



Contents lists available at ScienceDirect

Journal of Quantitative Spectroscopy & Radiative Transfer

journal homepage: www.elsevier.com/locate/jqsrt

Virtual atomic and molecular data centre

M.L. Dubernet^{a,b,*}, V. Boudon^c, J.L. Culhane^d, M.S. Dimitrijevic^e, A.Z. Fazliev^f, C. Joblin^g, F. Kupka^h, G. Letoⁱ, P. Le Sidaner^j, P.A. Loboda^k, H.E. Mason^l, N.J. Mason^m, C. Mendozaⁿ, G. Mulas^o, T.J. Millar^p, L.A. Nuñez^q, V.I. Perevalov^r, N. Piskunov^r, Y. Ralchenko^s, G. Rixon^{bb}, L.S. Rothman^t, E. Roueff^b, T.A. Ryabchikova^u, A. Ryabtsev^v, S. Sahal-Bréchet^w, B. Schmitt^x, S. Schlemmer^y, J. Tennyson^z, V.G. Tyuterev^{aa}, N.A. Walton^{bb}, V. Wakelam^{cc}, C.J. Zeippen^w

^a Laboratoire de Physique Moléculaire pour l'Atmosphère et l'Astrophysique, UMR7092 CNRS/INP, Université Pierre et Marie Curie, Case 76, 4 Place Jussieu, 75252 Paris Cedex 05, France

^b Laboratoire Univers et Théories, UMR8102 CNRS/INSU, Observatoire de Paris, Section Meudon, 5 Place Janssen, 92195 Meudon Cedex, France

^c Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 5209 CNRS-Université de Bourgogne, 9 Avenue Alain Savary, BP 47 870, F-21078 DIJON Cedex, France

^d Mullard Space Science Laboratory, University College London, Holmbury St. Mary, Dorking, Surrey RH5 6NT, UK

^e Astronomical Observatory, Volgina 7, 11060 Belgrade, Serbia

^f V.E. Zuev Institute of Atmospheric Optics, Siberian Branch, Russian Academy of Sciences, 1, Academician Zuev square, Tomsk 634021, Russian Federation

^g Centre d'Etude Spatiale des Rayonnements, UMR5187 CNRS/INSU, Université Paul Sabatier, 9 Avenue du Colonel Roche, F-31028 Toulouse cedex 9, France

^h Faculty of Mathematics, University of Vienna, Nordbergstrasse 15, 1090 Wien, Austria

ⁱ Istituto Nazionale di Astrofisica—Osservatorio Astrofisico di Catania, Via Santa Sofia 78, I-95123 Catania, Italy

^j Division Informatique de l'Observatoire, VO-Paris Data Centre, UMS2201 CNRS/INSU, Observatoire de Paris, 5 Place Janssen, 92195 Meudon Cedex, France

^k Russian Federal Nuclear Centre-All Russian Institute of Technical Physics (RFNTC-VNIITF), 13 Vasilyeva St., Snezhinsk, Chelyabinsk Region 456770, Russia

^l Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK

^m Open University, Faculty of Science, Walton Hall, Milton Keynes MK7 6AA, UK

ⁿ Centro de Física, Instituto Venezolano de Investigaciones Científicas (IVIC), P.O. Box 20632, Caracas 1020A, Venezuela, and Centro Nacional de Cálculo Científico Universidad de Los Andes (CeCalCULA), Corporación Parque Tecnológico de Mérida, Mérida 5101, Venezuela

^o Istituto Nazionale di Astrofisica—Osservatorio Astronomico di Cagliari, strada 54 loc. Poggio dei Pini, Capoterra (CA) I-09012, Italy

^p School of Mathematics and Physics, Queen's University Belfast, University Road, Belfast BT7 1NN, UK

^q Centro Nacional de Cálculo Científico Universidad de Los Andes (CeCalCULA), Corporación Parque Tecnológico de Mérida, Mérida 5101, Venezuela, and Grupo de Investigación en Relatividad y Gravitación (GIRG) Esc. Física, Universidad Industrial de Santander, Bucaramanga, Colombia

^r Uppsala University, Department of Physics and Astronomy, Lägerhyddsvägen 1, Uppsala 75120, Sweden

^s National Institute of Standards and Technology, Atomic Physics Division, 100 Bureau Dr., Stop 8422, Gaithersburg, MD 20899-8422, USA

^t Harvard-Smithsonian Center for Astrophysics, Atomic and Molecular Physics Division, MS 50, 60 Garden Street, Cambridge, MA 02138-1516, USA

^u Institute for Astronomy RAS, Pyatnitskaya 48, Moscow 119017, Russian Federation

^v Institute for Spectroscopy RAS, Physical 5, Troitsk 142190, Russian Federation

^w Laboratoire d'Etude du Rayonnement et de la Matière en Astrophysique, UMR8112 CNRS/INSU, Observatoire de Paris, 61, Avenue de l'Observatoire, 75014 Paris, France

^x Laboratoire de Planétologie de Grenoble, UMR5109 CNRS/INSU, Université Joseph Fourier, BP53, 38041 Grenoble Cedex 9, France

^y University of Cologne, I. Physikalisches Institut, Zùlpicher Strasse 77, Köln 50937, Germany

^z Department of Physics and Astronomy, University College London, London WC1E 6BT, UK

^{aa} Groupe de Spectroscopie Moléculaire et Atmosphérique, UMR6089 CNRS/INP, Université de Reims, U.F.R. Sciences Exactes et Naturelles, Moulin de la Housse B.P. 1039, 51687 Reims Cedex 2, France

^{bb} Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB30HA, UK

^{cc} Laboratoire d'Astrophysique de Bordeaux, UMR5804 CNRS/INSU, Université de Bordeaux, BP89, 33271 Floirac Cedex, France

ARTICLE INFO

ABSTRACT

The Virtual Atomic and Molecular Data Centre (VAMDC, <http://www.vamdc.eu>) is a European Union funded collaboration between groups involved in the generation,

* Corresponding author at: Laboratoire de Physique Moléculaire pour l'Atmosphère et l'Astrophysique, UMR7092 CNRS/INP, Université Pierre et Marie Curie, Case 76, 4 Place Jussieu, 75252 Paris Cedex 05, France. Tel.: +33 1 44 27 70 34; fax: +33 1 44 27 70 33.

