VAMDC - The Virtual Atomic and Molecular Data Centre - A New Way to Disseminate Atomic and Molecular Data - VAMDC Level 1 Release

G. Rixon[1], M. L. Dubernet[2,3], N. Piskunov[4], N. Walton[5], N. Mason[6], P. Le Sidaner[7], S. Schlemmer[8], J. Tennyson[9], A. Akram[5], K. Benson[10], J. Bureau[11], M. Doronin[11], C. Endres[8], U. Heiter[4], C. Hill[9], F. Kupka[12], L. Nenadovic[11], T. Marquart[4], G. Mulas[13], Y. Ralchenko[14], A. Shih[7], K. Smith[15], B. Schmitt[16], D. Witherick[9], V. Boudon[17], J. L. Culhane[10], M. S. Dimitrijevic[18], A. Z. Fazliev[19], C. Joblin[20], G. Leto[21], P. A. Loboda[22], H. E. Mason[23], C. Mendoza[24], T. J. Millar[15], L. A. Nunez[25], V. I. Perevalov[19], L. S. Rothman[26], E. Roueff[3], T. A. Ryabchikova[27], A. Ryabtsev[28], S. Sahal-Bréchot[29], V. G. Tyuterev[30], V. Wakelam[31] and C. J. Zeippen[32]

[1]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK*
[2]*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*[1]
[3]*Laboratoire Univers et Théories, UMR8102 CNRS/INSU, Observatoire de Paris, Section Meudon, 5 Place Janssen, 92195 Meudon Cedex, France*
[4]*Uppsala University, Department of Physics and Astronomy, Lägerhyddsvägen 1, Uppsala 75120, Sweden*
[5]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB30HA, UK*
[6]*Open University, Faculty of Science, Walton Hall, Milton Keynes, MK7 6AA, UK*
[7]*Division Informatique de l'Observatoire, VO-Paris Data Centre, UMS2201 CNRS/INSU , Observatoire de Paris, 5 Place Janssen, 92195 Meudon Cedex, France*
[8]*University of Cologne, I. Physikalishes Institut, Zuelpicher Strasse 77, Koeln 50937, Germany*
[9]*Department of Physics and Astronomy, University College London, London WC1E 6BT, UK*
[10]*Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK*
[11]*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*
[12]*Faculty of Mathematics, University of Vienna, Nordbergstrasse 15, 1090 Wien, Austria*
[13]*Istituto Nazionale di Astrofisica - Osservatorio Astronomico di Cagliari, strada 54 loc. Poggio dei Pini, Capoterra (CA), I-09012, Italy*
[14]*National Institute of Standards and Technology, Atomic Physics Division, 100 Bureau Dr., Stop 8422, Gaithersburg, MD 20899-8422, USA*
[15]*School of Mathematics and Physics, Queen's University Belfast, University Road, Belfast BT7 1NN, UK*
[16]*Laboratoire de Planétologie de Grenoble, UMR5109 CNRS/INSU, Université Joseph Fourier, BP53, 38041 Grenoble Cedex 9, France*
[17]*Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 5209 CNRS-Université de Bourgogne 9 Avenue Alain Savary, BP 47 870, F-21078 DIJON Cedex*

7th International Conference on Atomic and Molecular Data and Their Applications - ICAMDATA-2010
AIP Conf. Proc. 1344, 107-115 (2011); doi: 10.1063/1.3585810

