



IAEA

International Atomic Energy Agency

INDC(NDS)- 0532

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

15–16 May 2008

Prepared by

D. Humbert

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

August 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@iaeaand.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

August 2008

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 distribution 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 A+M Data 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. Such a meeting was held on 15, 16 May 2008, and the discussions and progress made in the schema are considered within this report.

August 2008

TABLE OF CONTENTS

1. Presentation of the Schema at Various Technical Meetings.....	7
2. Review of Status of XML Schema.....	8
3. Processes.....	10
4. SPECTR-W ³	11
5. Namespace, Registry, DAL, Schema Name.....	11
6. Draft Document.....	11
7. Collaboration.....	11
8. ICAMDATA, China, October 2008.....	12
9. Working Method, Working Plan and Time Schedule.....	12
10. Future Meetings.....	12
11. Naming and Schema Structure Rules.....	12

Appendices

Appendix A: List of Participants.....	15
Appendix B: Meeting Agenda.....	17

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

15–16 May 2008, IAEA Headquarters, Vienna, Austria

Present

Mr D. Humbert, IAEA (*Scientific Secretary*)

Mr A.L. Nichols (Opening of the meeting)

Attendees:

Ms M-L. Dubernet (France)

Ms E. Roueff (France)

Mr P. Loboda (Russian Federation)

Mr S. Gagarin (Russian Federation)

Mr Yu. Ralchenko (United States of America)

Mr D.R. Schultz (United States of America)

Mr R.E.H. Clark (IAEA)

Welcome and Opening of the Meeting

Participants were welcomed to the IAEA by Mr Nichols on behalf of the IAEA and the Nuclear Data Section. He spoke about the obvious good progress made on the project, and commented on a growing interest of the nuclear data developers in the Section, as well as IT people in the Agency. With the announced first release of the schema, he underlined the necessity to continue to keep a clear timetable with well-defined milestones.

Denis Humbert welcomed two new collaborators: Messrs Peter Loboda and Sergey Gagarin from the All-Russian Institute of Technical Physics, Russia.

1. Presentation of the Schema at Various Technical Meetings

Progress on the project was presented to the IFRC A+M subcommittee by D. Humbert in April 2008. This committee has the mandate and responsibility to review the activity of the IAEA A+M Data Unit periodically, and to recommend policies and programme priorities. The committee recognizes the importance of the project and renewed support for the Agency to organize technical meetings on XML. The need for the A+M Data Unit to establish and maintain databases in support of the fusion energy research and to keep a good time balance between the different projects were also underlined. Yu. Ralchenko asked if there is a possibility to organize a CRP on XML for AM/PSI data. A new CRP on XML could not be quickly started. A series of specific CRPs are already scheduled for the next few years and topics for new CRPs are addressed by the IFRC subcommittee which will meet in 2010.

M-L Dubernet reminded participants that it is possible to find new support through the I3-FP7. I3-FP7 is a European initiative to support multidisciplinary project in I-science. Foreseen participants are CDMS, ADAS, Chianti. USA and Russia can join an I3-FP7 project as long as the project leader is European. The IAEA, as an international organisation, could not be part of this project. The next deadline to submit a proposal is the end of September 2008.

2. Review of Status of XML Schema

The schema is split into different “containers”: *biblio* has been renamed *sources*, and *autoionization* changed to *nonRadiative*.