E-mail address: Marie-Lise.Dubernet-Tuckey@upmc.fr (M.L. Dubernet).

Keywords:

Atomic data
Molecular data
Astrophysics
Atmospheric physics
Fusion
Radiation
Lighting
Space physics
Planetology

evaluation, and use of atomic and molecular data. VAMDC aims to build a secure, documented, flexible and interoperable e-science environment-based interface to existing atomic and molecular data. The project will cover establishing the core consortium, the development and deployment of the infrastructure and the development of interfaces to the existing atomic and molecular databases. It will also provide a forum for training potential users and dissemination of expertise worldwide. This review describes the scope of the VAMDC project; it provides a survey of the atomic and molecular data sets that will be included plus a discussion of how they will be integrated. Some applications of these data are also discussed.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Atomic and molecular (A&M) data are of critical importance across a wide range of applications such as astrophysics, atmospheric physics, fusion, environmental sciences, combustion chemistry and in industrial applications from plasmas to lighting. Currently these vital and fundamental A&M data resources are highly fragmented and only available through a variety of highly specialized and often poorly documented interfaces, thus limiting the full exploitation of their scientific worth. This in turn hinders research across a wide range of topics including space exploration (the characterization of extra-solar planets [1], understanding the chemistry of our local solar system and of the wider universe); the study of the terrestrial atmosphere and quantification of climate change; the development of the international fusion programme for energy, and our understanding of radiation damage within biological systems, to give just a few examples.

The Virtual Atomic and Molecular Data Centre (VAMDC) aims to build a unified, secure, documented, flexible and interoperable e-science environment-based interface to existing A&M data. VAMDC will combine the expertise of existing A&M databases, data producers and service providers with the specific aim of creating an infrastructure that is easily tuned to the requirements of a wide variety of users in academic, governmental, industrial or public communities. The project will cover the construction of the core consortium, the development and deployment of the infrastructure and the development of interfaces to existing A&M databases.

VAMDC brings together a wide range of partners represented by the authors of this paper. These partners have expertise in both the provision and use of A&M data and in the technical development and use of key e-infrastructures (e.g. the Euro-VO, <http://www.euro-vo.org> and EGEE, Enabling Grids for E-science, <http://www.eu-egee.org/>). VAMDC is organised following inclusive and open principles. It will interact widely with the general scientific community to ensure that the priorities for deployment of A&M services are well founded (selecting its data from validated, broadly used and acknowledged data sets), that training programmes are widely available, and that the benefits of access to this information infrastructure is fully accessible to the wider scientific community. Internally the delivery of this scientific data infrastructure is divided into service

activities focussed on deployment and support, and research activities related to data content, transport, presentation. These efforts will be aimed at designing the necessary technical extensions to the baseline infrastructure which are required to meet the demands of the A&M user communities. VAMDC will interact with other underlying infrastructures – such as EGEE and the Euro-VO – and also participate in relevant standardisation activities such as the International Virtual Observatory Alliance (IVOA [2]), to ensure continued interoperability of VAMDC resources with related e-infrastructures.

2. Objectives

There are many groups worldwide who generate atomic and molecular data by measurement and/or calculation. Such data feeds several databases and information services in what is currently a very fragmented fashion. However, in some scientific areas, A&M data producers have already established large, trans-national networks, such as the European Theoretical Spectroscopy Facility (ETSF, <http://www.etsf.eu>), a project developed within the Nanoquanta Network of Excellence (<http://www.nanoquanta.eu>). Several major scientific collaborations plan to use such services for their own objectives and needs. Examples include the planetary science community through Europlanet (<http://www.europlanet-ri.eu>), the astronomy community through Euro-VO, the fusion community through ITER (<http://www.iter.org>) and EURATOM (<http://www.euratom.org>), and the radiation science community for modelling radiotherapy and effect of low doses on human health [3]. However within these initiatives there is little activity to ensure the interoperability of A&M databases. This means at present that every time the same A&M database is used for a new application, the output of the database has to be cumbersome adapted. For example automatic tools for the visualisation of simulations of planetary, stellar or the interstellar medium spectra are planned. These tools will require automatic access to different A&M databases, cross-matching the retrieved data as well as checking the quality of data. Yet to date there is no coherent nor sufficiently general infrastructure to perform such tasks.

The VAMDC is building such a secure, documented, flexible, easily accessible and interoperable e-infrastructure for A&M data based upon existing A&M databases. It has the specific aim of creating an infrastructure that on the one hand can directly extract data from the existing

depositories, while on the other hand be sufficiently flexible to be tuned to the needs of a wide variety of users from academic, governmental, industrial communities or by the general public.

The VAMDC can be understood as a publisher infrastructure, see Fig. 1. The VAMDC will deploy yellow pages (registries) in order to find resources, design user applications in order to meet the user needs, build data access layers above databases to provide unified outputs from these databases, care about asynchronous queries with workflows and the storage of large quantity of data with VOspace, and connect its infrastructure to the grid.

Central to the VAMDC is the task of overcoming the current fragmentation of the A&M database community. The VAMDC will alleviate this by:

- developing the largest and most comprehensive atomic and molecular e-infrastructure to be shared, fed and expanded by A&M scientists.
- providing a major distributed infrastructure which can be accessed, referenced and exploited by the wider research community.

In fulfilling these aims, the VAMDC project will organise a series of networking activities (NAs) laying the foundations for a long-lasting and self-sustaining infrastructure. NAs are specifically aimed at

- Engaging data providers.
- Coordinating activities among existing database providers.
- Ascertaining and responding to the needs of different user communities.

- Providing training and awareness of the VAMDC across the international A&M community and other use communities such as planetary scientists.

3. Databases

The core of the VAMDC e-infrastructure is the databases upon which it is based. These databases include those detailed in the following subsections.

