

The CEDAR Project

J. M. Butterworth, S. Butterworth and B. M. Waugh, UCL, London, UK
W.J. Stirling and M.R. Whalley*, IPPP, Durham, UK

Abstract

We describe the plans and objectives of the CEDAR project (Combined e-Science Data Analysis Resource for High Energy Physics) newly funded by the PPARC e-Science programme in the UK. CEDAR will combine the strengths of the well established and widely used HEPDATA database of HEP data and the innovative JetWeb data/Monte Carlo comparison facility, built on the HZTOOL package, and will exploit developing grid technology. The current status and future plans of both of these individual sub-projects within the CEDAR framework are described, showing how they will cohesively provide (a) an extensive archive of Reaction Data, (b) validation and tuning of Monte Carlo programs against these reaction data sets, and (c) a validated code repository for a wide range of HEP code such as parton distribution functions and other calculation codes used by particle physicists. Once established it is envisaged CEDAR will become an important Grid tool used by LHC experimentalists in their analyses and may well serve as a model in other branches of science where there is a need to compare data and complex simulations.

THE PHYSICS PROBLEM

Particle physics experiments at high-energy accelerators provide a wealth of data on the final state in electron-positron, lepton-proton and proton-(anti)proton interactions. These data represent a triumph for the Standard Model, particularly in precision electroweak measurements and the verification of the QCD sector to an impressive degree of precision.

Despite these successes, several aspects of high-energy collisions are still poorly understood, often due to technical difficulties in the calculation of non-perturbative or complex perturbative effects. Such lack of understanding can be a limiting factor in the accuracy of new measurements. Examples with particular relevance to the LHC include parton distribution functions (PDFs) in hadrons, hadronisation in the final state, multijet production and “underlying events”. Fig. 1 illustrates some of these processes in a complex high energy event. All these processes are calculated and/or modelled in Monte Carlo simulation and calculation programs that employ fits to existing data. Consistent tuning of the free parameters of these models, and confirmation of the physics assumptions they contain, is a non-trivial matter since the measurements are made with a vari-

ety of colliding beams, in many different regions of phase space, and for many complex observables.

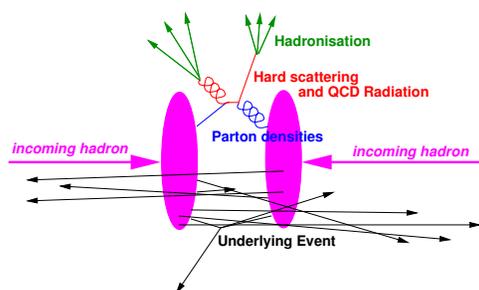


Figure 1: The complexity of a collision process.

THE SOLUTION – CEDAR

A solution to the above problem will be provided by CEDAR, a project newly funded by the PPARC e-Science programme in the UK, which will construct a resource for particle physics enabling the predictions of Monte Carlos, and other calculation programs, to be easily compared with real data. CEDAR will allow the parameters of the models to be varied and to be simultaneously compared to as wide a range of data distributions as is necessary to maintain global consistency. These global comparisons are vitally important as it is quite possible to obtain a good fit to a new set of data but at the same time lose the quality of fit to other existing data.

CEDAR will combine the strengths of the established and widely used HEPDATA database [1] of high-energy physics data and the innovative JetWeb data comparison facility [2], and will exploit developing Grid technology.

In short:-

$$CEDAR = JetWeb + HEPDATA + more...$$

The individual parts, including the “...more” are described below.

JetWeb

JetWeb is a WWW interface, developed at UCL, which provides a facility for direct comparison between the predictions of Monte Carlo programs and measured physics distributions from experiments. It is based upon the HZTOOL program [3] which generates the physics distributions from a given Monte Carlo program by calculating the

