Proceedings of the First EELA-2 Conference
B. Marechal et al. (Eds.)
CIEMAT 2008

Virtual Atomic and Molecular Data Centre

C. Mendoza[1] and L.A. Nuñez [2]

[1]*Centro de Física, IVIC, Caracas, Venezuela, and CeCalCULA, Mérida, Venezuela*
claudio@ivic.ve
[2] *Centro de Física Fundamental, ULA, and CeCalCULA, Mérida, Venezuela*
nunez@ula.ve

Abstract

A large-scale computational project has been recently launched to develop the Virtual Atomic and Molecular Data Centre (VAMDC). It involves several research groups from the European Union (Austria, France, Germany, Italy, Sweden, UK), the Russian Federation, Serbia and Venezuela. VAMDC intends to deploy an advanced and secure cyber-infrastructure for providing an interoperable e-science environment for existing distributed atomic and molecular databases. These databases are used in a variety of research and industrial fields such as astrophysics, atmospheric physics, fusion, environmental sciences, lasers and lighting. Research groups from Venezuela, namely from IVIC and CeCalCULA, will make use of the EELA grids and storage elements for upgrading, testing and tuning atomic database services and applications such as TOPbase, TIPbase, OPserver and XSTAR. In the present report we will describe the project guidelines, especially in reference to the training of potential users and expertise dissemination in Latin America.

1. Introduction

Atomic and molecular (A&M) data are used in a myriad of research fields (astrophysics, fusion plasmas, atmospheric physics and chemistry and quantum optics) and technological applications (lighting, semiconductor manufacturing, environmental sciences, molecular biology and nanotechnology). Data are obtained from both laboratory measurements and computations, including long-term international collaborations such as the Ferrum Project [1], the Opacity Project [2] and the Iron Project [3]. Due to the large data volumes usually involved, publication entitle devising set of tables, compilations and databases, which immediately brings to mind the three-volume reference source, *Atomic Energy Levels,* published by Charlotte E. Moore between 1948−58. It has been described by her colleague at the National Bureau of Standards (NBS) William C. Martin as "one of the most highly respected and frequently cited sources of basic atomic data ever published" [4].

The development of A&M databases has generally kept up with the fast computer evolution, the advent of the Internet and, more recently, with the ubiquitous establishment of the WWW. However, the intensive use of huge distributed data reservoirs in a growing number of scientific fields, e.g. high-energy physics, astronomy, genomics and proteomics, is rapidly leading to a new way of doing science: e-science. E-science may be visualized as a research enterprise driven by global and dynamic collaborations on a second-generation Internet. Within this context, a project to develop a “virtual atomic and molecular data center”, to be referred hereafter as VAMDC, has been recently launched to upscale and integrate A&M data activities and services to the new paradigm. In the present report, we describe the scientific institutions and the databases taking part, some of the endemic problems encountered in data collecting and dissemination, the tasks and issues involved in the setting up of the VAMDC cyber-infrastructure and the use of the EELA-2 [5] grid by the Venezuelan node to optimize database services and access by distributed applications.

2. Participating institutions

VAMDC will integrate research groups from a distinguished list of scientific institutions mainly from the European Research Area (ERA).

- **Centre National de la Recherche Scientifique** (CNRS), France, including institutes from sites in Paris, Reims, Grenoble, Bordeaux, Dijon and Toulouse.
- **University of Cambridge**, UK, with the Institute of Astronomy and the Department of Applied Mathematics and Theoretical Physics.
- **University College London**, UK, through the Department of Physics and Astronomy and Mullard Space Science Laboratory.
- **The Queen’s University Belfast**, UK.
- **The Open University**, UK.
- **Observatory of Vienna**, University of Vienna, Austria.
- **University of Uppsala**, Sweden.
- **University of Cologne**, Germany.
- **Istituto Nazionale di Astrofisica**, Italy, with the Osservatorio Astronomico of Cagliori and the Osservatorio Astrofisico di Catania.
- **Astronomical Observatory of Belgrade**, Serbia.
- **Institute of Astronomy**, Russian Academy of Sciences, Russian Federation.
- **Institute of Atmospheric Optics,** Russian Academy of Sciences, Russian Federation.
- **Institute for Spectroscopy**, Russian Academy of Sciences, Russian Federation.
- **Russian Federal Nuclear Centre All-Russian Institute of Technical Physics**, Russian Federation.
- **Venezuelan Institute for Scientific Research** and **Centro Nacional de Cálculo Científico Universidad de Los Andes**, Venezuela.

3. A&M databases

We provide here a portfolio of the database resources that will become the core of VAMDC.

- **VALD** [6]. A collection of atomic and molecular data of astronomical interest maintained by seven European institutes. Main nodes are in Vienna, Uppsala and Moscow, they serve about 1000 users from 70 countries and process over 200 requests per day.