3.1. Vienna atomic line database (VALD)

The VALD database [4] of atomic data provides a robust and consistent analysis of radiation from astrophysical objects. Created in 1995 in Vienna, it is developed and maintained by researchers at 7 European institutes. The main nodes are in Vienna, Uppsala and Moscow, where the three database servers (mirror sites) are located. VALD provides an e-mail and web-based user interface to a vast collection of spectral line parameters (central wavelengths, energy levels, statistical weights, transition probabilities, line broadening parameters) for all chemical elements of astronomical importance (over 50 million atomic transitions). Currently, VALD serves more than 1200 users from 47 countries and 5 international organisations. VALD processes nearly 100 requests per day.

3.2. CHIANTI

CHIANTI [5] is an atomic database for spectroscopic diagnostics of astrophysical plasmas. First released in 1997, it is a well-established atomic database for ions of astrophysical importance. Combined with interactive data

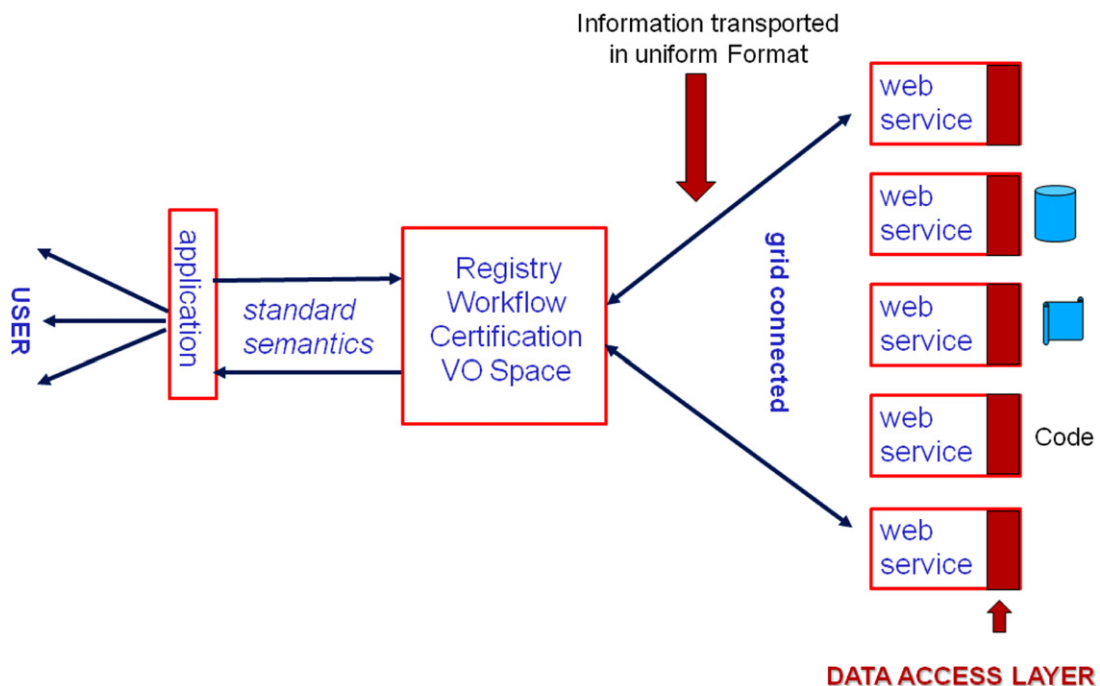


Fig. 1. Schematic diagram of the VAMDC infrastructure; note that it is a distributed system.

language (IDL) spectroscopic diagnostic programs, it is used in the analysis of optically thin collisionally ionised plasmas and is the preferred reference database in solar physics. The CHIANTI package also contains atomic structure data (experimental and calculated wavelengths and radiative data), and rates for electron and proton collisions. Version 6 of CHIANTI was released in 2008 [5]; it contains new rates for ionization and recombination, thus allowing the treatment of non-equilibrium plasmas. The CHIANTI data and programs are distributed via SolarSoft [6] to the solar community, but are also available on the web (www.damtp.cam.ac.uk/user/astro/chianti/) as a large number of ASCII files. To make them more accessible, some of the basic atomic data (line emissivities) were imported into Astrogrid's [7] VOTADA protocol in a MySQL database accessible from the Astrogrid Workbench.

3.3. EMol

The EMol database [8], under development at the Open University, aims to provide a comprehensive listing of measured and calculated cross sections for electron interactions with molecular systems. Combining several previously assembled databases and data reviews (e.g. Ref. [9]) in an on-line open-access database, EMol provides recommended cross-section values for data users in the plasma industry, and for the disciplines of discharge physics, fusion, aeronomy and radiation chemistry. The data are reviewed by an expert international panel and are regularly updated. Future data needs are considered [10]. Access is also provided to a suite of semi-empirical theoretical methods so that cross sections may be evaluated for targets for which there are currently no experimental data.

3.4. Cologne database for molecular spectroscopy (CDMS) [11] and Jet Propulsion Laboratory Submillimeter Catalog [12]

The CDMS (<http://www.ph1.uni-koeln.de/vorhersagen/>) provides recommended values of spectroscopic transition frequencies and intensities (including error estimates) for atoms and molecules of interest to the astronomical community and for studying the terrestrial atmosphere in the frequency range 0–10 THz, i.e. 0–340 cm^{-1} . It is a regularly updated database and a very thorough effort is carried out to give the maximum information on the fitting procedures, the quality of the fit, and the references to the original data. The CDMS is cross correlated with its US counterpart, the JPL submillimeter catalogue (<http://spec.jpl.nasa.gov/>), VAMDC will provide a vehicle for explicitly cross-linking these two activities.

3.5. BASECOL

The BASECOL database [13,14] (<http://basecol.obspm.fr>) provides excitation rate coefficients for ro-vibrational excitation of molecules by electrons, He and H_2 including error estimates. The BASECOL database can be automati-

cally accessed by software through a web service and data can be downloaded in several different formats such as HTML, ASCII or the VO's VOTable format. To date, BASECOL has been mainly used for the study of interstellar, circumstellar and cometary atmospheres.