[18]Astronomical Observatory, Volgina 7, 11060 Belgrade, Serbia
[19]V.E. Zuev Institute of Atmospheric Optics, Siberian Branch, Russian Academy of Sciences, 1, Academician Zuev square, Tomsk 634021, Russian Federation
[20]Centre d'Etude Spatiale des Rayonnements, UMR5187 CNRS/INSU, Université Paul Sabatier, 9 avenue du Colonel Roche, F-31028 Toulouse cedex 9, France
[21]Istituto Nazionale di Astrofisica, Osservatorio Astrofisico di Catania, Via Santa Sofia 78, I-95123 Catania, Italy
[22]Russian Federal Nuclear Centre- All Russian Institute of Technical Physics (RFNTC-VNIITF), 13 Vasilyeva St., Snezhinsk, Chelyabinsk region, 456770, Russia
[23]Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK
[24]Centro de Física, Instituto Venezolano de Investigaciones Científicas (IVIC), PO Box 20632, Caracas 1020A, Venezuela, and Centro Nacional de Cálculo Científico Universidad de Los Andes (CeCalCULA), Corporación Parque Tecnológico de Mérida, Mérida 5101, Venezuela.
[25]Centro Nacional de Calculo Científico Universidad de Los Andes (CeCalCULA), Corporacion Parque Tecnologico de Mérida, Mérida 5101, Venezuela, and Grupo de Investigacion en Relatividad y Gravitacion (GIRG) Esc. Física, Universidad Industrial de Santander, Bucaramanga Colombia.
[26]Harvard-Smithsonian Center for Astrophysics, Atomic and Molecular Physics Division, MS 50, 60 Garden Street, Cambridge, MA 02138-1516, USA
[27]Institute for Astronomy RAS, Pyatnitskaya 48, Moscow 119017, Russian Federation
[28]Institute for Spectroscopy RAS, Physical 5, Troitsk, 142190, Russian Federation
[29]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
[30]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
[31]Laboratoire d'Astrophysique de Bordeaux, UMR5804 CNRS/INSU, Université de Bordeaux, BP89, 33271 Floirac Cedex, France
[32]Laboratoire d'Etude du Rayonnement et de la Matiére en Astrophysique, UMR8112 CNRS/INSU, Observatoire de Paris, 61, Avenue de l'Observatoire, 75014 Paris, France

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, evaluation, and use of atomic and molecular data. VAMDC aims to build a reliable, open, flexible and interoperable e-science 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. This paper describes the organisation of the project and the achievements during its first year.

Keywords: Atomic & molecular physics, Computing, Archives
PACS: 30, 32, 33, 34, 89.20f, 89.20h

INTRODUCTION

The Virtual Atomic and Molecular Data Centre (VAMDC[2]) is a European-Union-funded collaboration between groups involved in the generation, evaluation, and use

[1] Corresponding author:marie-lise.dubernet@obspm.fr
[2] `http://www.vamdc.eu/`

of atomic and molecular data, as well as in the technical development and use of key e-infrastructures (e.g. the Euro-VO[3]) and the European Grid Initiative[4]. The authors of [1] constitute the core partners of the project. The VAMDC e-Infrastructure involves 15 legal institutes from 6 European Union member states, partners in non-EU countries (the Russian Federation, Serbia and Venezuela) and external partners in the US. It is a 42-month project that started on 1 July 2009.

VAMDC aims to build a reliable, open, flexible and interoperable e-science interface to existing atomic and molecular data. Initially, the core of the VAMDC e-infrastructure is based on the databases detailed in [1]; VAMDC welcomes the addition of other database resources in due course. VAMDC key objectives are 1) to implement VAMDC interface for accessing major existing databases containing heterogeneous data and aimed at different users, 2) to enable data queries across multiple databases that are focussed on specific research topics, 3) to enable data publishing/quality control process for major Atomic and Molecular (A&M) data-producers, 4) to involve wide user and producer communities in development and use of VAMDC. User communities include astrophysics, atmospheric sciences, plasma physics, combustion media to lighting and etching industries, with various approaches such as simulations, observations and diagnostics.

The project is organized in Networking activities, Service Activities and Joint Research Activities whose objectives are described in the following sections together with the achievements during the first year.

NETWORKING ACTIVITIES (NAs)

Objectives

The NAs foster a culture of cooperation between A&M scientists, database providers and data users throughout Europe. Three work packages (WP) are active: WP1 for internal management of VAMDC, including financial control of the project, reporting to the EU, and formal packaging of deliverables; and two NAs. NA1 provides the scientific and technical direction necessary for the development of the VAMDC e-infrastructure, while NA2 provides the links between VAMDC and the wider user community, being responsible for training and dissemination.

Achievements during Year 1

NA1 has focussed on coordination with key external standards groups such as the International Virtual Observatory [5], the XSAMS group [6], and EGI; and on coordina-

[3] http://www.euro-vo.org/
[4] http://www.egi.eu/
[5] http://www.ivoa.net/
[6] http://www-amdis.iaea.org/xml/

tion with key external domain groups, e.g. Euro-VO (Astronomy - VO technology), Gaia/GREAT[7] (Galactic Astronomy), HELIO[8] (Solar/STP) and EuroPlanet[9] (Planetary science). These actions benefit VAMDC through feedback from users of A&M data and by keeping in touch with technical developments in other projects, and shall continue in Period 2. Another aspect to WP2 was to put together policies related to standards and publication in VAMDC. Some simple steps have been achieved, such as having a reference paper published in JQSRT [1].

