INDC International Nuclear Data Committee

XML Schema for Atoms, Molecules and Solids (XSAMS)

Summary Report of an IAEA Consultants' Meeting

National Institute of Standards and Technology (NIST)

Gaithersburg, MD, United States of America

3-5 October 2011

Prepared by

B. J. Braams

December 2011

Selected INDC documents may be downloaded in electronic form from
j wr⊲lly y y /pf u0kc gc 0qt i lt gr qt w/pgy lkpf e/t gr qt wl 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
Vienna International Centre
A-1400 Vienna
Austria

Printed by the IAEA in Austria

December 2011

XML Schema for Atoms, Molecules and Solids (XSAMS)

Summary Report of an IAEA Consultants' Meeting

National Institute of Standards and Technology (NIST)

Gaithersburg, MD, United States of America

3-5 October 2011

Prepared by

B. J. Braams

Abstract

A Consultants' Meeting on "XML Schema for Atoms, Molecules and Solids (XSAMS)" was held at the National Institute of Standards and Technology (NIST) in Gaithersburg, MD, United States of America, 3-5 October 2011. Objectives of the meeting were to review and discuss developments of the Schema made during 2011 in connection with implementations on databases associated with the Virtual Atomic and Molecular Data Centre (VAMDC) and to agree on the adoption of an international standard XSAMS version 1.0. The proceedings of the meeting are summarized here.

December 2011

TABLE OF CONTENTS

1. Introduction

The XML Schema for Atoms, Molecules and Solids (XSAMS; see http://www-amdis.iaea.org/xsams/) had its initial release, version 0.1, in September 2009. The first implementations were carried out for the Atomic Spectra Database at NIST, the ALADDIN database at IAEA, the BASECOL database at the Paris Observatory and the Spectr-W3 database at RFNC-VNIITF. In 2010 the Virtual Atomic and Molecular Data Centre (VAMDC; see http://www.vamdc.org) decided to adopt XSAMS as their standard for data exchange and implementations commenced on all databases represented in VAMDC. In connection with these implementations VAMDC has extended the schema so that there are now two principal branches of XSAMS: the IAEA XSAMS, version 0.1.1 following minor changes made in January 2011, and VAMDC XSAMS, now version 0.3, which contains a different ("case-by-case") treatment of molecular spectroscopy, enhancements in the area of environmental (plasma) influences and lineshapes, and many smaller changes.

The present Consultants' Meeting was organized in order to review the experience gained through the ongoing implementations of XSAMS, review and discuss modifications made to the Schema by VAMDC, discuss adoption of an international standard XSAMS version 1.0 and make plans for further developments of XSAMS.

The meeting was hosted by the NIST and this made it possible to have several additonal participants from the United States fusion and A+M data community. The IAEA is grateful to Dr Yuri Ralchenko and the NIST Atomic Spectroscopy Group for hosting the meeting. The meeting programme included a tour of the NIST Electron-Beam Ion Trap (EBIT) experiment. EBIT experiments have revolutionized experimental atomic spectroscopy and it was valuable to see an example and to learn how the device is used.

Section 2 provides an Executive Summary of the Consultants' Meeting. Section 3 contains the meeting proceedings and Section 4 contains the conclusions of the meeting. Appendix 1 contains the list of participants and Appendix 2 contains the meeting agenda.

2. Executive Summary

The presentations at the meeting are available on the A+M Data Unit web site http://www-amdis.iaea.org/meetings/ via the link to Meeting Reports and Presentations.

The meeting heard of the ongoing implementations of XSAMS on approximately 25 databases associated with the Virtual Atomic and Molecular Data Centre (VAMDC) and discussed developments of the Schema made by VAMDC. Participants are pleased with the success of this experience. In the course of their implementations the VAMDC team made several choices that extend XSAMS-0.1.1 to VAMDC XSAMS version 0.3. Major issues have been the treatment of the environment, lineshapes and molecular spectroscopy, but many smaller developments took place as well. For documentation of IAEA XSAMS please see http://www-amdis.iaea.org/xsams/documents/ and for documentation of VAMDC XSAMS please see http://www.vamdc.org/documents/standards/. The participants agree in general terms with the choices made by VAMDC.

The following decisions were reached.

There will be an XSAMS Steering Committee meeting and we will work with VAMDC to organize an XSAMS technical meeting or technical session in conjunction with the VAMDC annual meeting at the University of Vienna, Tue-Fri 21-24 Feb 2012. We will use the Monday too, also on the university campus rather than at IAEA.

We intend to define precisely and adopt XSAMS-1.0 at that meeting in February. We expect that it will follow VAMDC XSAMS-0.3 or the then current VAMDC XSAMS quite closely, but maybe

some features will be left out of XSAMS-1.0. In any case, VAMDC XSAMS should be compatible with XSAMS-1.0.

We endorse VAMDC XSAMS-0.3 for present implementations of XSAMS also outside VAMDC.

We defer new PMI work in XSAMS until after the Feb 2012 meeting; first we want to show XSAMS-1.0 to the world and we want to see the user experience with VAMDC XSAMS.