3.6. GhoSST

The GhoSST (Grenoble astrophysics and planetology solid spectroscopy and thermodynamics, <http://ghosst.obs.ujf-grenoble.fr>) [15] database service, under development in Grenoble, provides spectroscopic laboratory data on molecular and atomic solids and liquids from the near UV to the far-infrared. These atomic and molecular data are contained in a sub-database that includes a spectroscopic band list (band central wavelength, band width, peak and integrated intensities, vibration mode, quantum numbers, isotopic species, reference, accuracies) for molecular and atomic solids and liquids in different condensed phases (amorphous and crystalline solid ice phases and in the liquid phase) and at several temperatures. Around 2000 bands from more than 30 different molecular solids in about 50 different phases will be included in the first phase of the database. Another sub-database will contain the physical properties of atomic and molecular solids (crystallography, thermodynamic properties, spectroscopic band list). Such a compilation is already completed for the vapour pressure of 53 molecular solids with almost 2000 data points tabulated at temperatures (T) from 15 to 280 K [16]. Access to these databases is by a web-based interface. Visualization of bands (intensity versus position) will be provided as well as spectral and intensity unit transformation. Downloads in different standard formats will be provided.

3.7. University of Manchester Institute of Science and Technology (UMIST) database for astrochemistry

The UMIST database for astrochemistry [17] (<http://www.udfa.net/>), which was created by Millar et al. [18] in 1991, provides a fundamental set of reaction rate data and related software for use in chemical kinetic modelling of astronomical regions. It was last updated in 2006.

3.8. Kinetic database for astrochemistry (KIDA)

KIDA is a database that will contain all chemical reactions used in the modelling of the chemistry in the interstellar medium and in planetary atmospheres. KIDA is in a consultation phase; a preliminary version was released in June 2009 (<http://kida.obs.u-bordeaux1.fr>) and a more complete version is due in 2010. The data in KIDA will be regularly updated by the users themselves, under the control of a group of experts, and will be regularly evaluated by its steering committee.

3.9. Polycyclic aromatic hydrocarbon spectral database

The CESR (Centre d'Etude Spatiale des Rayonnements/CNRS)-Cagliari spectral database of polycyclic aromatic hydrocarbons (PAHs) and carbon clusters (<http://astrochemistry.ca.astro.it/database/>) makes available a number of properties for a sample of presently about 60 species in four charge states: anion, neutral, cation and dication [19]. The properties were determined in a homogeneous way using state-of-the-art quantum-chemistry techniques, and include: (i) general energetics such as electron affinity and ionisation energies, static polarizability, permanent dipole moment, van der Waals coefficients, symmetry, multiplicity and optimised geometry of the ground electronic state; (ii) harmonic vibrational analyses, i.e. normal modes, their frequencies and IR activities; (iii) vertical electronic photoabsorption cross-sections and complex frequency-dependent electronic polarisabilities in the linear regime.

3.10. LASP database

The LASP (Laboratorio di Astrofisica Sperimentale) database (<http://web.ct.astro.it/weblab/dbindex.html#dbindex>) is under development at the INAF—Catania Astrophysical Observatory. It provides (i) infrared (IR) spectra of molecules in the solid phase ($T=10\text{--}100\text{ K}$) for both pure species and their mixtures before and after processing with energetic ions (30–200 keV) and UV photons (10.2 eV) [20,21,22]; (ii) IR optical constants of molecules in the solid phase as well as optical constants of frozen molecules after processing with energetic ions (30–200 keV) [23,24]; (iii) band strengths of the most relevant IR absorption bands [25,26]; and (iv) density values of frozen samples [26,27]. This database is particularly useful for detailed studies of the profile (shape, width and peak position) of absorption bands of solid phase molecules observed in astronomical spectra recorded in the direction of interstellar molecular clouds, as well as estimating the column density of the observed species. The optical constants will be useful in those instances for which a direct comparison between laboratory and astronomical spectra is not possible.

3.11. Stark-B

The Stark-B (<http://stark-b.obspm.fr>) [28] database is a collaborative project between the Astronomical Observatory of Belgrade and the Laboratoire d'Etude du Rayonnement et de la Matière en Astrophysique. This is a database of calculated widths and shifts of isolated lines of atoms and ions due to electron and ion collisions. This database is devoted to modelling and spectroscopic diagnostics of stellar atmospheres and envelopes. In addition, it is also relevant to laboratory plasmas, laser equipment and technological plasmas. Hence the range of temperatures and densities covered by the tables is wide and depends on the ionization degree of the considered ion. The temperature can vary from several thousands Kelvin for neutral atoms to several hundred thousands of Kelvin for highly charged ions.

The electron or ion density can vary from 10^{12} (in the case of stellar atmospheres) to several 10^{19} cm^{-3} (in some white dwarfs and some laboratory plasmas).

3.12. Spectr- W^3

The Spectr- W^3 project [29] is a collaboration between the Russian Federal Nuclear Centre All-Russian Institute of Technical Physics (RFNC VNIITF) and the Institute for High Energy Densities of the Joint Institute for High Temperatures of the Russian Academy of Sciences (IHED JIHT RAS). To date, the Spectr- W^3 atomic database contains about 450,000 records and includes experimental, calculated and compiled data on ionization potentials, energy levels, wavelengths, radiation transition probabilities and oscillator strengths, and also parameters for analytic approximations for electron-collision cross-sections and rates for atoms and ions. These data were extracted from publications or provided directly by the authors. All the information is supplied with references to the original sources and comments, elucidating the details of experimental measurements or calculations.

The Spectr- W^3 information-reference system is implemented as an on-line web resource providing free access to the Spectr- W^3 atomic database and the relevant additional information. The Spectr- W^3 site (<http://spectr-w3.snz.ru>) has been operating on the web since 2002. The Spectr- W^3 web-resource averages more than 50 visiting sessions per day. A number of experts also use the fully functional local version of the Spectr- W^3 database, Spectr-CD, created for the off-line operation on Windows PCs. The Spectr-CD setup package is available on the Spectr- W^3 homepage.