NA2 has established links worldwide in Asia, Russia, South America, USA and within many different communities of producers and users of A&M data. Details of all presentations to the community can be found on the VAMDC web-site. [10]

SERVICE ACTIVITIES (SAs)

Objectives

The key objective of the SAs is to provide access to an inclusive range of high quality data and applications services to the research community. The VAMDC partners represent major data producers. The SAs make these data available on the WWW in a consistent and supported form.

SA1, Infrastructure Deployment, establishes web services at the sites of VAMDC partners. These services provide access to A&M databases; supply metadata informing the use of those databases; and allow higher-order data products to be derived from the archived data by execution of applications at the archive site. Where a partner holds data that are not in a suitable form for remote querying, SA1 assists in the creation of suitable databases. Further, where a small data-producer does not want to run their own database, SA1 can arrange for hosting of those data and services at a VAMDC site. SA1 also provides a web portal and desktop utilities for access to the services.

SA2, Support to Infrastructure, supports the operation of the services deployed by SA1. Email support is available both to data producers and to scientists using the data. SA2 monitors the health of the deployed services. Some support is available for users who want to adapt their own software to the grid. Since VAMDC makes A&M data from different sources more easily comparable, SA2 is able to assess quality by looking for discrepancies between database.

Achievements during Year 1

A level-1 infrastructure was release, including a registry service for the metadata and services for a selected set of databases which served as a test of the technology. The

---

[7] http://www.ast.cam.ac.uk/GREAT/
[8] http://www.helio-vo.eu/
[9] http://www.europlanet-eu.org/
[10] http://voparis-twiki.obspm.fr/twiki/bin/view/VAMDC/TalksVamdc

data services respond to database queries and emit data extracts in the XSAMS format. Monitoring of the deployed services is operational. VAMDC beta-testers can access the grid at Paris Observatory.

In Year 2, there will be a level-2 release with data services for all data-sets held by VAMDC partners and some prototypes of the derived-data services. SA2 will arrange access to EGI for VAMDC users and will start the quality assessment of the data.

JOINT RESEARCH ACTIVITIES (JRAs)

Objectives

The Joint Research Activities build the complete set of "tools" necessary to create the VAMDC infrastructure, creating new specifications and creating/adapting/integrating new software. All the VAMDC software and supporting libraries will be available under free-software licenses.

JRA1, Interoperability, defines standards necessary to build a consistent infrastructure. It specifies data models, query languages, service protocols and dictionaries of standard terms.

JRA2, Publishing Tools, provides the software by which SA1 can deploy archive-data services for data producers. The tools cover generation of relational databases from ASCII files and the web services that respond to remote queries on those databases, following the standards developed by JRA1.

JRA3, New Mining and Integration Tools, develops software for cross-matching and cross-federation of heterogeneous resources and application services wrapping complex work flows combining AM data access, manipulation, and integration into user processing chains.

Achievements during Year 1

JRA1 has prototyped new schemata for the molecular part of XSAMS. These extensions proved very useful in the design of VAMDC services and shall become part of the XSAMS core during 2011. A separate data model has been developed for solid spectroscopy which is not totally included in XSAMS.

JRA1 has defined the TAP-XSAMS protocol for data services. This extends the IVOA Table Access Protocol with the XSAMS data-model for A&M data and the query language VAMDC SQL sub-set 1 (VSS1). VSS1 in turn uses the standard names established in the VAMDC Dictionary.

JRA2 has developed two prototypes of the publishing tools, investigating the benefits of different languages and libraries and of two different approaches to the development.

The first prototype is designed to be reusable at many VAMDC sites. It is flexible and can be configured to suit an existing database, but also includes the code to generate its own database from data in ASCII files. The adaptability comes from the use of the Django framework for web-services and the software is therefore written in Python.

The second prototype is written specifically for the BASECOL database, in Java. The code is co-developed with the database (the database was given extra tables to better support XSAMS) and aims for efficiency rather than adaptability to other data-sets. Figure 1 shows the internal architecture of this service. Initially, the service was implemented with specialized, Java code that depends on the structure of the BASECOL database. That approach allowed a working prototype to be quickly developed but left much code to be maintained. The hand-written code has been progressively replaced with a data-access layer generated, by tooling, from the database schema. This refinement takes longer to achieve but makes maintenance much easier.