With respect to solids in XSAMS we note the work on SSDM (solid state data model) and the GhoSST database at Grenoble. We wish to hear about solid state spectroscopy in relation to XML at our meeting in February.

We will advertise to our data centre partners (e.g., NFRI, NIFS, CRAAMD) the VAMDC tutorials for XSAMS implementors and users. We note that it is desirable to hold a training session somewhere in Asia some time, and we should work with the Asia-Pacific Atomic Data Network (APAN) on this.

We intend to attach some XSAMS event to ICAMDATA 2012, which will be held at NIST from Sunday evening 30 September to Thursday 04 October 2012. We may use Sunday 30 September or Friday 05 October.

3. Meeting Proceedings

Opening

The meeting was opened at 09:30 on Monday by Dr Carl J. Williams, Chief of the Quantum Measurement Division of the Physical Measurement Laboratory at NIST. The Atomic Spectroscopy Group, which hosted our XSAMS meeting, is a part of the Quantum Measurement Division. Dr Williams noted the long history of database work within the Atomic Spectroscopy Group on atomic energy levels and on transition probabilities and line shapes. The Atomic Spectra Database (now Version 4) is the largest product of the group. It provides critically evaluated data on energy levels, wavelengths and transition probabilities. The database represents many years of work and serves a large international community of users. Dr Williams expressed the wish for a fruitful meeting on developments in data exchange.

Dr Yu. Ralchenko, guest researcher in the Atomic Spectroscopy Group and local organizer of the meeting, added his welcome. The meeting has 2.5 days of work ahead of it. In particular we wish to review the experience that was gained during the past year with implementation of XSAMS on a diverse set of databases, discuss the developments in the schema in the area of environment, line shapes, molecular spectroscopy, discuss further developments, especially in molecular processes, and agree on the adoption of XSAMS version 1.0.

Participants then briefly introduced themselves and mentioned their connection with XSAMS. Dr N. Piskunov of the University of Uppsala is one of the coordinators of the Vienna Atomic Lines Database (VALD) and is task force leader for publishing tools for the Virtual Atomic and Molecular Data Centre (VAMDC). Dr V. Popova is a physicist and Dr S. Gagarin is a programmer with the SPECTR-W3 atomic spectroscopy project at the Russian Federal Nuclear Centre All-Russian Institute of Technical Physics (RFNC-VNIITF) in Chelyabinsk region. Dr C. Ballance of the Department of Physics at Auburn University, Alabama, USA, works on development of an R-matrix suite of codes in connection with the Atomic Data and Analysis Structure (ADAS) project. Dr Yu. Ralchenko of NIST works on atomic physics and plasma spectroscopy and is one of the originators of XSAMS. Dr Ch. Hill of the Department of Physics and Astronomy at UCL works on computational molecular spectroscopy and on the development of XSAMS. Dr E. Roueff of the Paris Observatory, Meudon, studies molecular spectroscopy of the interstellar medium and has participated in the development of XSAMS since the start. Dr D. P. Stotler of Princeton Plasma Physics Laboratory uses atomic and molecular data in fusion plasma modelling, especially in connection with the Degas-2 Monte Carlo

neutral gas transport code. Dr B. J. Braams is head of the Atomic and Molecular Data Unit at IAEA, which coordinates the development of XSAMS as an international standard.

Presentations

Yu. Ralchenko: XSAMS in 15 Minutes

There are a large number of publically available databases or data collections in the field of atomic, molecular and particle-surface interaction (AMPSI) processes and the typical size of these datasets has increased greatly over the years. A standard for exchange of AMPSI data is needed so that scientists can develop interfaces and tools that can be used across databases. Twenty years ago the ALADDIN format was developed for this purpose, but it is outdated. New methods based on the Extensible Markup Language (XML) are available and are well tested.

The XML Schema for Atoms, Molecules and Solids (XSAMS) was developed in order to serve as a modern standard for exchange of AMPSI data. To the extent that it is practical an XSAMS file encodes the whole A+M (+PSI) physics that underlies the data, including the precise quantum mechanical state or states of the reactants. Thereby XSAMS provides a recipe to create good XML for A+M+PSI applications. The initial proposal for XSAMS was discussed at the IAEA Atomic and Molecular Data Centres Network (DCN) meeting in 2003 with follow-up at the ICAMDATA conference in Toki City, Japan, in 2004. There was a steady, but slow development 2003-2009, with regular meetings coordinated by the IAEA A+M Data Unit. In 2009 the Virtual Atomic and Molecular Data Centre (VAMDC) became an important, and in fact leading, participant in XSAMS development. XSAMS Version 0.1 was released in September 2009.

In an XSAMS dataset every data value is embedded in hierarchically nested labelled fields that identify the nature of the value. Constraints can be imposed that are validated as the file is created or read, for instance that an entry in a field PrincipalQuantumNumber must be a positive integer. An XSAMS file is finally ASCII text and readable by humans, although it is meant to be written with use of a special-purpose editor. The file is large relative to the amount of data that is stored in it, but it can be efficiently compressed for transmission.

The present status of XSAMS is that development is lead by the VAMDC team. The XSAMS Steering Committee makes the final decisions on the Standard, but it meets only about once per year. Many VAMDC partners have now developed an XSAMS interface to their databases; more interaction between VAMDC and non-VAMDC database owners is needed.