3.13. Databases maintained at IAO

The V.E. Zuev Institute of Atmospheric Optics (IAO) in Tomsk (<http://www.iao.ru/>) hosts a number of molecular databases accessible via either the internet or ftp including the following:

- The carbon dioxide spectroscopic databank (CDSDB) [30] (<http://cdsd.iao.ru> and <ftp://ftp.iao.ru/pub/CDSDB-2008>) which contains calculated spectral line parameters for seven isotopologues of carbon dioxide in the 5–12,784 cm^{-1} wavenumber range.
- The spectroscopy & molecular properties of ozone (S&MPO) relational database [31], see <http://ozone.iao.ru> (<http://ozone.univ-reims.fr/>), is a user-friendly information system, developed in collaboration with the University of Reims. It contains a well-established and unique database of spectral line parameters for the ozone molecule, experimental UV cross-sections, information on ozone's molecular properties, updated reference lists classified by type as well as programs and extended facilities for user applications involving original FTS experimental data in the IR range.
- "Spectroscopy of atmospheric gases" (<http://spectra.iao.ru>): is a compilation that contains the well-known databases such as HITRAN [32], GEISA [33] and HITEMP

[34]. Both “spectroscopy of atmospheric gases” and S&MPO have the programs for simulation of synthetic spectra from microwave to visible wavelengths.

- (d) Water internet @ccessible distributed information system (W@DIS) (<http://wadis.saga.iao.ru>) is a system based on data collected by the International Union of Pure and Applied Chemistry (IUPAC) Water-Vapor Task Group [35] which assimilates experimental water-vapour spectroscopy data from the literature and calculated line lists. W@DIS contains energy levels, transition positions and line intensities, and line profile characteristics. This project aims to include both the best theoretical as well as experimental data in the gas phase that have been generated for water (including all its important isotopologues).

3.14. Databases under the management of IVIC/CeCalCULA

These include:

- (a) TIPTOPbase [36], located at the Centre de Données astronomiques de Strasbourg, France (<http://cdsweb.u-strasbg.fr/topbase/home.html>), contains the atomic data computed in the Opacity Project and Iron Project, namely energy levels, *gf*-values, *A*-values, photoionization cross sections and electron impact cross sections and rates for light chemical elements with atomic number, $Z=1-28$.
- (b) OPserver [37], located at the Ohio Supercomputer Center, USA, (<http://opacities.osc.edu/>), is a remote, interactive server for the computation of mean opacities for stellar modelling using the monochromatic opacities computed by the Opacity Project. It can be accessed through a web portal or from a stellar modelling code, in the latter case by downloading locally a linkable subroutine library. (c) xstarDB [38] is the atomic database used by the XSTAR code for modelling photoionised plasmas.

3.15. High-resolution TRANsmission molecular absorption database (HITRAN) [32]

The atmospheric and planetology research communities make extensive use of the HITRAN [32] and GEISA [33] databases. The widely used HITRAN database lists individual line parameters for molecules in the gas phase from the microwave through to the UV spectral region as well as photoabsorption cross-sections for many molecules. It also lists refractive indices of several atmospheric aerosols. The line-by-line portion of the database contains spectroscopic parameters for 42 molecules (and their significant isotopologues). HITRAN lists line positions, intensities, Einstein *A*-coefficients, halfwidths and associated properties, lower-state energies, statistical weights of the upper and lower levels of the transitions, and a system for the representation of quantum identifications and enhanced referencing and uncertainty codes. In addition, there is a provision for making corrections to the broadening of line transitions due to line mixing.

This portion of the HITRAN compilation currently contains about 2×10^6 individual transitions.

The cross-section portion of HITRAN contains experimentally determined absorption cross-sections for many gases that are not amenable to a line-by-line representation. These gases include the fluorochlorocarbons, heavy oxides of nitrogen, and molecular species with very low-frequency fundamental modes.

There is also a portion of the HITRAN compilation that consists of aerosol indices of refraction for many particulates found in the terrestrial atmosphere.

A new version of HITEMP, a high temperature extension to HITRAN, has just been released [34]. So far HITEMP only contains data for five species (water, CO₂, CO, NO and OH), but the extension to elevated temperatures leads to a vast increase in the number of spectral lines in the database.

3.16. The National Institute of Standards and Technology (NIST) atomic database

NIST has been a major centre for A&M data compilation and has developed a number of numerical and bibliographic A&M databases. The largest of these, the atomic spectra database [39] contains evaluated data on about 77,000 energy levels and 144,000 spectral lines from atoms and ions of 99 elements. It is being continually updated and new versions appear several times a year. The Handbook of Atomic Spectroscopic Data [40] contains energy levels and prominent spectral lines for neutral and singly ionised atoms. The Spectral Data for the Chandra X-Ray Observatory database [41] presents critically compiled wavelengths (20–170 Å), energy levels, line classifications, and transition probabilities for several astrophysically important elements. Three bibliographic databases contain references on atomic energy levels and spectral lines, transition probabilities and spectral line shapes and line broadening and are updated daily. Finally, there are two non-local thermodynamic equilibrium (NLTE) population kinetic databases containing theoretical benchmark data [42,43].

4. User communities

Atomic and molecular data are indispensable in many scientific and technological areas: indeed the limited availability of accurate qualitative information on the properties of such A&M processes and on spectroscopic characteristics of participating atomic and molecular species remains a major restriction for the development of many of these scientific and technological fields. Here we present a few examples drawn from recent reviews for different research fields as examples of how VAMDC may address such a ‘roadblock’ and hence impact these and other fields of research.