- **CHIANTI** [7]. Critically evaluated set of up-to-date atomic data to analyze the spectra from optically thin astrophysical plasmas. It lists experimental and calculated wavelengths, radiative data and rates for electron and proton collisions.
- **CDMS** [8]. It provides recommendations for spectroscopic transition frequencies and intensities for atoms and molecules of astronomical interest and for studying the Earth atmosphere.
- **BASECOL** [9]. Excitation rate coefficients for ro-vibrational excitation of molecules by electrons, He and H_2, mainly used for the study of interstellar, circumstellar and cometary atmospheres.
- **STSP** [10]. Spectroscopic laboratory data of molecular and atomic solids and liquids comprising from the near UV to the far-infrared
- **UMIST** [11]. Reaction rate data for chemical kinetic modeling of astronomical regions.
- **KIDA**. This database will list chemical reactions that are included in the modeling of the chemistry of the interstellar medium and planetary atmospheres.
- **PAH** [12]. It lists general energetics, ground-state optimized geometries, harmonic vibrational frequencies and IR activities and photoabsorption cross-sections for polycyclic aromatic hydrocarbons and carbon clusters.
- **LASP** [13]. This database contains: (i) IR spectra of molecules in the solid phase for both pure species and their mixtures before and after processing with energetic ions and UV photons; (ii) IR optical constants of molecules in the solid phase and of frozen molecules after processing with energetic ions; (iii) band strengths of IR absorption bands and (iv) density values of frozen samples.
- **BELDATA** [14]. Calculated widths and shifts for isolated atomic lines due to electron and ion collisions. They are used in the modeling and diagnostics of stellar atmospheres and envelopes.
- **Spectr-W^3** [15]. Experimental, calculated, and compiled data on ionization potentials, energy levels, wavelengths, radiation transition probabilities and oscillator strengths and parameters for analytical approximations of electron-collisional cross-sections and rates for atoms and ions.
- **CDSD** [16]. Calculated spectral line parameters for seven isotopologues of carbon dioxide.
- **OZONE** [17]. Spectral line parameters, simulated and observed absorption spectra (microwave to IR) and other relevant information for the ozone molecule.
- **SPECTRA** [18]. A compilation of the well-known databases HITRAN, GEISA, HITEMP and other line lists.
- **W@DIS** [19]. Levels, transition and line profile characteristics of the water molecule and its isotopologues.
- **TOPbase** [20]. Opacity Project LS-coupling energy levels, gf-values and photoionization cross sections for light elements ($Z \leq 26$) of astrophysical interest.
- **TIPbase** [21]. Intermediate-coupling energy levels, A-values and electron impact excitation cross sections and rates for astrophysical applications ($Z \leq 28$). Computed by the IRON Project.
- **OPserver** [22]. Remote, interactive server for the computation of mean opacities for stellar modeling using the Iron Project monochromatic opacities.
- **XSTAR** [23]. An atomic database for the modeling of photoionized plasmas based on the database currently used by the code XSTAR.
- **HITRAN** [24]. It lists individual line parameters for molecules in the gas phase (microwave through to the visible), photoabsorption cross-sections for 39 molecules and refractive indices of several atmospheric aerosols.

- **NIST** [25]. It lists evaluated data on about 77000 energy levels and 144000 spectral lines from 99 chemical elements. Additional databases contain critically compiled wavelengths (20−170 Å), energy levels, line classifications and transition probabilities for elements of astrophysical interest and NLTE population kinetics data. It maintains three bibliographic databases with references on atomic energy levels and spectral lines, transition probabilities, spectral line shapes and line broadening parameters.

4. Problems in data A&M activities

Some of the highlights and problems in the development of A&M database have been previously discussed [26]. While some lines of A&M research have high profiles, for instance Bose-Einstein condensations, quantum dots, entanglement, spintronics and nanoscience, the actual production and dissemination of A&M data, in spite of its relevance in important fields, have been traditionally low-budget and overlooked. Several experimental groups, in particular, have lost momentum or have closed indefinitely as a result, causing severe unattended client needs.

Due to the lack of standards and common guidelines, outstanding problems in existing A&M database are interoperability and data interfaces which hamper productive searches and data mining. Attempts to implement A&M search engines such as GENIE and DANSE [27,28] have been interesting, but perhaps did not receive sufficient support from database developers. Data exchange has also been carried out in a somewhat informal manner (emails, undocumented ASCII files, peer-to-peer arrangements) even though standard formats from specific client fields (e.g. the FITS astronomical data format [29]) have been incorporated. The data exchange tool ALADDIN [30,31] has played a well-recognized role, but it is based on a set of Fortran-77 programs rather than on more current XML schemata. For this reason, an Atomic and Molecular Data Markup Language (AMDML) [32] has been recently presented although its adoption by the A&M data community still remains an open question.

