UCRL-PROC-228275

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

HEND: A Database for High Energy Nuclear Data

David Brown, Ramona Vogt

February 23, 2007

Quark Matter 2006
Shanghai, China
November 14, 2006 through November 20, 2006

International Journal of Modern Physics E

HEND: A Database for High-Energy Nuclear Data

David Brown

Lawrence Livermore National Laboratory,
P.O. Box 808
Livermore, CA 94550, USA
brown170@llnl.gov

Ramona Vogt

Lawrence Berkeley National Laboratory,
1 Cyclotron Road,
Berkeley, CA 94720
USA
and
University of California, Davis
Davis, CA 95616
USA
RLVogt@lbl.gov

Received (received date)
Revised (revised date)

We propose to develop a high-energy heavy-ion experimental database and make it accessible to the scientific community through an on-line interface. The database will be searchable and cross-indexed with relevant publications, including published detector descriptions. It should eventually contain all published data from older heavy-ion programs such as the Bevalac, AGS, SPS and FNAL fixed-target programs, as well as published data from current programs at RHIC and new facilities at GSI (FAIR), KEK/Tsukuba and the LHC collider. This data includes all proton-proton, proton-nucleus to nucleus-nucleus collisions as well as other relevant systems and all measured observables. Such a database would have tremendous scientific payoff as it makes systematic studies easier and allows simpler benchmarking of theoretical models to a broad range of experiments. To enhance the utility of the database, we propose periodic data evaluations and topical reviews. These reviews would provide an alternative and impartial mechanism to resolve discrepancies between published data from rival experiments and between theory and experiment. Since this database will be a community resource, it requires the high-energy nuclear physics community's financial and manpower support.

1. Background and Potential Impact

We propose to create and maintain a database for high-energy nuclear physics. This central database will be web-accessible and searchable. As with Evaluated Nuclear Data File (ENDF/B) and EXchange FORmat (EXFOR) databases[1] and HEPDATA website[2], we will store cross sections, particle yields, and single particle spectra. We will also store data specific to higher energy reactions such as multi-particle spectra, flow and correlation observables. In short, we seek to archive whatever is needed to characterize a high-energy heavy-ion reaction. Initially we will compile published measurements but eventually we will cross-link the data with experiment descriptions. We also envision evaluating high-energy nuclear data and reporting these results in periodic topical reviews of subsets of the data. Several journals have expressed interest in publishing the topical reviews.

Such a database would organize existing data, allowing easier cross-experiment comparisons, theory benchmarking and development of systematics. In addition to basic science needs, there is a growing list of applications for high energy nuclear data. These applications include backgrounds in proton radiography; heavy-ion driven inertial confinement fusion; secondary ν and μ beam source development for MINOS and the International Linear Collider; and cosmic ray dose rates for space exploration. Most of these applications do not use the data directly but, instead, use evaluated representations of the data in application codes.

There is surprisingly no national or international effort to collect and maintain such a database. The US Nuclear Data Program (USNDP)[1] has compiled low-energy nuclear reaction data in the ENDF/B and other databases for decades. Similarly, the HEPDATA, Particle Data Group (PDG), arXiv.org and SLAC-SPIRES websites serve the high-energy particle physics community. The high-energy nuclear physics community is only partially served by these data sources. While most experiments make their published data available by, for example, posting published data tables on the collaboration website, this inevitably leads to a proliferation of data formats and web sites. Furthermore, after experiments end, their web servers may no longer be maintained. Thus, there is a very real risk that the data could be lost, as has, in some cases, already occured. Given the data volumes generated by experiments at the RHIC and future experiments at the LHC, GSI and elsewhere, this oversight should be rectified.

We first explain that the proposed database should be a community effort. Next, we motivate the need for data evaluation and topical reviews and provide some technical details. Finally, we conclude with the status of the database proposal.

2. Database Management Philosophy

Since this database is intended to be a community resource, we propose a community driven management model such as the arXiv.org preprint server: the "consumers" of arXiv.org are also its "suppliers." Physicists submit their preprints to arXiv.org to better disseminate their latest results. They browse arXiv.org because they know others are submitting their latest results there. In this way, data collection is farmed out to the data producers – a tactic we wish to employ.

The proposed database would, however, differ from arXiv.org in two key respects. First, the proposed database could make available not only published data but also auxiliary or supporting data sets that may be too large for publications such as Physical Review Letters. Second, in order to assure that the data is of the highest quality, we would like database submission to be linked to the journals' peer-review process. Ideally, the experiments would be able to submit both the published and auxiliary supporting data to the database when submitting papers for publication through *e.g.* links on the journal submission pages. Preliminary discussions indicate that the journal editors would cooperate in this endeavor.

For experimental collaborations to support this project with more than abstract enthusiasm, they must play an active role in steering the database development and be given a financial stake in the eventual product. We thus propose holding annual workshops to guide development, discuss the topical reviews and propose new subjects for review.

