Atomic Data Needs for X-ray Astronomy. 22 June 2000
http://heasarc.gsfc.nasa.gov/docs/heasarc/atomic/
M. A. Bautista, T. R. Kallman, and A. K. Pradhan, eds.

Atomic Databases

Claudio Mendoza
Centro de Física, Instituto Venezolano de Investigaciones Científicas (IVIC)
PO Box 21827, Caracas 1020A, Venezuela
E-mail: *claudio@taquion.ivic.ve*

Abstract. We describe some of the highlights in the development of atomic databases in the past five years and the ingredients that must be taken into account in order to ensure a successful venture. Also, important issues that have to be addressed in this growing field are briefly discussed.

1. Introduction

Atomic and molecular data are required in a variety of fields ranging from the traditional astronomy, atmospherics and fusion research to fast growing technologies such as lasers, lighting, low-temperature plasmas, plasma assisted etching and radiotherapy. In this context, there are some research groups, both theoretical and experimental, scattered round the world that attend to most of this data demand, but the implementation of atomic databases has grown independently out of sheer necessity. In some cases the latter has been associated with the data production process or with data centers involved in data collection and evaluation; but sometimes it has been the result of individual initiatives that have been quite successful. In any case, the development and maintenance of atomic databases call for a number of skills and an entrepreneurial spirit that are not usually associated with most physics researchers.

In the present report we present some of the highlights in this area in the past five years and discuss what we think are some of the main issues that have to be addressed.

2. Highlights

A summary of the highlights of the field of atomic databases in the last five years must start with the tremendous boost caused by the World Wide Web (WWW). The WWW is certainly leading to a paradigm change in the way scientific information and data are exchanged. Perhaps we are not yet fully aware of the extent of these changes as we just seem to be adjusting to the onset of a major information technology revolution. It is worth pointing out that atomic physics has had a portal (`http://plasma-gate.weizmann.ac.il/API.html`) since November 1994 which, in true web fashion, was the result of an individual initiative, namely by Yuri Ralchenko of the Weizmann Institute in Israel. This portal currently provides up-to-date information about conferences, databases, software, a who's who, jobs, links, resources, atomic servers by country, and includes an electronic preprint archive. Regarding atomic databases, it lists all active sites and classify them by type: numerical and bibliographic databases, data collections and data productions. Although the portal shows a table of data types, it does not display the sites where they are to be found which is useful for first-timers. A review of atomic data and databases on the Internet up to 1996 (`http://plasma-gate.weizmann.ac.il/~fnralch/app.dvi`) can be downloaded from its homepage. There are other portals of more specialized nature linking atomic databases such as *Databases of Atomic & Molecular Data for Astronomy* (`http://cfa-www.harvard.edu/amdata/ampdata/amdata.html`) at the Harvard–Smithsonian Center for Astrophysics.

Another important advance in the field has been the launching of the biennial International Conference on Atomic and Molecular Data and their Applications (ICAMDATA; `http://physics.nist.gov/Divisions/Div842/Icamdata/Homepage/icamdata.html`) that took place at NIST in September 1997, to be followed by a meeting at Keble College, Oxford, in March 2000. This series of conferences is addressing the main issues in the implementation and applications of atomic and molecular databases. It

