dr. David R. Schultz
Physics Division
MS-6372, Bldg. 6003
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN 37831-6372
U.S.A.
Tel.: +1-865-576-9461
Fax: +1-865-576-1118
E-mail: schultzd@ornl.gov

I.A.E.A.

Dr. R.E.H. Clark
IAEA Atomic and Molecular Data Unit
Wagramerstrasse 5
P.O. Box 100
A-1400 Vienna
AUSTRIA
Tel.: +43-1-2600-21731
Fax: +43-1-26007
E-mail: r.e.h.clark@iaea.org

Dr. D. Humbert
IAEA Atomic and Molecular Data Unit
Wagramerstrasse 5
P.O. Box 100
A-1400 Vienna
AUSTRIA
Tel.: +43-1-2600-21729
Fax: +43-1-26007
E-mail: d.humbert@iaea.org

Appendix B

IAEA Consultants' Meeting: XML Schema for Atomic and Molecular Data

1–2 October 2007, building-C, floor-7, room-51 (C0751), IAEA Headquarters, Vienna, Austria

---

MEETING AGENDA

Monday, October 1

09:15 Opening

Morning

1. NIST Workshop on A+M/PSI Data Exchange, 3-4 May 2007. *Evelyne*

2. Review of present AMDML status
- Overview: *Yuri*
- Atomic data *Yuri*
- Solid *Dave*
- Single reactant and multi reactants processes *Denis, Yuri, Bob*
- Biblio *Dave*
- Functions, methods *Denis*
- Types and attributes *Denis*
- Actions from last meetings *Yuri*

Afternoon

3. Comments, corrections, new developments and unresolved issues from last meeting
- list of processes
- element "dataXY"
- list of data types: cross sections, reaction rates…
- structure consistency
- units, link with UNITSML
- isoelectronic series

Estimated end of first day 17h

Tuesday, October 2

09:00 Meeting Continued

Morning

4. Web site *Denis*
5. Space name for "containers" *(pending item)*
6. Data Access Layer *(pending item)*
7. AMDML web registry *(pending item)*
8. Draft document *(pending itme)*

Afternoon

9. Collaboration
10. Call of the EU FP7 concerning building an infrastructure for exchange of atomic and molecular data
11. Work plan and milestones
    a. Version 1.0
    b. Next Workshop on "AMD/PSI Data Exchange", spring 2008
    c. Dates for next meetings

Estimated close time 16h

Nuclear Data Section
International Atomic Energy Agency
P.O. Box 100
A-1400 Vienna
Austria

e-mail: services@iaeand.iaea.org
fax: (43-1) 26007
telephone: (43-1) 2600-21710
Web: http://www-nds.iaea.org