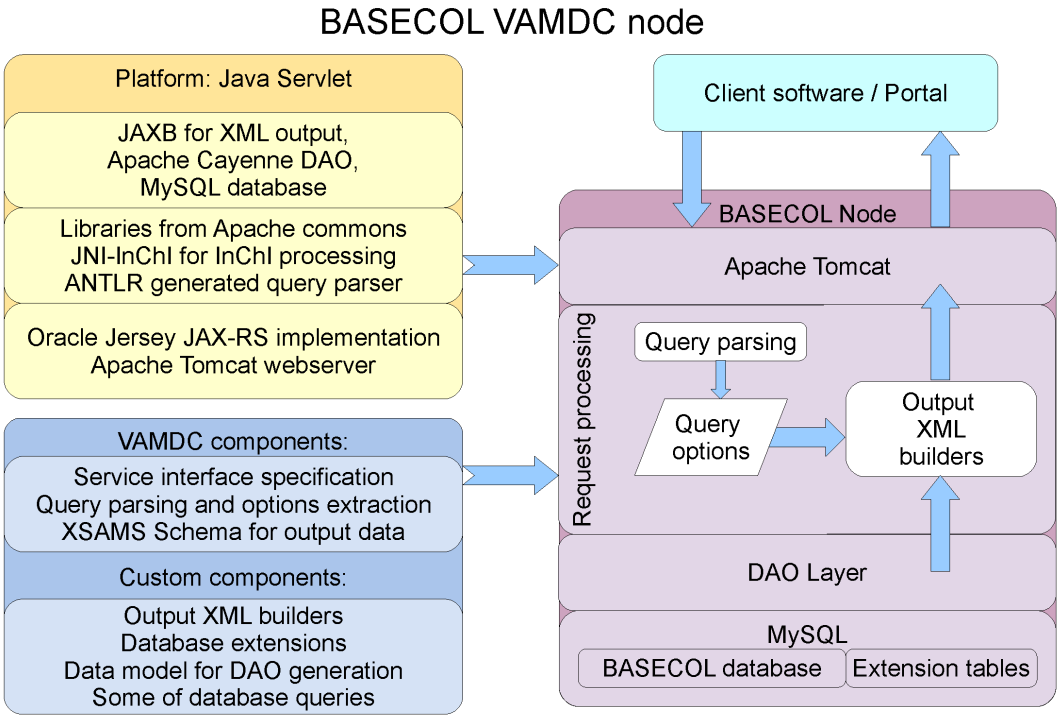


FIGURE 1: Internal architecture of BASECOL archive-data service for VAMDC standards. The client/portal component is a separate installation and is not specific to BASECOL.

Both prototypes are considered successful and are used in the level-1 release. The BASECOL prototype proves that it is feasible for publishers to write a local implementation of the VAMDC standards. The Dango-based prototype proves that it is possible for VAMDC to provide code to data producers that can be adapted to their data sets with little extra development. In year 2, the Django-based software will be refined and released to data producers. The BASECOL service-software will be made available as a guide to local development.

JRA3 has focused on designing a tool for handling XSAMS outputs. This is a useful 'user test-bench' for the evaluation of the standards in JRA1, and a refined version will be released in year 2. Specification have been prepared for a tool allowing handling both gas phase and solid spectroscopy for planetology and interstellar medium applications, to be built in year 2.

VAMDC LEVEL 1 RELEASE

The level-1 release provides archive-data services for a selection of databases: VALD [2], XstarDB [3], BASECOL [4, 5], CDMS [6]. A registry of metadata and a web portal support these services. There is also a 'broker' service, following the Simple Line Access Protocol of IVOA, that combines data from all the archive services.

The set of databases is deliberately restricted in the level-1 release. The level-2 release, planned for 2011, will connect far more data-sets, including CHIANTI [7], eMOL [8], GhoSST [9], HITRAN [10], SPECTRW [11], SMPO [12], TOPbase [13], UDfA [14] and a database of polycyclic hydrocarbons [15].

Users are able to investigate the level-1 release using the provided web-portal, and could also use the services in their own application. To prove the latter approach, we produced a simple web-site that generates spectral-line lists from the XSAMS output of the archive services and presents them as web pages in HTML. Figure 2 shows the query form and Fig. 3 the results.