N. Piskunov: Overview of VAMDC and VSS1/2 Demo

This presentation is based on the one given on 2011-09-16 at the EU second-year review of VAMDC. Please note the main VAMDC web site: http://portal.vamcd.eu/.

XSAMS is the common language for communicating data in the Virtual Atomic and Molecular Data Centre, VAMDC. It is the basis of the query language and the language in which query results are returned; however, tools are provided that make it possible to flatten an XSAMS structure into tables or arrays. XSAMS is also used for automatic registry updates.

VAMDC-TAP (Transaction Access Protocol) is the interface between an application and a data service. The query language is an SQL subset, currently VSS1 (VAMDC SQL Subset #1), but a second iteration, VSS2 is due soon and is expected to constitute the final version of the query language for the VAMDC project. VAMDC-TAP returns results in XSAMS format. Dictionaries define the translation between the database entry format (not specified by VAMDC) and the XSAMS entry format. Database owners can choose to use a Java or a Python version of the XSAMS generator. Certain entries are classified as restrictables in the dictionary, thereby identifying them as potential

search terms. In VSS2 also certain entries can be classified as returnables to restrict the range of XSAMS content.

VSS1/2 serves as a universal query language across all VAMDC databases. Various data conversions are handled automatically, for example change of units, and different XSAMS views are possible including a conversion to plain tabular data. This was illustrated via a live demonstration on the Vienna Atomic Line Database (VALD). The element and wavelength range are specified (wavelength in any of several units); it can be refined by ionization stage or initial or final energy.

Several VAMDC use cases for VSS and XSAMS were discussed in more detail. A query to the central registry provides content statistics over all connected databases. A collection of test queries has been prepared to provide quality assurance: responses to these test queries are analyzed to test new database interfaces or new database content. XSAMS output is passed through the Spectroscopy Made Easy (SME) package to generate synthetic spectra for a Gaia space astrometry application. Finally it was shown how VSS and XSAMS can be used to match output from two databases -- Basecol for collisional rate coefficients and CDMS for Einstein coefficients -- and a separate table of experimental energies.

D.P. Stotler: Molecules in Fusion Plasma Diagnostics and Fuelling

Following a brief reminder of the tokamak concept and the magnetic divertor configuration two examples are discussed of the use of atomic and molecular data in the study of magnetically confined fusion plasma.

The first example involves the use of a fast camera and a hydrogen or helium gas puff to image edge plasma turbulence on the National Spherical Torus Experiment (NSTX) [S.J. Zweben, R.J. Maqueda, D.P. Stotler et al., Nucl. Fusion 44 (2004) 134]. Due to the strong magnetic field the turbulence is highly anisotropic with rapid variation across the field and long structures along the field. A sheet of neutral gas is injected perpendicular to the local field, with the camera view directed along the field. The perturbations in plasma density and temperature resulting from the turbulence lead to corresponding changes in the emission of line radiation from the gas and this allows the turbulence to be visualized. The repetition rate of the camera can be as high as 10^6 frames/s although more often a time resolution of about 10 μs is used.

Quantitative analysis of the gas puff imaging data relies on Monte Carlo modelling of the neutral gas in a given background plasma. The neutral gas particles, D_2 or He, undergo dissociation, charge exchange, excitation and ionization and the simulation provides a 3-D neutral gas profile and a synthetic camera image. In the work to-date this image is matched to a time-average of the movie data. The quality of the A+M data is important. For $e + D_2$ processes the data of [R.K. Janev, W.D. Langer, K. Evans Jr., D.E. Post, Jr., Elementary Processes in Hydrogen-Helium Plasmas, Springer-Verlag, 1987] have been used. Since 1987 there is increased interest in higher density and lower temperature divertor conditions in which vibrational excitation of molecules is important. Aspects of this are described in a combined atomic and molecular collisional-radiative model [U. Fantz, D. Reiter, B. Heger and D. Coster, J. Nucl. Mater. 290-293 (2001) 367]. Further effects to be considered are associated with the energy distribution of dissociation products in dependence on vibrational state of the molecule; this is asking for state-resolved fully differential cross-sections. The Monte Carlo code can handle it, but one needs the data.

The second example of use of A+M data in fusion plasma is provided by molecular cluster injection (MCI). This is proposed as a fuelling method intermediate between gas puff injection, for which the ionization region is close to the outer wall, and frozen pellet injection, for which the ionization region is in the core plasma. At Princeton Plasma Physics Laboratory MCI is being developed for use on the Lithium Tokamak Experiment (LTX). LTX is a small experiment and the particle inventory in the plasma is only about $3*10^{19}$, which implies the use of small pellets (diameter 0.6-0.8 mm) if frozen pellet injection is used. These small pellets are difficult to produce and to handle. MCI is being

explored on a test stand and the principal diagnostic is a measurement of the visible Balmer-alpha line emission. The key molecular process is dissociation by electron impact producing excited H atoms, $e + H_2 \rightarrow e + H(n=1) + H(n=3)$. The test stand and diagnostic system are modelled with use of Monte Carlo methods for the neutral gas as in the previous example, taking account of molecular dissociation, dissociative excitation and ionization and atomic processes. The modelling leads to the conclusion that MCI with clusters of about 1000 particles each can be effective for fuelling LTX.