- general schema *Yuri Ralchenko*
- sources *Dave Schultz*
- functions *Denis Humbert*
- methods *Denis Humbert*
- types and attributes *Yuri Ralchenko*
- statesAtoms *Yuri Ralchenko*
- statesMolecules *M-L Dubernet*
- statesSolids *Dave Schultz*

- processes
 - radiative *Evelyne Roueff*
 - non-radiative *Evelyne Roueff*
 - collisions *Denis Humbert*

A general clean-up of the schema has been made by Yu. Ralchenko. Annotations have been added. Patterns have been introduced for some attributes and elements. As an example, for the chemical elements, the following pattern is used: $\{Lu\}\{Ll\}?$, uppercase followed by an optional lowercase letter. As the discussion proceeded, corrections were agreed and future actions mentioned:

- Abbreviations should be used as little as possible in the naming of elements and attributes (see naming rules). A British English dictionary is used for the naming.
- For attributes of ID type and Ref type, the naming is *nameID* for the target and *nameRef* for the reference. Ex: *methodID* and *methodRef*. Pattern rule for the naming of ID is the following “[SMBF].+”, where S refers to state, M to method, B to source, and F to function.
- sourceRef and methodRef are optional everywhere, but one of each should at least be defined in the xml tree, although unable to impose in an xml schema. sourceRef and methodRef should only refer to numerical data except for methods, where a bibliographic reference can be given. Marie-Lise and Yuri will give examples to explain where this information is required.
Action: Marie-Lise and Yuri.
- Check in the schema what should be optional and mandatory. Sources and states are the only mandatory elements at the primary level.
- The naming solid is preferred to surface.

Action: Check the consistency of the schema according to the previous remarks (all).

AttrType (attributes and types)

Recognized that to have global types and attributes gives greater legibility to the schema and an easier maintenance. For types referred only once or twice in the schema (collParamType, methodCategory), the choice between local or global has to be carefully examined.

Some types are labelled “ext” for extension – should be removed if unnecessary.

Action: Denis and Yuri to find out the rules and adequacy in using the option “extension” when defining a type.

Sources

The biblio element is renamed *sources*. “source” is preferred to “reference” or “bibliographicReference” to avoid ambiguity with the term reference used in computer science. The leading “bib” in the naming of each element is removed.

Methods

A method can have a bibliographic reference.

Data types

- *Accuracy or uncertainty*: For the schema, “accuracy” refers to an average uncertainty on a whole data set or fit function. For example the element “fitAccuracy” gives the precision on how the function fits the data points. “uncertainty” refers to single points. The word “uncertainty” is preferred to “error” or “errorBar”. An error can be absolute or relative.
- For units, UNITSML from NIST will be used when available.

Atom

- *shellType*: is Kappa needed? **Action** Yuri,
- *shellPairType*: is shellPairID(Ref) is needed as attribute? **Action** Yuri,
- for a better definition of isotopes, the element *massNumber* is added.

Molecules

Marie-Lise presented the new schema for molecules. Developing a detailed model which includes all possible descriptions is impossible, as the physics is still evolving. Nevertheless, description of rotational and electronic excited states is complete. More expertise is needed for vibrational excited states. The description of an excited state is linear: first electronic, then vibrational, and finally rotational.

The quantum state may be given in two ways, and both are optional:

- element *description* is a character string, with an optional attribute “encoding”– possible values are ASCII, html, latex, user...
- in symbolic form, using boxes.

Also noted that the mixing coefficient type should be the same for atoms and molecules. The naming “hyperfineMomentum” will be checked. The molecule data model will be tested before the release of version 0.1.

Action:

- verify the naming “hyperfine momentum” (Marie-Lise)
- test the data model validity for molecules on BASECOL (Marie-Lise)
- global mixing coefficient type (Marie-Lise, Yuri)

Solid, surface

Yuri presented a simple data model for the solids. This first approach should fulfil the elementary needs in fusion energy research, astrophysics and industrial plasmas. A solid is formed of one or multiple layers, each layer having the same data structure. No solid properties with numerical data are for the moment implemented and for consistency a method and bibliographic reference may be given. The wording “solid” is preferred to “surface”. The data model will be tested on the IAEA ALADDIN database before the release of version 0.1 of the schema.