A&M data are stored in a variety of relational database management systems (DBMS) using very diverse data models. Many of these DBMS are not SQL standard packages but local developments that, in the long run, can compromise data integrity and regular updating procedures. Furthermore, in most cases data are not even housed in databases but as plain files in data centers, servers belonging to scientific journals or departmental, project and personal web pages. In order to implement efficient access methods within such heterogeneous data domains, one begins to speak of “dataspaces” rather databases and of “dataspace support platforms” rather than DBMS [33]. Therefore, XML schemata become key not only for data exchange but also for data identification when implementing a new generation of search engines that must look “everywhere” in order to map the A&M universe and provide successful answers to user queries.

The growing sophistication of web technology has also become a problem for database developers who in most cases are physicist and chemists with a limited command of the many new computer languages and tools or short of funds to hire computer engineers. This situation is expected to become more acute in the e-science environment denoted by network applications, virtualization and grids.

5. VAMDC

The VAMDC proposal will address many of the issues raised in Section 4. It will become an accessible and interoperable e-infrastructure for AM data, upgrading and integrating an extensive portfolio of database services (see Section 3) and catering for the needs of a wide variety of data users from both academia and industry. Starting points will be the infrastructure and capabilities currently being developed by AstroGrid [34], Euro-VO [35] and EGEE [36], the creation of the core consortium and the programming of a series of network and service activities to establish self-sustainable computational and data mining services. Training of potential users and regular data dissemination in the ERA and worldwide have high priority. In Venezuela, three workshops are being planned which will start within the Congreso Latinoamericano de Computación de Alto Rendimiento (CLCAR-2009) in Choroní in September 2009.

6. EELA-2 collaboration

The Venezuelan node of VAMDC will be include the Computational Physics Laboratory of the Venezuelan Institute for Scientific Research (IVIC) and the Centro Nacional de Cálculo Científico Universidad de Los Andes (CeCalCULA). These two groups have been long-standing collaborators in scientific computing projects, especially to do with the EELA [5] initiative to integrate European and Latin American grids. Therefore, part of the VAMDC activities in Venezuela will make extensive use of the EELA-2 infrastructure, particularly with reference to database access from distributed applications and the gridification and virtualization of data mining and modeling codes. More precisely, for upgrading, testing and tuning TOPbase, TIPbase, OPserver and XSTAR.

7. Concluding remarks

In our opinion VAMDC is a topical example of the global collaborations and upscale innovations that are being brought about in e-science, and it is a welcome sign that the European Union funding agencies are taking into account data activities with consonant priority. VAMCD is expected to become a major European cyber-infrastructure with worldwide impact, and also a tribute to the many scientific groups, both experimental and theoretical, that generated tons of A&M data with the sole interest of being rather of use than famous.

References

[1] The Ferrum Project. Retrieved January 24, 2009, from *http://www.astro.lu.se/Research/astrophys/research/projects/ferrum.html*

[2] The Opacity Project. Retrieved January 24, 2009, from *http://cdsweb.u-strasbg.fr/topbase/op.html*

[3] The Iron Project. Retrieved January 24, 2009, from *http://www.usm.uni-muenchen.de/people/ip/iron-project.html*

[4] Charlotte E. Moore biography. Retrieved January 24, 2009, from

http://www.bookrags.com/biography/charlotte-e-moore/

[5] EELA-2. E-science grid facility for Europe and Latin America. Retrieved January 24, 2009, from *http://www.eu-eela.eu/*

[6] VALD: Vienna Atomic Line Database. Retrieved January 21, 2009, from *http://ams.astro.univie.ac.at/~vald/*

[7] CHIANTI: An atomic database for spectroscopic diagnostics of astrophysical plasmas. Retrieved January 21, 2009, from *http://www.damtp.cam.ac.uk/user/astro/chianti/*

[8] CDMS: Cologne Database for Molecular Spectroscopy. Retrieved January 21, 2009, from *http://www.astro.uni-koeln.de/site/vorhersagen/*

[9] BASECOL: Ro-vibrational collisional excitation database and utilities. Retrieved January 21, 2009, from *http://basecol.obspm.fr*

[10] B. Schmitt , P. Volcke, V. Gouanère, E. Quirico, N. Fray , A. Pommerol, "STSP: data bases of Spectroscopy and Thermodynamics of Planetary Solids", EPSC Abstracts, Vol. 3, EPSC2008-A-00560 (2008). Retrieved January 21, 2009, from *http://www.cosis.net/abstracts/EPSC2008/00560/EPSC2008-A-00560-1.pdf*