A. Kramida: NIST Physical Reference Data

Following a recent reorganization the Physical Measurement Laboratory (PML) is one of the major operating units of the NIST and it is the home of precision measurement services, standards and data. One of the branches of the PML is the Quantum Measurement Division, which is responsible for development of databases on atomic spectroscopy. The most important product, representing many years of effort, is the Atomic Spectra Database, ASD, currently version 4. ASD provides critically evaluated data on atomic energy levels, wavelengths and transition probabilities. Data are included for observed transitions of 99 elements and energy levels of 89 elements. ASD contains data on more than 1000 spectra from about 0.4 Å to 500 μm, with about 92,500 energy levels and 180,000 lines, 73,400 of which have transition probabilities (Please see "NIST Atomic Spectroscopy Databases": http://www.nist.gov/pml/data/atomspec.cfm). The new interface for ASD 4 offers dynamic transition plots (Grotrian diagrams) with ability to zoom into particular energy levels and their transitions, and synthetic plasma emission spectra for Saha or for Local Thermodynamic Equilibrium (LTE) including lineshape effects.

Each critical compilation is a multi-year effort that starts with a bibliographical study. For this purpose NIST maintains several atomic spectra bibliographical databases: one on atomic transition probabilities, one on atomic spectral line broadening and shifts and one on atomic energy levels and spectra. These bibliographical databases are updated about once per month; an effort that starts with a prepared batch of scripted searches through ISI Web of Knowledge.

Each article that is part of a critical compilation must be carefully evaluated. The published wavelengths and intensities are reviewed and may be rejected; in any case adjustments may be made and uncertainty estimates are assigned. A complete line list is built up based upon all relevant publications, and this line list is analyzed via ab initio calculations, line spread function fitting, isoelectronic comparisons and series formulae. Then the levels are optimized and final uncertainties are assigned. The final compiled list contains levels, energies, uncertainties, fractional compositions, lines with their observed and Ritz wavelengths and their uncertainties, intensities and level populations, and atomic ionization potentials. The result is published and the data are entered into the online database.

C. Ballance: An Overview of Large Scale Collisional Calculations and Processing the Resulting Datasets

The work with T. G. Lee, S. D. Loch and M. S. Pindzola at Auburn and N. R. Badnell at Strathclyde is motivated by the Atomic Data and Analysis Structure (ADAS) project. ADAS requires comprehensive data in order to be useful for plasma modelling. The talk describes current capabilities and future directions for electron-impact processes including excitation, ionization and dielectronic recombination. These capabilities are based upon a massively parallel Dirac R-matrix code with calculations driven by scripts. Some problems of interest are high-n shell ionisation, excitation of Fe-peak elements and heavier, and dielectronic recombination for open f shells.

Data (for example for electron impact excitation) is stored in a well-prescribed format that includes the atomic configurations, the energy levels, the A-values for all E1, E2, M1 and M2 transitions, Maxwellian averaged collision strengths for a range of temperatures and the Born-Bethe infinite energy limit points. The Maxwellian averaging is done to remove detailed resonance information that

is of no interest in applications. The data go into ADAS and OpenADAS (please see http://www.adas.ac.uk/openadas.php) in the appropriate standard ADAS format.

For the calculations a massively parallel code is used, and steadily developed, that relies on the R-Matrix with Pseudostates (RMPS) method; the pseudostates represent the continuum above the ionization threshold. The part that is computationally most demanding is the calculation of ionization from excited states, which is often important in conditions relevant to fusion. When there are several valence electrons, for example in the case of boron-like systems, the matrices become very large and the mode of parallelization becomes critical. In the Auburn work an adaptive form of parallellization is employed that is described in more detail in the talk. At first this was implemented for semi-relativistic R-matrix calculation and applied to the F-, Na-, Ne- and Li-like sequences as described in a sequence of articles by Witthoeft, Liang and others. In order to handle larger atomic numbers a similar scripted parallel version of the Dirac R-matrix Code (DARC) was developed; see [M. A. Bautista, C. P. Ballance and P. Quinet, Astrophys. J. Lett. 718 (2010) L189–L193] for an application to Fe III.

One current project is the calculation of dielectronic recombination of W^{20+} and of Au^{20+}. Experimental studies by Schippers et al show many resonances up to 100 eV or more and it requires level-resolved DR to reproduce these resonances. This is a very large computational challenge that also involves major modifications to the AUTOSTRUCTURE code (work of N. R. Badnell).

V. Popova and S. Gagarin: Spectr-W3: Status and Perspectives

The Spectr-W^3 atomic database provides experimental, calculated, and compiled data on ionization potentials, energy levels, wavelengths, radiation transition probabilities and oscillator strengths for use in spectroscopy and other applications. With support from the International Science and Technology Center (ISTC) it has been qualitatively updated and extended in recent years. The interface was upgraded to allow direct upload of authors' data, software and equipment of the website were upgraded, functionality and content of the web pages were extended and a new version of the off-line (Spectr-CD) setup package was generated. In addition the scientific content of Spectr-W^3 has continued to grow and with more than 450000 records it is the largest database of its kind.