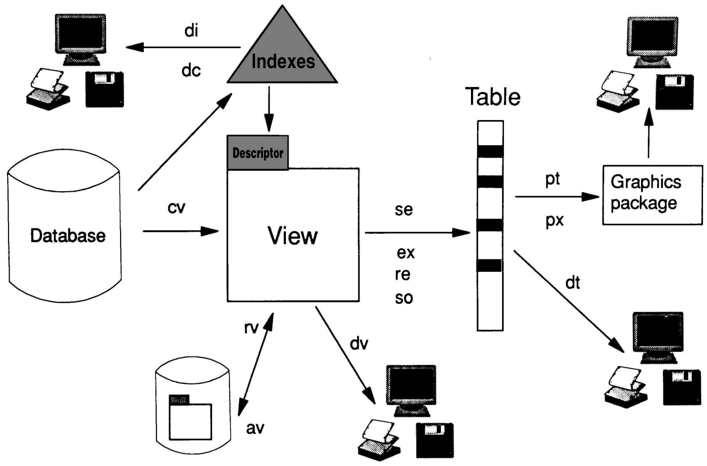

Figure 1. TOPbase database structure

aims at bringing together data producers and users, database developers, data center managers, administrators, and perhaps for the first time, prospective clients from industry (e.g. lighting, semiconductors). One of the key aspects discussed at the first meeting was the general lack of funding for data activities in government agencies. The conference proceedings include a comprehensive review on atomic collision databases and data services by Schultz (1998).

A further aspect that is bound to shape the evolution of atomic databases in the near future is a change in emphasis in high performance computing from number crunching to data intensive applications. This is due to the massive flood of experimental data that is swamping fields such as high energy physics and astronomy and to the huge commercial possibilities of bioinformatics. This situation is rapidly transforming the field of scientific databases from informal inhouse ventures to a new, fast growing, computer engineering discipline. As recently pointed out by Sincell (1999), the data volume and complexity are leading to a double paradigm shift in data storage and dissemination. On the one hand, the heavy data usage is suggesting new subtle ways of data organization and presorting, and on the other, it could transform the existing client–server network architecture to a hierarchical computer grid. In conclusion, it is envisioned that "data mining" will become a new and promising research mode in the near future.

3. Database Developments

When we were involved in the development of TOPbase (Cunto and Mendoza 1992; Cunto et al. 1993), the Opacity Project (The Opacity Project Team 1995) atomic database, a model was proposed for a database structure (see Figure 1). Data stored in secondary memory (disk) are retrieved as a result of a user query by at most two disk accesses through a properly structured set of indexes which can also be used as a table content. The retrieved data give rise to a subdatabase (a *view*) in main memory. The view can be displayed on the monitor, printed or stored as a disk file. Each view is identified by means of a *descriptor* that lists the search parameters that define it. Data in a view can be further massaged in main memory through the *table* structure to satisfy the user ultimate output requirements. The final table can then be displayed, printed, stored in a disk file or graphically processed.

This scheme was devised before the client–server architecture of the web became established and its transcription is not always smooth (refer to the TOPbase web version (`http://heasarc.gsfc.nasa.gov/topbase`). For instance, the index structure can be smartly exploited to allow, in addition to the query search mode, to a browsing search mode through hyperlinks; the *view descriptor* can be exported to a second web page to become an interactive *view manager*; graphic processing can be carried out efficiently at the client end through a Java applet. However, data massaging in main memory is very difficult to implement as the browser sockets are disconnected after every page access thus limiting interaction. This means that the user must specify precisely before each query the input and output requirements.

With this model in mind, we proceed to make quick comments on the current atomic database initiatives. The web site containing the *Atomic Spectral Line Database* (`http://cfa-www.harvard.edu/amdata/ampdata/kurucz23/sekur.html`) by Kurucz is a good example of a database that allows the user to specify detailed input/output requirements. The *Atomic and Molecular Database* (`http://amdata.nifs.ac.jp/`) at the National Institute of Fusion Science, Japan, makes extensive use of plots to display data, although it does not as yet allow interactive graphic processing at the client end. An interesting page is that of *Atomic Data for Astrophysics* (`http://www.pa.uky.edu/ ~verner/atom.html`) where a comprehensive database has been built up entirely from hyperlinks from different data reservoirs. Although the popular *CHIANTI* database (`http://wwwsolar.nrl.navy.mil/chianti.html`) cannot be accessed directly on the web, as its software and data must be both downloaded and installed locally, its success is perhaps due to the environment that it provides for data processing.