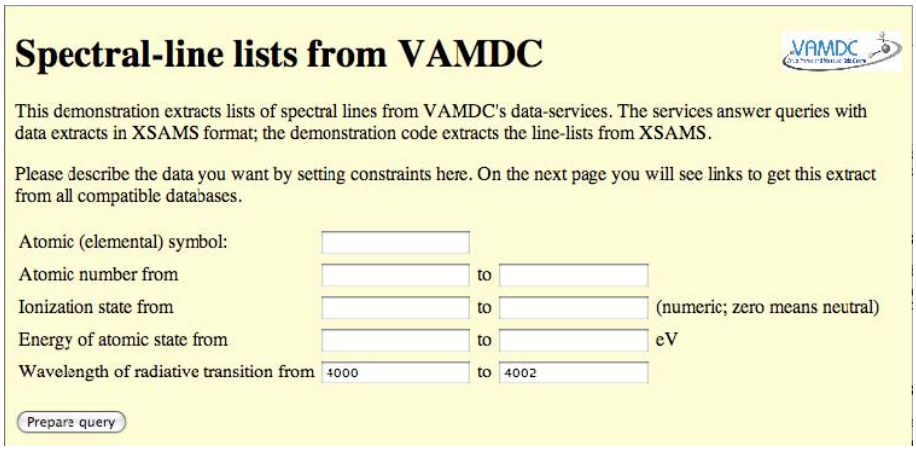
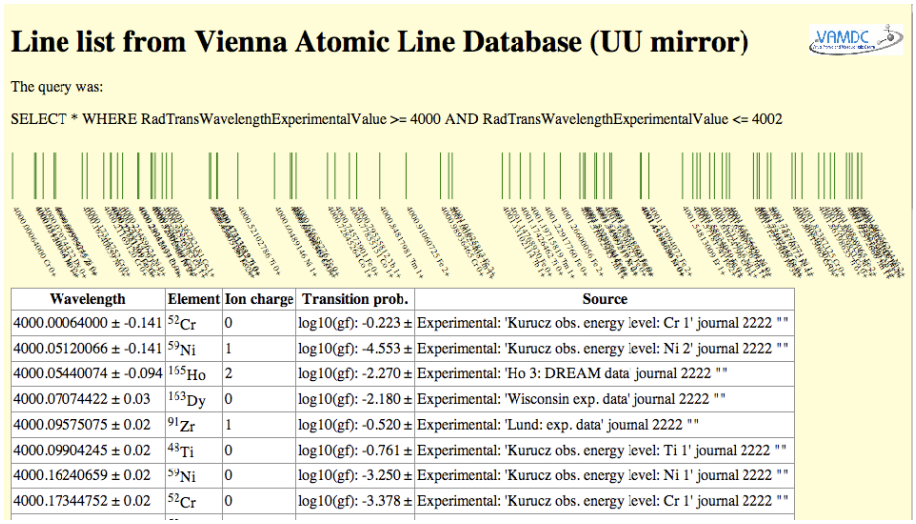


FIGURE 2: Query form for spectral-line application



Wavelength	Element	Ion charge	Transition prob.	Source
4000.00064000 ± -0.141	^{52}Cr	0	log10(gf): -0.223 ±	Experimental: 'Kurucz obs. energy level: Cr 1' journal 2222 ""
4000.05120066 ± -0.141	^{59}Ni	1	log10(gf): -4.553 ±	Experimental: 'Kurucz obs. energy level: Ni 2' journal 2222 ""
4000.05440074 ± -0.094	^{165}Ho	2	log10(gf): -2.270 ±	Experimental: 'Ho 3: DREAM data journal 2222 ""
4000.07074422 ± 0.03	^{163}Dy	0	log10(gf): -2.180 ±	Experimental: 'Wisconsin exp. data' journal 2222 ""
4000.09575075 ± 0.02	^{91}Zr	1	log10(gf): -0.520 ±	Experimental: 'Lund: exp. data' journal 2222 ""
4000.09904245 ± 0.02	^{48}Ti	0	log10(gf): -0.761 ±	Experimental: 'Kurucz obs. energy level: Ti 1' journal 2222 ""
4000.16240659 ± 0.02	^{59}Ni	0	log10(gf): -3.250 ±	Experimental: 'Kurucz obs. energy level: Ni 1' journal 2222 ""
4000.17344752 ± 0.02	^{52}Cr	0	log10(gf): -3.378 ±	Experimental: 'Kurucz obs. energy level: Cr 1' journal 2222 ""

FIGURE 3: Visualization of spectral-line output from VALD.