Currently the Spectr-W^3 atomic database (ADB) homepage has been included in the family of specialized atomic databases maintained at the Weizmann Institute (http://plasma-gate.weizmann.ac.il/DBfAPP.html) and has been integrated into the IAEA General Internet Search Engine for Atomic Data GENIE (www-amdis.iaea.org/GENIE). Spectr-W3 is freely accessible round-the-clock on the Web at http://spectr-w3.snz.ru/ and its web pages are regularly visited by users at research institutions and universities all over the world (about 30 visits per day). The web site of the Spectr-W^3 atomic database is hosted on a dedicated web server located at the Center for International Science and Technology Cooperation of RFNC VNIITF (http://www.vniitf.ru).

Access to Spectr-W^3 from the VAMDC portal through the AstroGrid registry has been implemented. A local version of the Spectr-W^3 database, Spectr-CD, is available from the Spectr-W^3 homepage for use on Windows PCs. The access through VAMDC returns data in XSAMS format for energy levels, ionization potentials and spectral lines; XSAMS output format is also available through SpectrCD. A parsing algorithm was developed to convert the notation for atomic level structure from XSAMS to the more human-readable notation as in $((2p^4\ ^3P)3d)^4P_{1/2}$.

Future Spectr-W^3 activities are concentrated on expansion of the database and of the XSAMS i/o capabilities. There is also interest to do a full scale export/import that involves parsing of detailed spectroscopic information.

Discussions

Atomic data issues

The discussions on atomic data issues in XSAMS were introduced and led by N. Piskunov. Following the Consultants' Meeting on XSAMS at IAEA in November 2010 minor changes (mainly bug-fixes) were made to XSAMS version 0.1 and the result was called version 0.1.1. VAMDC then made larger developments in two areas: the case-by-case approach for molecular spectroscopy (discussed by Ch. Hill in the present meeting) and the treatment of lineshapes and the plasma environment.

XSAMS is maintained under the SourceForge project management system at http://xsams.svn.sourceforge.net/. There are three principal tracks in the system starting from the common base that is IAEA XSAMS version 0.1. The trunk is the international version, now IAEA XSAMS 0.1.1. The two main branches are VAMDC XSAMS (successively versions 0.2 and 0.3) and NIST XSAMS. It is intended to merge the branches together into a common VAMDC 1.0 in the near future.

Within VAMDC XSAMS there are established procedures for bug reporting and for handling such reports through SourceForge. When logged in, the "Tracker" menu offers forms for submitting bug reports, feature requests and patches, as well as a mechanism to follow the developers' progress in resolving them. There are also discussion forums under SourceForge for developers and users. Similar procedures need to be put in place for IAEA XSAMS, especially as VAMDC will conclude at the end of 2012.

Major changes to VAMDC XSAMS are approved by the VAMDC Board. The past major changes (versions 0.2 and 0.3) and a possible final (within the lifetime of the present VAMDC) future change address issues of consistency and completeness of the schema. A summary of all changes can be found on the VAMDC web site at http://www.vamdc.org/documents/standards/. The largest changes in the atomic data part are the new sections on Environments, Broadening and Shiftings. In addition there are some smaller, but still significant changes:

AtomicNumericalData: new elements describe hyperfine structure caused by interaction with magnetic dipole and electric quadrupole nuclear moments.

InChI and InChIKey elements added for atomic ions; InChIKey is mandatory.

States element renamed into Species.

PartitionFunction element was added; it can be a table and can depend on the environment.

For the forbidden radiative transitions the probability can now have tags M1, E2, M2 etc.

Cross sections now include an element AbsorptionCrossSection.

The Environment section has optional elements Temperature, TotalPressure, TotalNumberDensity and Composition. Composition can in turn be described using elements PartialPressure, MoleFraction or Concentration.

Broadening consists of a named attribute identifying the process leading to the line width ("pressure", "doppler", "natural", etc.) and a reference to the environment causing the broadening. Within each Broadening element is a Lineshape element, identifying the lineshape by name ("lorentzian", "voigt", etc.) and, optionally, by reference to a FunctionType explicitly giving the functional form for the lineshape. Each Lineshape element contains one or more LineshapeParameters, named values which can be static (traditional DataTypes, e.g. a specific half-width for a Lorentzian line under the specified environment) or parameters to a named function (e.g. giving the half-width for a range of temperatures

or pressures through defined Function). Shifts are treated similarly to broadening. Finally satellite lines can be specified.

VAMDC must reflect the needs of the data publishers and data users. Data publishers must be able to use XSAMS to describe all the data that they intend to publish. They must be able to define a complete dictionary and there must be a mechanism and policies for adding new parameters to the dictionary and to the schema. The procedures for this need to be agreed beyond the VAMDC project in order to sustain XSAMS. Another need for data publishers is to obtain citations for use of their data. Although this is outside the direct control of VAMDC or XSAMS the design of the schema can facilitate it. Currently XSAMS provides means for including bibliographic information for various components. This must be complemented with a statistic collection tool for analyzing an XSAMS structure and extracting a unique list of data sources.