4. What makes a successful atomic database?

To answer this important question, we refer to our personal experience with the publication in 1983 of what has become a very successful paper (Mendoza 1983). containing tables of energy levels and radiative and collisional rates for nebula modeling. The actual success of this rudimentary database is not linked to its completeness (it was not extensive), accuracy (it listed the best numbers at the time) or regular updating. The tables have not been updated since the original publication, and in fact, independent attempts to do so have not managed to displace it which, after 17 years, is still regularly cited as an atomic data source of choice. After some consideration of this question, we have arrived at the following conclusions:

1. The development of an atomic database must address the needs of prospective users;
2. The availability of the database has to be timely;
3. The data have to become standard reference.

Item (1) suggests that database functionality must follow closely the way users go about in their work, the data must be labeled and in units and structures familiar to them; therefore development must involve user participation. Point (2) regarding timing is key, and can be illustrated with a a frequent situation in astronomy. Usually atomic database developments are demanded when a new detector becomes operational not taking into account that the time spent in computing the data and organizing the database may be as long as that taken to design and construct the detector, and is usually underfunded. As a consequence, users tend to gather, as needed, the best they can find and not bother when the more accurate values appear as the data search stage has already been completed. Relating item (3), it is often found that users prefer to share a common set of data that facilitates model comparison and the tracking of discrepancies; the emphasis is then correctly on the physics rather than on the data.

5. Current issues

As coordinator of an on-line atomic database service for almost ten years, we have found that there is a definite world-wide demand for accurate atomic data, but at the same time little scope for commercial ventures. This means that most projects will start as inhouse initiatives and financing would have to rely on the traditional funding agencies. Moreover, in spite of the present trends to promote liaisons with industry, and we would expect data activities to be one of them, there are at present scarce incentives for the latter thus limiting progress.

In 1996 copyright legislation to protect databases was introduced by the European Union, and a similar Antipiracy Act is being discussed in the USA. There has been an intensive debate (see, for instance, Gardner and Rosenbaum 1998; Maurer and Scotchmer 1999; Liedes 1997) on its actual benefits

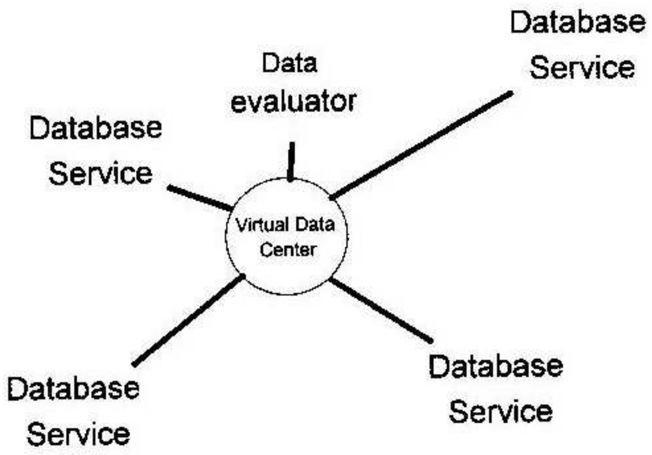

Figure 2. Proposed virtual data center

to the scientific community that has always relied on the open access of data. Proponents of the bill are the American Intellectual Property Law Association, the Information Industry Association and the American Association of Publishers. Critics include the AAAS, the National Academy of Sciences, the Association of Research Librarians, the Digital Future Coalition and the Electronic Freedom Foundation. We are of the opinion that the the Antipiracy Act would harm science by generally restricting data access and by discouraging the implementation of nonprofit databases such as the ones of our concern here.

We have found that there are definite advantages offering database services through a data center although the concept of the latter is rapidly evolving with the growth of the Internet. In Figure 2 we propose the structure of a "virtual data center" which is now envisioned, instead of a data gatherer, as a contractor of distributed database services developed and maintained by data producers. This proposal is in tune with the new web paradigm that information is to be accessed where it is produced. Data evaluation is then integrated as an on-going parallel process rather than a requisite for publication thus speeding up significantly the process of data dissemination. The best example of this concept is the A+M/PMI Data Centre Network (`http://www-amdis.iaea.org/dcn.htm`) that federates 15 national data centers for fusion applications coordinated by the International Atomic Energy Agency (IAEA).