Action: test the data model validity for solids on ALADDIN (Denis),

Processes

This container has three sections:

- transition probabilities for radiative processes (*Data type*)
- transition probabilities for non-radiative processes (*Data type*)
- collisions (*DataXY type*)

Radiative and non-radiative processes are separate to reflect their own specificity. This organisation was again shortly discussed. Participants were not convinced that this procedure reflects the physics in reality. However, the same organization was maintained because no better solution was proposed. Autoionization is renamed non-radiative process.

Radiative and non-radiative processes: reviewed by Evelyne. Air broadening and Stark broadening will be introduced into the schema later. No change is needed in the actual version.

Collisions (*collDataType*): renewed by Denis, who proposed a simplified schema. “dataDescriptor” (former parameter and collParamType) stays as an attribute. Values for “dataDescriptor” are collision strength, cross sections... As the general structure does not give full satisfaction (e.g. place of dataDescriptor), Denis will review this part.

Action: Denis to review the collision part.

Labelling and data model structure

Denis presented a compilation of rules formulated during the previous meeting for the data model structure. This activity is important to the maintenance of the consistency of the data model. Some unclear formulations were removed or modified (see 10. Naming and Schema Structure Rules).

Unresolved issues

Some data representations such as isoelectronic series, isonuclear sequences, quantum defects, average atom, branching ratio, band properties and solid spectroscopy do not comply with the current schema. These topics are deferred to a future release of the schema.

3. Processes

The element *process* is complex with two optional elements. A user-defined element and a list of codes (for example: dissociative recombination will have two values: dissociation and recombination). Two documents are available: “Specifications for the IAEA ALADDIN Surface Database Interface” and “IAEA Classification of Processes”.

Action: Dave to propose a simplified list of processes for the surface interactions and for the collisions.

4. SPETR-W³ Presentation

Petr Loboda presented the SPECTR-W³ project. SPECTR-W³ is an atomic numerical database, available on the Web at <http://spectr-w3.snz.ru>. The local version for the off-line operation (SPECTR-CD) can also be downloaded from the SPECTR-W³ project homepage upon the user's request. SPECTR-W³ includes spectroscopic and collisional data.

5. Namespace, Registry, DAL, Schema Name and Logo

These topics are still pending. Except for the schema name, they are not a priority in the project time schedule. Proposed names for the schema are SAMS or XSAMS, which stand for [XML] Schema for Atoms, Molecules and Solids. A final decision will be taken in July.

6. Draft Document

The first release of the schema, previously scheduled in June 2008, is postponed to September 2008. A draft documentation should be ready by the time of the next meeting. Yuri quoted that XMLSpy has an option to generate a documentation from an XML schema.

- overview document *Yuri Ralchenko*
- sources *Dave Schultz*
- functions *Denis Humbert*
- methods *Denis Humbert*
- types and attributes *Yuri Ralchenko*
- statesAtoms *Yuri Ralchenko*
- statesMolecules *M-L Dubernet*
- statesSolids *Dave Schultz*
- processes
 - radiative *Evelyne Roueff*
 - non-radiative *Evelyne Roueff*
 - collisions *Denis Humbert*

From minutes of the December meeting: IOVA document presentation will be used (e.g. Marie-Lise document).

Action: Marie-Lise to send a Latex template.

7. Collaboration

We were pleased to welcome colleagues from Russia. They showed a strong interest in the schema and are ready to produce outputs in XML for the SPECTR-W³ database, once the first release available. Their contribution to the meeting was very valuable. They will follow closely progress made on the schema and will participate in further meetings.

Marie-Lise reported on possible collaboration through the European I3-FP7 projects (see 1. Presentation of the Schema at Various Technical Meetings).

There is still a strong interest from China to participate in the project. The next ICAMDATA is an opportunity to extend this collaboration.

8. ICAMDATA, China, October 2008

As well as an invited talk to the conference, the schema will be publicized through several posters:

- Version 0.1 of the schema,
- XML application on the IAEA ALADDIN Database,
- XML application on the NIST ASD Database,
- XML application on the Observatoire de Meudon, BASECOL Database.