Data users require a consistency of the interface with the transferred/stored data; they must be able to trust that data access through XSAMS gives them the complete dataset. Data users require flexibility and configurability of query form and query options. Many users are experienced with the pre-XSAMS interface to the database and it is important to obtain their comments and suggestions in order not to lose the possibilities already in place. Finally, many users don't want to see the XSAMS output; they need to have it converted into tabular format compatible with their applications.

N. Piskunov concluded his presentation with a summary of work yet to be done. The treatment of lineshapes is not fully satisfactory yet; further work is needed on transitions and lineshape parameters. Another issue is to reconcile different linguistic traditions that cause confusion, for example between lower-upper states and initial-final states. Also words accuracy, precision, uncertainty and quality need to be more carefully defined.

The presentation by N. Piskunov of the changes in the atomic data part of XSAMS was met with much interest, requests for clarification and discussion. In the end only the choice to make the InChIKey mandatory was really controversial, because it forces changes in existing implementations. VAMDC found it essential to introduce this unique identifier to overcome different choices of specification among databases. For example, one database may use charge state and another may use spectral identifier (Fe 0 vs Fe I). In some communities it is common to specify the number of electrons instead of the charge state using language such as "Li-like iron" for Fe^{23+}.

The change from States to Species also affects existing implementations. SpeciesID now refers to atoms and ions as well as molecules. The SpeciesID does not include information about the electronic state.

Molecular data issues

Ch. Hill introduced and led the discussion on molecular data issues. The Schema is maintained on SourceForge under http://xsams.svn.sourceforge.net/; one finds there the trunk and branches for NIST, UCL, VAMDC and also a VAMDC-Working (development) branch. Ch. Hill recommends that the trunk should have a proper namespace at IAEA, certainly when XSAMS version 1.0 is adopted.

The most significant development in the molecular data part of XSAMS is the "case-by-case" approach to molecular spectroscopy. However, some smaller issues are briefly noted as well; see http://www.vamdc.org/documents/vamdc-xsams-changelog_v0.2.pdf for more detail.

Source references for an element have been moved from attributes to sub-elements.

The type DataType is extended to include fit-function parameters.

New vector and matrix data types have been introduced and extended with a ref element. These prove useful for describing, for example, atomic displacements in normal modes (a set of vectors) and the relaxation matrix which appears in the description of line-coupling in a set of radiative transitions.

New types DataList and LinearSequence have been introduced.

Species identification (InChIKey) has been added and made mandatory for atoms and molecules.

Molecular structure may now be described using tags imported from the Chemical Markup Language (CML).

The case-by-case approach to molecular symmetry and spectroscopy is described in detail in the VAMDC document http://www.vamdc.org/documents/cbc_v0.2.pdf. One general case and 14 special cases have been defined. They are: general (gen), diatomic closed shell (dcs), Hund's case (a) diatomics (hunda), Hund's case (b) diatomics (hundb), closed-shell linear triatomics (ltcs), closed-shell non-linear triatomics (nltcs), closed-shell symmetric top molecules (stcs), closed-shell linear polyatomics (lpcs), closed-shell asymmetric top molecules (asymcs), open-shell asymmetric top molecules (asymos), closed-shell spherical top molecules (sphcs), open-shell spherical top molecules (sphos), open-shell linear triatomics (ltos), open-shell linear polyatomics (lpos) and open-shell non-linear triatomics (nltos). XML validation of the quantum numbers for such cases is trivial if the constraint involves only a single item and non-trivial if more than one item is involved; for example the constraint $K_a + K_c \in \{J, J+1\}$ in case asymcs. Non-trivial validation is achieved through additional software that can, for a specific molecule, check state degeneracies against the quantum numbers and confirm the consistency of the symmetry and parity labels of a state.

Environmental parameters were already discussed in connection with atomic lineshapes; these parameters (envID and envRef and their components describing temperature, pressure and composition) are also relevant in the molecular data part. We recall the principal line broadening types: pressure, natural, Doppler and instrumental. In addition Zeeman splitting (line shift) should be included, but it is not yet properly treated.

Treatment of data accuracy in XSAMS has been developed further with the introduction of a new type AccuracyType. It allows to distinguish between absolute and relative errors and between systematic and statistical errors, and offers some further refinements. There remains the question, for the VAMDC developers, to what extent data attributes such as accuracy and quality belong in XSAMS; perhaps they belong rather in the original papers.

There is a new type NormalModes for numerical description of molecular excitations; it can describe symmetry, frequency, intensity and displacement vectors.

Management issues

The discussion on management issues of XSAMS was led by B. Braams and Yu. Ralchenko. A key consideration is the relation between the international XSAMS project that was initiated in 2003 and coordinated by the IAEA A+M Data Unit and the VAMDC project that started in 2009 and that has adopted XSAMS as the language for data exchange among the participating databases. In practice, at the present time all active implementations of XSAMS are done for databases that are associated with VAMDC and VAMDC has its own development path for XSAMS. There is no desire for an independent development path in "IAEA XSAMS" and therefore the IAEA meetings such as the present one serve to some extent as a check on the developments that are chosen and tested in VAMDC.