Data centers must play a leading role in the specification of standards and guidelines on IT matters; for instance, the IAEA has established ALADDIN (A Labelled Atomic Data Interface) for the exchange and management of data. However, ALADDIN is implemented in the old fashion way by means of a set of routines rather than actual data structure specifications. Following the distributed model of Figure 2, what is required is the implementation of inter-service data loans or, further, seamless navigation across the databases such that a researcher seeking data in one system can automatically link to the others to complete the process. However, establishing standards can be slow and expensive. Thus, proposals have been mainly concerned with standardizing interfaces (e.g. ALADDIN) rather than the internal functions of the databases (Williams 1997; Pennisi 1999).

The issues concerning accuracy rankings, evaluation and recommendation are among the most difficult to solve. Regarding the accuracy of radiative data, say, an assignment must be given for each transition rather than a transition array, let alone an ion. For instance, in Figure 3 we show the $2s2p^2-2s^22p$ transition array in boron-like ions. The accuracy of the A-values for the $^2L_J-^2P^o_{J'}$ allowed transitions and the forbidden transition within the $^2P^o_J$ ground term is much higher than that for the sensitive $^4P_J-^2P^o_{J'}$ intercombination transition. For the case of the low members of the sequence, e.g. C II and N III, the latter shows both theoretically and experimentally a very broad scatter that

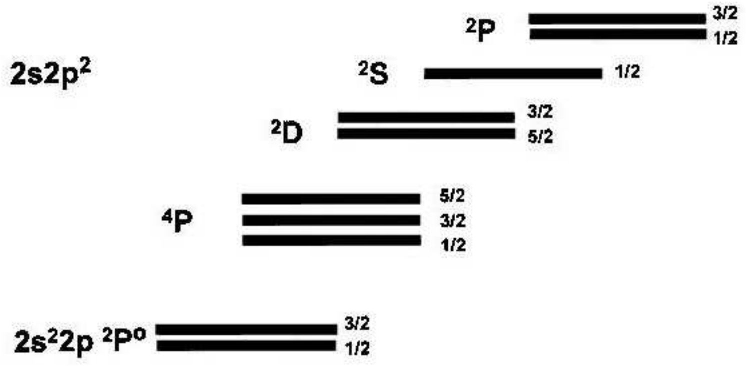

Figure 3. Level structure of the ground and first excited configurations of boron-like ions.

has only been recently sorted out by an experimental benchmark (Träbert et al. 1999, and references therein). This frequent situation makes data evaluation and recommendation a very slow process as every transition must be examined individually.

Finally, it must be stressed that data should always be documented by formal publication and the reference source listed in all database projects.

References

Cunto W, Mendoza C, 1992, Rev. Mexicana Astron. Astrofis. 23, 107

Cunto W, Mendoza C, Ochsenbein F, Zeippen CJ, 1993, Astron. Astrophys. 275, L5

Gardner W, Rosenbaum J, 1998, Science 281, 786

Liedes J, 1997, Science 276, 223

Maurer SM, Scotchmer S, 1999, Science 284, 1129

Mendoza C, 1983, in Planetary Nebulae, Flower DR (ed), Reidel, Dordrecht, 143

Pennisi E, 1999, Science 286, 449

Schultz DR, 1998, in Atomic and Molecular Data and Their Applications, Mohr PJ, Wiese WL (eds), AIP Conference Proceedings 434, American Institute of Physics, Woodbury, New York, p 119

Sincell M, 1999, Science 286, 1840.

The Opacity Project Team, 1995, The Opacity Project, Vol. 1, Institute of Physics Publications, Bristol

Träbert E, Gwinner G, Knystautas EJ, Tordoir X, Wolf A, 1999, J. Phys. B 32, L1

Williams N, 1997, Science 275, 301