4.1. Atmospheric science

Understanding the physics and chemistry of the terrestrial atmosphere and thence climate change with

all its social and economic consequences has required the construction of some of the largest, most complex computational models in the world today [44]. The development of such models is performed as an iterative process that is often driven by experimental observations. Such field observations are predominantly based on spectroscopic techniques and hence require a rich and diverse database of A&M photoabsorption cross sections both to determine the absolute concentrations in the terrestrial atmosphere and to quantify their role in the atmospheric radiative transport problem. Such data are then used to construct the models which predict the influence of natural or anthropogenic phenomena in our atmosphere. The models presently typically include about 100 species, with data drawn from both experiments and theory which must be sourced, validated, tabulated and cross referenced. One key exemplar is water vapour which has a central role in the terrestrial atmosphere as both the major absorber of incoming sunlight and the dominant greenhouse gas [45]. A modern water spectroscopy information system W@DIS is under construction as part of an IUPAC task group [35].

4.2. Astrochemistry (and planetary science)

All astronomical and space observations and most planetary explorations are necessarily remote observations requiring spectroscopic and collisional data. Because of the extraordinary range of physical conditions of temperature density and radiation environments encountered in different astrophysical objects and events, there is perhaps no scientific field in which the need for access to reliable A&M data is greater. Almost all the information about the universe is brought to us by the photons and processes by which they are created, modified and detected, being largely the domain of atomic, molecular and optical physics. For example cross sections and branching ratios for photo-destruction processes are essential if we are to understand the complex astro-chemical processes occurring in stellar systems harbouring extra-solar planets.

4.3. Plasma technologies

Plasma-based technologies make up one of the largest commercial industries in the world today. The selective and highly anisotropic etching of materials and the controlled deposition of thin films in the fabrication of microelectronic structures are among the most important methods of plasma-assisted materials processing, plasma polymerisation and plasma-assisted surface modification. Plasma-based processes are used in some 40% of the steps in the manufacture of semi-conductor chips. For modelling the behaviour and properties of such chemically active plasmas, detailed quantitative information is required for the most important collisional and radiative processes taking place both in the gas phase and on the surfaces requiring an enormously detailed database [46]. Indeed the recognition of this need by industry is such that several commercial databases have been developed but they are often incomplete and, crucially, many lack

rigorous academic review. The A&M databases for most technological plasma applications therefore remain inadequate for a full understanding of the corresponding dynamics and full exploitation of the optimization of these reactive plasmas. A coordinated international effort to improve the current status of the A&M database relevant to such plasma technologies has been highlighted by both academic and industrial communities for the future advancement of technological development.

4.4. Lighting

General purpose lighting consumes about 25% of all electricity produced, hence there is a great interest in improving lighting technologies. However despite the long and distinguished history of the lighting research community in establishing the required databases, the present status is still unsatisfactory particularly for the innovative technology developments using new working gas species (for example mercury-free fluorescent lamps, pure Xe, phosphorous and metal alloys such as InSb). Lapatovich [47] recently presented the need for a broad, well-coordinated joint effort by the A&M and lighting research communities to establish the necessary databases for the development of future light sources.

4.5. Fusion energy research

Atomic and molecular collisional and spectroscopic processes play an important role in magnetic fusion devices as they influence the energy balance of the confined plasma, the plasma transport and radiation properties, and are used as the basis for many plasma diagnostic methods. The design and operation of some vital fusion device systems, such as neutral beam heating, impurity control and thermal power and particle exhaust systems, require large amounts of A&M collisional and spectroscopic data. The need for coordination of A&M data acquisition efforts on an international scale became evident at the beginning of the 1970s. Assembling A&M data for fusion has been part of the central programme of the International Atomic Energy Agency (IAEA). The ITER programme brings new technological and scientific challenges and thence new requirements for the ALADDIN database (see <http://www-amdis.iaea.org/ALADDIN/>) maintained by the IAEA.

4.6. Radiation sciences

The effect of increased exposure to low doses of radiation in our daily lives has led to concerns that this may have long-term health effects, e.g. through damage of cellular DNA. Similarly, the use of ionizing radiation in medical therapy relies on a better understanding of how to control or induce damage of DNA within cancerous tissue. Both problems are explored by the development of complex radiotherapy models on the nanoscale, incorporating fundamental A&M data, for example, low-energy electron collisions with the DNA components (e.g. nucleobases) and other biomolecules such as nucleotides, amino acids and in the study of cellular walls and lipids.

However the database both for the spectroscopy and collisional cross sections for such compounds is scarce and to date has not been assembled in a single validated database. The VAMDC will provide the infrastructure within which this may be enacted.

5. Current efforts for A&M standards

Some prototype services and tools (<http://voparis-molecular.obspm.fr>) have been designed to automatically access the BASECOL [48,49] and CDMS [50,51] databases. These developments are based on the definition of standards [52,53] for the exchange of atomic and molecular physics data which are carried out within the context of the International Virtual Observatory Alliance (IVOA, <http://www.ivoa.net/>) for line lists [54]. The VAMDC's activities are based on the use of an XML schema XSAMS (XML Schema for Atomics & Molecular Spectroscopy) [55] which has been developed under the auspices of the IAEA and aims at describing atomic, molecular and particle-surface interaction data in distributed databases around the world. The VAMDC will test these standards and propose new extensions to cover the large range of different data types discussed above when necessary.

6. Conclusions

Atomic and molecular data play a key role in many scientific areas. The tendency over many decades has been for each area to develop its own specialised data repositories and associated protocols for accessing these data. This is inefficient as improved data do not permeate all communities rapidly and lead to significant duplication of effort. There are associated issues with data accuracy and validation: for example old and inaccurate data often remain in use in some areas even after other areas have adopted improved values. Use of disparate data sources also leads to difficulties comparing or linking different models.

The Virtual Atomic and Molecular Data Centre (VAMDC) aims to provide unifying structures which will allow the interlinking of databases and information systems worldwide. The VAMDC will develop the appropriate structures and protocols for this activity, while leaving the provision of the actual data in the hands of the experts in each particular field. This paper has set out the initial scope of the VAMDC activity and the tools that will be used to achieve its aims.

Acknowledgments

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.