Topics for our meeting are to review and adopt changes to the international standard XSAMS and also to develop new areas not covered by VAMDC such as data for solid state.

The meeting discussed the adoption of the international standard XSAMS version 1.0. Recall that version 0.1 was adopted in September 2009 and minor correction were implemented immediately after the meeting of November 2010 to create version 0.1.1. Starting from version 0.1/0.1.1 the VAMDC XSAMS branch developed through versions 0.2 and 0.3 to include environmental and lineshape parameters, the case-by-case approach to molecular spectroscopy, and many smaller changes. At this time VAMDC XSAMS version 0.3 is the standard version that is being implemented on all databases associated with VAMDC.

In the run-up to the present meeting it was expected that the meeting would agree to adopt the present VAMDC XSAMS as the international standard version 1.0. However, this plan was finally not followed. The meeting agreed that with so many implementations still under way, and with significant user feedback only expected starting nearer the end of 2011, it would be advisable to defer agreement on international XSAMS version 1.0 until the time of the next VAMDC annual meeting, 21-24 February 2012 at the University of Vienna.

At the same time the meeting strongly endorsed the implementation work based on VAMDC XSAMS version 0.3 and recommends this version also for any new parties that choose to implement XSAMS. There are no major issues with this version and it is likely that the proposal for international XSAMS version 1.0 in February 2012 will look very much like VAMDC XSAMS version 0.3 now. In essence, the decision to adopt an XSAMS version 1.0 is a statement that the project is now out of the testing phase. We find it appropriate to have the benefit of the user feedback that is expected to be available by Feb 2012 before this decision is made.

The topic of XSAMS development for plasma-material interaction data was briefly discussed. Such data can be described in XSAMS, but with very much less resolution of detailed physics than is available for atomic and molecular data. The meeting agreed that that is adequate for the present time. We want to see the implementations and the user acceptance of XSAMS for atomic and molecular data before taking up further development of the representation of plasma-material interaction data.

Another candidate topic for XSAMS development is that of solid state spectroscopy, which at this time is not represented in XSAMS at all. Active work is being carried out by B. Schmitt and colleagues at Grenoble on an XML-based Solid Spectroscopy Data Model (SSDM). This is in connection with the Grenoble Astrophysics and Planetology Solid Spectroscopy and Thermodynamics database service, also known as GhoSST [http://ghosst.obs.ujf-grenoble.fr/]. We should review if this work can be part of XSAMS.

User interface issues

The discussion on the XSAMS user interface was introduced and led by N. Piskunov and Ch. Hill. The recent effort has been on the specification of the query language VAMDC SQL Subsets 1 and 2 (VSS1 and VSS2): http://www.vamdc.org/documents/querylanguage_v11.05.pdf and http://vamdc-standards.readthedocs.org/en/latest/queryLanguage/vss2.html.

VSS1 and VSS2 are based on SQL92, but only allow SELECT statement queries, not modifications of the database. At present most VAMDC databases have implemented VSS1, but VSS2 implementations are underway and it is intended that VSS2 will replace VSS1; VSS2 is a superset of VSS1. Queries are submitted to a VAMDC database web-service that, using its dictionary, converts the query into an SQL query or set of queries.

A few examples were shown: searches of the HITRAN and Cologne databases for spectroscopy of water and of CO. Most of the VAMDC databases are just now starting to make the XSAMS output option available to users, and we expect user feedback now through the VAMDC annual meeting in Feb 2012.

Technical issues with VAMDC XSAMS 0.3

A discussion on technical issues with the current VAMDC XSAMS branch was introduced and led by Yu. Ralchenko, mainly to raise issues for the attention of VAMDC. In the following, references are to the VAMDC-XSAMS reference guide version 11.12: http://www.vamdc.org/documents/vamdc-xsams-guide_v11.12.pdf.

With respect to the documentation is is requested to have a more careful attribution to the earlier international XSAMS documents.

In section 3.2, XSAMSData, there is an anomaly that all elements are optional; shouldn't there be at least one mandatory element?

In subsection 4.3.1, DataSeriesType, there occurs a DataFile optional element. It is never used, but its purpose is to describe data that are contained in some external location rather than being part of the database and transmitted through XSAMS. It is questioned if such an element should be introduced before being used. More substantially it is questioned if this element should ever be used: XML is meant to transmit validated data and not links to data.

In section 15, Processes.Radiative, it is recommended to use the more specific expression "absorption cross section" rather than just cross section. (This was done in version 0.3.)

In section 16.1, NonRadiativeTransition, autoionization and predissociation are mentioned as examples, but it is not clear what is the role of these terms in the schema. We don't know of any database that has information about predissociation states and rate coefficients. We may view this section as a placeholder and return to the issue when an implementation requires it.

In section 21.3, lineshapes dictionary, the list of lineshapes names and parameters appears to be incomplete; it contains only Lorentzian (parameter gammaL) and Voigt (parameters gamma and sigma).

4. Conclusion

The meeting participants are very pleased with the success of VAMDC to implement XSAMS on their approximately 25 databases. In the course of these implementations the VAMDC team made several choices that extend IAEA XSAMS version 0.1.1 to VAMDC XSAMS version 0.3. The most important choices concern the treatment of the environment, lineshapes and molecular spectroscopy. The participants agree in general terms with the choices made by VAMDC and endorse VAMDC XSAMS version 0.3 for the present implementations.