[11] UMIST: Database for astrochemistry. Retrieved January 21, 2009, from *www.udfa.net/*

[12] PAH: Theoretical spectral database of Polycyclic Aromatic Hydrocarbons and pure carbon clusters. Retrieved January 21, 2009, from *http://astrochemistry.ca.astro.it/database/*

[13] LASP: Laboratorio di Astrofisica Sperimentale. Retrieved January 21, 2009, from *http://web.ct.astro.it/weblab/*

[14] L.Č. Popović, M.S. Dimitrijević, N. Milovanović, "BelData - An Atomic Database for Astrophysical Purpose", Atomic and Molecular Data for Astrophysics: New Developments, Case Studies and Future Needs, 24th meeting of the IAU, Joint Discussion 1 (2000)

[15] Spectr-W^3: Atomic database for plasma spectroscopy and other applications. Retrieved January 21, 2009, from *http://spectr-w3.snz.ru/*

[16] CDSD: Carbon Dioxide Spectroscopic Databank. Retrieved January 21, 2009, from *http://cdsd.iao.ru/*

[17] OZONE: Spectroscopy & molecular properties of ozone. Retrieved January 21, 2009, from *http://ozone.iao.ru/*

[18] SPECTRA: Spectroscopy of atmospheric gases. Retrieved January 21, 2009, from *http://spectra.iao.ru/*

[19] W@DIS: Water Internet @ccesible Distributed Information System. Retrieved January 24, 2009, from *http://wadis.saga.iao.ru/*

[20] TOPbase: Opacity Project online database. Retrieved January 24, 2009, from *http://cdsweb.u-strasbg.fr/topbase/*

[21] TIPbase: IRON Project online database. Retrieved January 24, 2009, from *http://cdsweb.u-strasbg.fr/tipbase/*

[22] OPserver: Opacity Project interactive opacity server. Retrieved January 24, 2009, from *http://opacities.osc.edu/*

[23] M.A. Bautista, T.R. Kallman, "The XSTAR atomic database", Astrophys. J. Suppl. Ser. 134, 139 (2001)

[24] HITRAN: HIgh-resolution TRANsmission molecular absorption database. Retrieved January 24, 2009, from *http://www.cfa.harvard.edu/HITRAN/*

[25] NIST: National Institute for Standards and Technology databases. Retrieved January 24, 2009, from *http://physics.nist.gov/PhysRefData/contents.html*

[26] C. Mendoza, "Atomic databases", in Atomic Data Needs for X-ray Astronomy, M.A. Bautista, T.R Kallman, A.K. Pradhan (eds), NASA/CP-2000-209968, p. 167 (2000)

[27] D. Humbert, Yu. Ralchenko, R.E.H. Clark, K. Katsonis, "GENIE and DANSE: two atomic and molecular data web search engines for fusion and plasma physics", 30th EPS Conference on Contr. Fusion and Plasma Phys., ECA Vol. 27A, P-4.160pd (2003). Retrieved January 24, 2009, from *http://epsppd.epfl.ch/StPetersburg/PDF/P4_160.PDF*

[28] GENIE: A GENeral Internet search Engine for atomic data. Retrieved January 24, 2009, from h*ttp://www-amdis.iaea.org/GENIE/*

[29] FITS: Flexible Image Transport System. Retrieved January 24, 2009, from *http://heasarc.gsfc.nasa.gov/docs/heasarc/fits_overview.html*

[30] R.A. Hulse, "The ALADDIN atomic physics database system", Atomic Processes in Plasmas, AIP Conference Proceedings, Volume 206, p. 63 (1990)

[31] ALADDIN: atomic physics database system. Retrieved January 24, 2009, from *http://www-amdis.iaea.org/ALADDIN/*

[32] Y. Ralchenko, R.E.H Clark, D. Humbert, D.R. Shultz, T. Kato, Y.J Rhee, "Development of the atomic and molecular markup language for Internet data exchange", J. Plasma Fusion Res. Ser., 7, 338 (2006). Retrieved January 24, 2009, from *http://www.jspf.or.jp/Journal/JPFR/itc14/jpfr2006_07-338.pdf*

[33] M. Franklin, A. Halevy, D. Maier, "From Databases to Dataspaces: A new abstraction for information management", SIGMOD Record, 34, 27, (2005)

[34] AstroGrid: Virtual Observatory software for astronomers. Retrieved January 26, 2009, from *http://www.astrogrid.org/*

[35] EURO-VO: The European Virtual Observatory. Retrieved January 26, 2009, from *http://www.euro-vo.org/pub/*

[36] EGEE: Enabling Grids for E-sciencE. Retrieved January 26, 2009, from *http://www.eu-egee.org/*