This exercise show the power of the XSAMS format. All necessary information is contained in the XSAMS structure and, because that structure is XML, the web presentation of the line-list, including the graphics, can be generated from the XSAMS using a stylesheet; because the presentation detail is captured in the stylesheet, the rest of the web site needs little data-handling code and is only 137 statements in Java. The stylesheet can easily be changed to provide a different view of the data, or to transcribe selected data into formats other then HTML.

To test the transcription of XSAMS into machine-readable formats, we used the broker service implementing IVOA's Simple Line Access Protocol (SLAP). This service reformats the SLAP query into VAMDC's VSS1 query-language, forwards the query to the VAMC archive services and translates the results into IVOA's format using a stylesheet. Both translations proved straightforward to write.

In year 1, the release is only available to selected beta-testers within VAMDC. The level-2 release in 2011 will be more widely available.

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. M. L. Dubernet, et al, *J. Quantitat. Spectrosc. Radiat. Transfer* **111**, 2151–2159 (2010).
2. F. Kupka, N. Piskunov, T. A. Ryabchikova, H. C. Stempels and W. W. Weiss, *Astron. Astrophys. Sup. Ser.* **138**, 119–33 (1999).
3. M. A. Bautista and T. R. Kallman, *Astrophys. J. Suppl.* **134**, 139–149 (2001).
4. M.L. Dubernet, A. Grosjean, D. Flower, E. Roueff, F. Daniel, N. Moreau and B. Debray, "Ro-vibrational Collisional Excitation Database BASECOL – http://basecol.obspm.fr/"in *Proceedings Joint Meeting ITC14 and ICAMDATA 2004, Toki, Japan*; *J. Plasma Fusion Res. Ser.* **7**, 356–357 (2006).
5. M. L. Dubernet, F. Daniel, N. Moreau, A. M. Vasserot and S. Marinakis (to be submitted to *Astron. Astrophys.*)
6. H. S. P. Müller, F. Schlöder, J. Stutzki and G. Winnewisser, *J. Mol. Struct.* **742**, 215–27 (2005).
7. K. P. Dere, E. Landi, P. R. Young, G. Del Zanna, M. Landini and H. E. Mason, *Astron. Astrophys.* **498**, 915–29 (2009).
8. Mason NJ., "Electron Induced Processing; Applications and Data Needs" in *Proceedings of ICAM-DATA06, AIP Conference Proceedings* **901**, 74–84 (2007).
9. B. P. Schmitt, E. Volcke, O. Quirico, N. Brissaud, W. Fray, J.-M. Grundy, L. Bernard, D. Bonal, A. Baklouti, A. Pommerol, F. Trotta and N. Bonnefoy, "GhoSST: the Grenoble Astrophysics and Planetology Solid Spectroscopy and Thermodynamics database service: RELEVANT Database", 2009; see http://ghosst.obs.ujf-grenoble.fr/ .
10. L. S. Rothman, I. E. Gordon, A. Barbe, D. C. Benner, P. F. Bernath, M. Birk, et al., *J. Quantitat. Spectrosc. Radiat. Transfer* **110**, 533–572 (2009).
11. A. Y. Faenov, A. I. Magunov, T. A. Pikuz, I. Y. Skobelev, P. A. Loboda, N. N. Bakshayev, S. V. Gagarin, V. V. Komosko, K. S. Kuznetsov, S. A. Markelenkov, S. A. Petunin and V. V. Popova, "Spectr-W-3 online database on atomic properties of atoms and ions" in *Atomic and Molecular Data and Their Applications* edited by D. R. Schultz, P. S. Krstic and F. Ownby, AIP Conference Proceedings 636, American Institute of Physics, Melville, NY, 2002, pp. 253–262.

12. S. Mikhailenko, A. Barbe, Y. Babikov and V. G. Tyuterev, “S&MPO – a databank and information system for ozone spectroscopy on the WEB”, see http://smpo.iao.ru/ .
13. W. Cunto, C. Mendoza, F. Ochsenbein and C. Zeippen, *Astron. Astrophys.* **275**, L5–L8 (1993).
14. J. Woodall, M. Agundez, A. J. Markwick-Kemper and T. J. Millar, *Astron. Astrophys.* **466**, 1197–2003 (2007).
15. G. Malloci, C. Joblin and G. Mulas, *Chem. Phys.* **332**, 353–359 (2007).