At the present time the highest priority is to bring the implementation of XSAMS for atomic and molecular data out from the testing stage, as is to be certified by adoption of an international standard XSAMS version 1.0. Developments in XSAMS for plasma-material interaction and for solid state should follow later.

The next XSAMS technical meeting and meeting of the Steering Committee is planned to be held during the week 20-24 Feb 2012 in conjunction with the VAMDC annual meeting in Vienna. We intend to define precisely and adopt XSAMS-1.0 at that meeting in February. We will advertise to our data centre partners (e.g., NFRI, NIFS, CRAAMD) the VAMDC tutorials for XSAMS implementors and users. After that we expect to meet again during the week 30 Sep - 04 Oct 2012 in conjunction with ICAMDATA.

Consultants' Meeting on "XML Schema for Atoms, Molecules and Solids (XSAMS)"

3−5 October 2011, NIST, Gaithersburg MD, USA

Scientific Secretary: Bas Braams
Local Organizer: Yuri Ralchenko

List of Participants

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 Valerya Popova
Russian Federal Nuclear Center
All-Russian Institute of Technical Physics
P.O. Box 245
13, Vasilyeva st., Snezhinsk
Chelyabinsk Region, 456770
RUSSIAN FEDERATION
Tel.: +7-351-46-54730; Fax: +7-351-46-55118
E-mail: daka_popova@mail.ru

Prof Nikolai Piskunov
Department of Physics and Astronomy
Uppsala University
Box 515
S-75120 Uppsala
SWEDEN
Tel.: +46-18-471-5958; Fax: +46-18-471-5999
E-mail: piskunov@fysast.uu.se

Dr. Christian Hill
Department of Physics & Astronomy
University College London
Gower Street
London, WC1E 6BT
UNITED KINGDOM
Tel.: +44-207-679-3478; Fax: +44-207-679-7145
E-mail: christian.hill@ucl.ac.uk

Dr Connor Ballance
Auburn University
Department of Physics
206 Allison Labs
Auburn, AL 36849
UNITED STATES OF AMERICA
Tel.: +1-334-844-8712; Fax: +1-334-844-4613
E-mail: ballance@physics.auburn.edu

Dr Alexander E. Kramida
Atomic Physics Division
100 Bureau Dr., Stop 8422
National Institute of Standards and
 Technology
Gaithersburg MD 20899-8422
UNITED STATES OF AMERICA
Tel.: +1-301-975-8074; Fax: +1-301-975-3038
E-mail: alexander.kramida@nist.gov

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

Dr Daren P. Stotler
Princeton University
Plasma Physics Laboratory
MS 27
P.O. Box 451
Princeton, NJ 08543
UNITED STATES OF AMERICA
Tel.: +1-609-243-2063; Fax: +1-609-243-2662
E-mail: dstotler@pppl.gov

IAEA

Dr Bastiaan J. Braams
IAEA Atomic and Molecular Data Unit
Vienna International Centre
P.O. Box 100
A-1400 Vienna
AUSTRIA
Tel.: +43-1-2600-21731; Fax: +43-1-26007
E-mail: bastiaan.j.braams@iaea.org

Consultants' Meeting on "XML Schema for Atoms, Molecules and Solids (XSAMS)"

3−5 October 2011, NIST, Gaithersburg MD, USA

Scientific Secretary: Bas Braams
Local Organizer: Yuri Ralchenko

Agenda

Monday, 3 October (09:00 – 17:00h)

09:30   Yu. Ralchenko and B. Braams: Opening and introductions

09:40   C.J. Williams (NIST): Welcome

10:00   Yu. Ralchenko: Brief introduction to XSAMS

10:15   *Coffee*

10:30   N. Piskunov: The VAMDC project

11:30   D. Stotler: Molecules in fusion plasma diagnostics and modelling

12:00   *Lunch*

13:30   Visit to the EBIT

14:15   *Coffee*

14:30   A. Kramida: The NIST atomic spectra database

15:00   C. Ballance: Large scale collisional calculations and processing for the ADAS project

15:30   S. Gagarin/V. Popova: Implementation and development of XSAMS tools in Spectr-W3

16:00   N. Piskunov: Survey of atomic data issues in VAMDC XSAMS

16:30   C. Hill: Survey of molecular data issues in VAMDC XSAMS

19:00   *Social dinner*

Tuesday 4 October (09:00 – 17:00h)

A. All: Review and discussion of all things atomic in XSAMS ... spectroscopy, structure, electron-atom collisions, environment, lineshapes.

B. All: Review and discussion of all things molecular in XSAMS ... spectroscopy, electron-molecule collisions, heavy particle collisions.

C. All: Review and discussion about user interface issues.

D. Steering Committee + All: "Business meeting". Possible adoption of VAMDC XSAMS as XSAMS-1.0.

Wednesday 5 October (09:00 – 14:00h)

All: Plans for XSAMS beyond version 1.0. Large molecules and particle-surface interaction issues.

14:00   *Close of meeting*