3. Evaluations and Topical Reviews

What makes our database different from other available data repositories such as collaboration web sites or HEPDATA is data evaluation. Evaluated data provides our "best guess" representation of a particular observable. An evaluation is often a combination of model calculations, systematics and fits to raw experimental data. However the evaluation is produced, it needs to be checked against existing experimental data and should be published in a peer-reviewed journal. Such a review can also provide an alternative and impartial mechanism for resolving discrepancies between published data from rival experiments and between theory and experiment.

As an example, consider charm production in heavy-ion collisions. Charm measurements not only can reveal important information about the early state of the system through measurements of their azimuthal angle correlations and transverse momentum spectra in proton-proton and nucleus-nucleus collisions but also provide an important baseline for J/ψ production in heavy-ion collisions. Charm hadrons, particularly D mesons, have been measured in a number of experiments over a wide energy range, from fixed-target experiments reporting total cross sections to high transverse momentum spectra from the $\overline{p}p$ collider at the Tevatron. The extrapolation of limited acceptance measurements to total charm cross sections vary considerably, even at the same energy, presumably since many of these measurements were taken before next-to-leading order perturbative QCD calculations became available. Early measurements were also hampered by small statistical samples. To determine the consistency of these data with each other and with new measurements at heavy-ion colliders, these data must be re-evaluated with all previous assumptions re-examined. It would, in fact, be useful to have such an evaluation already in hand, with comparison to next-to-leading order calculations since the heavy flavor data from $\sqrt{s_{NN}} = 200$ GeV pp, d+Au and Au+Au collisions at RHIC agree in shape but not in magnitude. Neither agree well with QCD predictions. While some of the data are reconstructed D^0's, the bulk are electrons from semileptonic decays of D and B mesons. The interpretation of the nucleus-nucleus data relies heavily on the accuracy of the D^0 measurements.

4. Technical Details

A central, yet often neglected, aspect of data archives is the technical details of the data storage format. The nuclear data community traditionally suffers from a multitude of relatively obscure data formats. For example, the format still used in the ENDF/B database was designed specifically to accommodate the limitations of now-obsolete punch cards. In some cases, the task of writing translation and visualization tools for these data sets requires a large, dedicated effort.

Given the importance of using a transparent and well-supported format, we have decided on XML (eXtensible Markup Language) as our data storage format. Documents stored in XML can be self-describing so that, with minimal effort, scientists/users 30 years or more from now can interpret the documents' contents. Furthermore, XML documents are represented by computationally convenient tree structures rather than the simple strings typically used to store nuclear data. XML is a mature technology with the support of thousands of programmers and web developers and is extensively supported by most common programming languages. Lastly, the many tools needed for web-based access and manipulation of XML databases have reached a state of maturity.

We are in the process of evaluating which tools to store the data and serve it to the web. The database itself will most likely be in a native XML database such as the Berkeley XML Database rather than a relational database. While relational databases have a performance advantage over a native XML database, they use a data paradigm that is quite different than the structure-based approach we are pursuing. The web application itself will be one of many Java web application frameworks that include some kind of templating engine. We wish to use a Java-based XML Schema to source code generator such as JAXB or Castor to translate the XML data into Java objects. Once we have Java objects, we can use the template engines to generate query and data displays on the fly with relatively simple coding. This scheme will also be used to simplify the communication with the Nuclear and Atomic Data System plotting tool[6].

5. Current Status

Elements of this proposed project are being developed for other uses, such as the XML data format, XEndl, and the Nuclear and Atomic Data System plotting tool[6]. The XEndl format is under active development and an outline of the draft format is published[7]. This format in use by a Livermore-Los Alamos collaboration seeking to develop restricted form of format for use on processed data libraries in transport codes such as Geant4, Mercury or MCNPX. The data access routines are being rewritten in C/C++ and Python and the processing codes written in Python.

We have submitted a white paper describing our proposal to the DOE-OS Heavy-Ion and Nuclear Theory programs and are circulating it in the STAR and PHENIX collaborations. Copies of the whitepaper are available upon request from the authors. Since we want this database to be a community resource, we strongly encourage members of the heavy-ion community to contact us with their questions, comments, wishes and ideas.

Acknowledgments

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48. R.V. was supported in part by the Director, Office of Energy Research, Office of High Energy and Nuclear Physics, Nuclear Physics Division of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

1. National Nuclear Data Center website *http://www.nndc.bnl.gov/index.jsp*.
2. HEPDATA website *http://durpdg.dur.ac.uk/HEPDATA/*.
3. Particle Data Group website *http://pdg.lbl.gov/*.
4. SLAC-SPIRES website *http://www.slac.stanford.edu/spires/*.
5. arXiv.org preprint server *http://arXiv.org/*.
6. LLNL Computational Nuclear Physics Group's Nuclear and Atomic Database System *http://nuclear.llnl.gov/CNP/nads/NADSApplet.html*.
7. J. Pruet, D.A. Brown, B. Beck, D. McNabb, "Alternative approach to nuclear data representation," *Nucl. Inst. Meth.* B **244** (2006) 473.