9. Working Method, Working Plan and Time Schedule

The two important milestones are the release of version 0.1, and an announcement at the ICAMDATA. A new meeting is scheduled in Vienna in the last week of August. Before the meeting all xml documents and reports will be posted on the AMDML Google group at (<http://groups.google.com/group/amdml>) to fulfil the requirements to release the first version of the schema:

- 30 Jun 2008 Proposal for processes list, surface interactions and collisions (*Dave*)
- 11 Jul 2008 Corrections and first agreement on the list (*everyone*)
- 11 Jul 2008 Agreement on the name. SAMS, XSAMS or other (*everyone*)
- 15 Jul 2008 NIST web site (*Yuri*)
- 15 Jul 2008 XML application for molecules at Meudon, BASECOL (*Marie-Lise*)
- 15 Aug 2008 Deadline for abstract submission at ICAMDATA
- 22 Aug 2008 Draft documentation posted on AMDML Google group (*everyone*)
- 26 Aug 2008 XML application for surfaces at IAEA, ALADDIN collisions (*Denis*)
- 26-27 Aug Meeting in Vienna
- **31 Aug 2008 Version 1.0 of the XML schema, 2008**
- End Sep Deadline for I3-FP7 proposal
- 27 Oct 2008 XML meeting in Beijing, China
- **28-31 Oct ICAMDATA, Beijing, China**

10. Future Meetings

- 26-27 August 2008, IAEA Vienna
- 27 October 2008, before the ICAMDATA, China

11. Naming and Schema Structure Rules

Naming of elements, types and attributes:

- Only letters [a-z,A-Z] are used for the naming – no space, dash or underscore
- First letter in lower case. Each new word begins with an uppercase
- The words are taken from a British English dictionary (*ex: transitionProbability, DOI*)
- Acronyms are kept in uppercase (*ex: DOI*)
- Full names are preferred to abbreviations (*ex: oscillatorStrength vs oscStrength*)

Structure;

- All groups must be defined with a begin tag and an end tag:

<particles>

<particle>

.....

</particle>

<particle>

.....

</particle>

</particles>

- Bounded or unbounded element and sequence: unbounded sequence, then bounded element.
- Global types are preferred to local types for better legibility.

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

15–16 May 2008, building-A, floor-23, room-13 (A2313), 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. Sergey V. Gagarin
Head of the Laboratory
Applied Programming and Database
Developments
Russian Federal Nuclear Center
All-Russian Institute of Technical Physics
13, Vasilyeva st., Snezhinsk
Chelyabinsk Region, 456770
RUSSIAN FEDERATION
Tel.: +7-351-46-56560
Fax: +7-351-46-55118
E-mail: s.v.gagarin@vniitf.ru

Dr. Peter A. Loboda
Chief Scientist
Theoretical Division
Russian Federal Nuclear Center
All-Russian Institute of Technical Physics
13, Vasilyeva st., Snezhinsk
Chelyabinsk Region, 456770
RUSSIAN FEDERATION
Tel.: +7-351-46-56560
Fax: +7-351-46-55118
E-mail: p.a.loboda@vniitf.ru

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

Friday, 16 May

09:00 Meeting Continued

Morning

6. AMDML applications
 - ASD *Yuri*
 - ALADDIN *Denis*
 - BASECOL *Marie-Lise*
7. Web site: IAEA (*Denis*), NIST (*Yuri*)
8. Draft document, (*Yuri*)
9. Space name for “containers” (*pending item*)
10. Data Access Layer (*pending item*)
11. AMDML web registry (*pending item*)
12. Name for the schema

Afternoon

13. Collaboration
14. Work plan and milestones
 - a. Version 1.0
 - b. ICAMDATA “AMD/PSI data Exchange” invited paper
 - c. Dates for next meetings

Estimated closing 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>