References

- [1] Swain R, Vasisht G, Tinetti G. The presence of methane in the atmosphere of an extrasolar planet. *Nature* 2008;452:329–31.
- [2] Quinn P, Barnes D, Csabai I, Cui C, Genova F, Hanisch B, et al. The International Virtual Observatory Alliance: recent technical developments and the road ahead. *SPIE* 2004;5493:137–45.
- [3] Mason NJ. Radiation induced damage in macromolecular systems; where do we go next? Radiation damage in biomolecular systems. *AIP Conf Proc* 2008;1080:3–20.
- [4] Kupka F, Piskunov N, Ryabchikova TA, Stempels HC, Weiss WW. VALD-2: progress of the Vienna Atomic Line Data Base. *Astron Astrophys Sup Ser* 1999;138:119–33.
- [5] Dere KP, Landi E, Young PR, Del Zanna G, Landini M, Mason HE. CHIANTI—an atomic database for emission lines IX. Ionization rates, recombination rates, ionization equilibria for the elements hydrogen through zinc and updated atomic data. *Astron Astrophys* 2009;498:915–29.
- [6] Freedland SK. Solarsoft, 1999, soho <www.nascom.nasa.gov/solarsoft>.
- [7] Walton NA, The AstroGrid Consortium. The astrogrid virtual observatory service. *Astron Soc Pacific Conf Series* 2008;394:251–4.
- [8] Mason NJ. Electron induced processing; applications and data needs. In: *Proceedings of ICAMDATA06 AIP Conference Proceedings*, vol. 901; 2007. p. 74–84.
- [9] Itikawa Y, Mason N. Cross sections for electron collisions with water molecules. *J Phys Chem Ref Data* 2005;34:1–22.
- [10] Mason NJ. The status of the database for plasma processing. *J Phys D: Appl Phys* 2009;42:194003.
- [11] Müller HSP, Schlöder F, Stutzki J, Winnewisser G. The Cologne database for molecular spectroscopy, CDMS: a useful tool for astronomers and spectroscopists. *J Mol Struct* 2005;742:215–27.
- [12] Pickett HM, Poynter RL, Cohen EA, Delitsky ML, Pearson JC, Müller HSP. Submillimeter, millimeter and microwave spectral line catalog. *JQSRT* 1998;60:883–90.
- [13] Dubernet ML, Grosjean A, Daniel F, Flower D, Roueff E, Daniel F, et al. Ro-vibrational collisional excitation database BASECOL—<<http://basecol.obspm.fr/>>. *J Plasma Fusion Res Ser* 2006;7:356–57.
- [14] Dubernet ML, Daniel F, Moreau N, Vasserot AM, Marinakis S. BASECOL 2010: an molecular and atomic database for astrophysical data analysis. *A&A*, in preparation.
- [15] Schmitt BP, Volcke E, Quirico O, Brissaud N, Fray W, Grundy-JM, et al. GhoSST: the Grenoble astrophysics and planetology solid spectroscopy and thermodynamics database service: “RELEVANT Database” 2009; see <<http://ghosst.obs.ujf-grenoble.fr/>>.
- [16] Fray N, Schmitt B. Sublimation of ices of astrophysical interest. A bibliographic review. *Planet Space Sci* 2009;57:2053–80.
- [17] Woodall J, Agundez M, Markwick-Kemper AJ, Millar TJ. The UMIST database for astrochemistry 2006. *Astron Astrophys* 2007;466:1197–2003.
- [18] Millar TJ, Rawlings JMC, Bennett A, Brown PD, Charnley SB. Gas-phase reactions and rate coefficients for use in astrochemistry—the UMIST ratefile. *Astron Astrophys Sup Ser* 1991;87:585–619.
- [19] Mallocci G, Joblin C, Mulas G. On-line database of the spectral properties of polycyclic aromatic hydrocarbons. *Chem Phys* 2007;332:353–9.
- [20] Strazzulla G, Leto G, Palumbo ME. Ion irradiation experiments. *Adv Space Res* 1993;13:189–98.
- [21] Palumbo ME, Baratta GA, Fulvio D, Garozzo M, Gomis O, Leto G, et al. Ion irradiation of astrophysical ices. *J Phys Conf Series* 2008;101:012002.
- [22] Leto G, Baratta GA. Ly-alpha photon induced amorphization of Ic water ice at 16 Kelvin effects and quantitative comparison with ion irradiation. *A&A* 2003;397:7–13.
- [23] Palumbo ME, Baratta GA, Collings MP, McCoustra MRS. IR Optical Constants of solid CO. *PCCP* 2006;8:279–84.
- [24] Brunetto R, Roush TL. Impact of irradiated methane ice crusts on compositional interpretations of TNOs. *A&A* 2008;481:879–82.
- [25] Mulas G, Baratta GA, Palumbo ME, Strazzulla G. Profile of CH₄ IR bands in ice mixtures. *A&A* 1998;333:1025.
- [26] Brunetto R, Caniglia G, Baratta GA, Palumbo ME. Integrated near-infrared band strengths of solid CH₄ and its mixtures with N₂. *ApJ* 2008;686:1480.
- [27] Fulvio D, Sivaraman B, Baratta GA, Palumbo ME, Mason NJ. Novel measurements of refractive index, density and mid-infrared integrated band strengths for solid O₂, N₂O and NO₂:N₂O mixtures. *Spectrochim Acta Part A: Mol Biomol Spectrosc* 2009;72:1007–13.
- [28] Jevremović D, Dimitrijević MS, LČ Popović, Dačić M, Protić Benišek V, Bon E, et al. The project of Serbian Virtual Observatory and data for stellar atmosphere modeling. *New Astron Rev* 2009;53:222–6.
- [29] Faenov AY, Magunov AI, Pikuz TA, Skobelev IY, Loboda PA, Bakshayev NN, et al. Spctr-W-3 online database on atomic properties of atoms and ions. *AIP Conf Proc* 2002;636:253–62.

- [30] Perevalov VI, Tashkun SA. CDS-296 (Carbon Dioxide Spectroscopic Databank): updated and enlarged version for atmospheric applications. In: 10th HITRAN database conference, Cambridge, MA, USA; 2008.
- [31] Mikhailenko S, Barbe A, Babikov Y, Tyuterev VG. S&MPO—a databank and information system for ozone spectroscopy on the WEB. See: <<http://smo.iao.ru>>.
- [32] Rothman LS, Gordon IE, Barbe A, Benner DC, Bernath PF, Birk M, et al. The HITRAN 2008 molecular spectroscopic database. *JQSRT* 2009;110:533–72.
- [33] Jacquinet-Husson N, Scott NA, Chedin A, et al. The GEISA spectroscopic database: current and future archive for Earth and planetary atmosphere studies. *JQSRT* 2008;109:1043–59.
- [34] Rothman LS, Gordon IE, Barber RJ, Dothe H, Gamache RR, Goldman A, et al. HITEMP, the high-temperature molecular spectroscopy database. *JQSRT* 2010;111:2139–50.
- [35] Tennyson J, Bernath PF, Brown LR, Campargue A, Carleer MR, Csaszar AG, et al. IUPAC critical evaluation of the rotational-vibrational spectra of water vapor. Part I. Energy levels and transition wavenumbers for H₂¹⁷O, H₂¹⁸O, and HD¹⁶O. *JQSRT* 2009;110:573–96.
- [36] Cunto W, Mendoza C, Ochsenbein F, Zeippen C. TOPbase at the CDS. *Astron Astrophys* 1993;275:L5–8.
- [37] Mendoza C, Seaton MJ, Buerger P, Bellorin A, Melendez M, Gonzalez J, et al. OPserver: interactive online computations of opacities and radiative accelerations. *Mon Not R Astr Soc* 2007;378:1031–5.
- [38] Bautista MA, Kallman TR. The XSTAR atomic database. *Astrophys J Suppl* 2001;134:139–49.
- [39] Ralchenko Yu, Kramida AE, Reader J, NIST ASD Team. NIST Atomic Spectra Database (version 3.1.5). <<http://physics.nist.gov/asd3>>. National Institute of Standards and Technology, Gaithersburg, MD; 2008.
- [40] Sansonetti JE, Martin WC, Young SL. Handbook of Basic Atomic Spectroscopic Data (version 1.1.2). <<http://physics.nist.gov/Handbook>>. National Institute of Standards and Technology, Gaithersburg, MD; 2005.
- [41] Podobedova LI, Musgrove A, Kelleher DE, Reader J, Wiese WL, Coursey JS, et al. Spectral Data for the Chandra X-Ray Observatory, (version 1.1). <<http://physics.nist.gov/chandra>>. National Institute of Standards and Technology, Gaithersburg, MD; 2003.
- [42] Ralchenko Yu. NIST SAHA Plasma Kinetics Database (version 1.0). <<http://nlte.nist.gov/SAHA>>. National Institute of Standards and Technology, Gaithersburg, MD; 2006.
- [43] Rubiano JG, Florido R, Bowen C, Lee RW, Ralchenko Yu. Review of the 4th NLTE code comparison workshop. *High Energy Density Phys* 2007;3:225–32.
- [44] Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, editors. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge, UK: Cambridge University Press; 2007.
- [45] Solomon S, Rosenlof KH, Portmann RW, Daniel JS, Davis SM, Sanford TJ, et al. Contributions of stratospheric water vapor to decadal changes in the rate of global warming. *Sci Express* 2010;327:1219–23.
- [46] Kushner M, et al. (editors). Database needs for modelling and simulation of plasma processing. National Research Council Report, National Academy Report, Washington DC; 1996.
- [47] Lapatovich WP. The role of molecules in low temperature plasmas for lighting. *AIP Conf Proc* 2007;901:65–73.
- [48] Moreau N, Dubernet ML. VO access to BASECOL database. In: Gabriel C, Arviset C, Ponz D, Solano E, editors. Astronomical data analysis software and systems XV ASP conference series, proceedings of the conference held 2–5 October 2005 in San Lorenzo de El Escorial, Spain. San Francisco: Astronomical Society of the Pacific; 2006. p. 351–91.
- [49] Moreau N, Dubernet ML. Automatic access to BASECOL database and scientific applications. In: Barret D, Casoli F, Lagache G, Lecavelier A, Pagani L, editors. SF2A-2006: Proceedings of the Annual meeting of the French society of astronomy and astrophysics; 2006. p. 95.
- [50] Dubernet ML, Moreau N, Osuna P, Guainazzi M, Salgado J, Roueff E, et al. Definition and implementation of VO standards for the access of atomic and molecular linelists. In: Barret D, Casoli F, Lagache G, Lecavelier A, Pagani L, editors. SF2A-2006: Proceedings of the Annual meeting of the French society of astronomy and astrophysics; 2006. p. 87.
- [51] Moreau N, Dubernet ML, Müller H. VO access to CDMS spectroscopic database. In: Guainazzi M, Osuna P, editors. Astronomical spectroscopy and virtual observatory, Proceedings of the EURO-VO workshop, held at the European Space Astronomy Centre of ESA, Villafraanca del Castillo, Spain, 21–23 March, 2007. European Space Agency; 2008. p. 195.
- [52] Salgado J, Osuna P, Guainazzi M, Barbarisi I, Dubernet ML, Tody D. Simple Line Access Protocol. In: Osuna P, Salgado J, editors. IVOA, Version 1.0, July 2009, <<http://www.ivoa.net/Documents/SLAP/>>.
- [53] Osuna P, Guainazzi M, Salgado J, Dubernet ML, Roueff E. Simple Spectral Lines Data Model. Osuna P, Salgado J, editors. IVOA, Version 1.0, July 2009, <<http://www.ivoa.net/Documents/SSLDM/>>.
- [54] Dubernet ML. The virtual observatory: its goals and the relevance of atomic and molecular data. *AIP Conf Proc* 2007;901:233–8.
- [55] Dubernet ML, Humbert D, Clark REH, Ralchenko Yu, Roueff E, Schultz DR. XSAMS: XML schema for Atomic, Molecular and Solid Data. In: Dubernet ML, Humbert D, Ralchenko Yu, editors. Version 0.1, September 2009, <<http://www-amdis.iaea.org/xml/>>.