* M.R. Whalley@durham.ac.uk

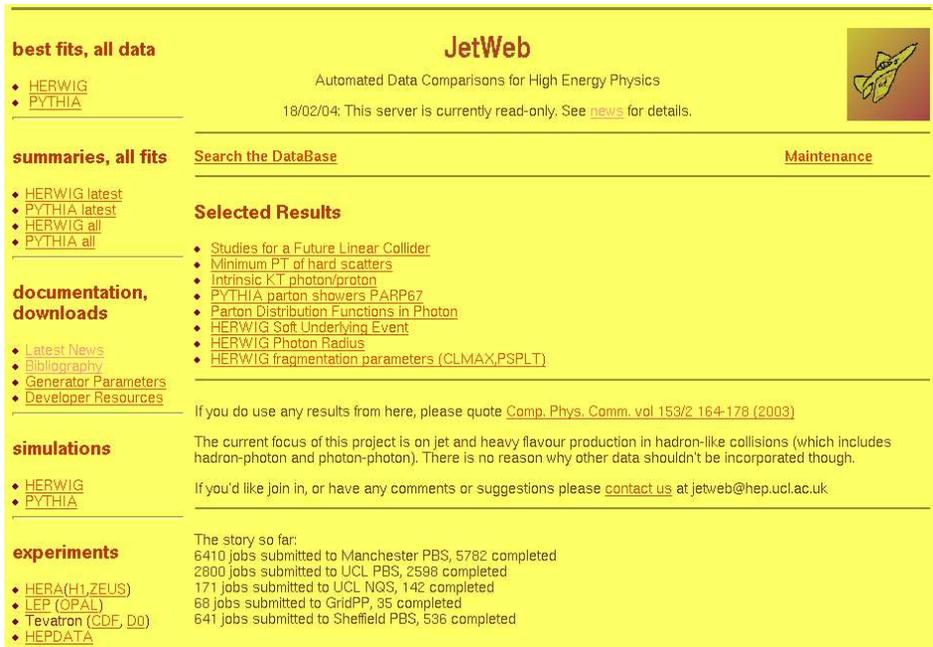


Figure 2: The JetWeb home web page

relevant variables and applying the cuts used in the published measurements.

The JetWeb interface allows the user to request and display quantitative comparisons of chosen Monte Carlo models with any number of measured data distributions. In this way the compatibility with the older data sets can be maintained when tuning to new data sets.

Since the generation of Monte Carlo events to calculate the predicted distributions to the required accuracy is a very compute intensive operation, JetWeb maintains a database (MySQL) of Monte Carlo generated data which can queried, or added to at user request, thus increasing the available statistics. JetWeb is already “grid-enabled” and has submitted its Monte Carlo jobs to GridPP with successful outcomes, as well as directly using the computer systems at UCL, Manchester and Sheffield.

Fig. 2 shows the JetWeb WWW home page which provides the user with the starting point to begin a new fit, or to search the JetWeb database for results from previous fits.

Fig. 3 show the main components of the JetWeb server. Interaction with the user is via two Java Servlets, running in a Tomcat servlet container linked to an Apache web server: the Searcher servlet provides the the standard functions available to a general user; the Maintainer provides additional functions, such as adding new data, to the JetWeb administrators.

At the heart of JetWeb is the Java Object Model (JOM) which contains the properties and interactions of the Model, Papers, Plots and Fits. (In this context a Model means a unique generator version and set of parameters.) Data is converted between formats via this model, e.g. from Database to HTML display. The JOM interacts with the JetWeb database, a MySQL database which stores exper-

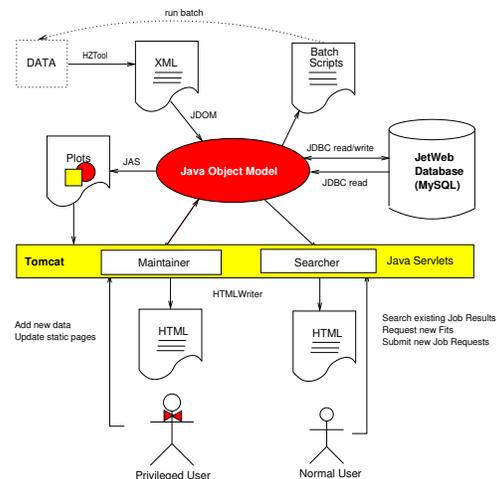


Figure 3: The various components of JetWeb

imental data as well as predictions from various models. The predictions are generated by running the Monte Carlo generators within the HZTOOL package, and the results are available to the user who requested the run and to any future user who makes a similar request.

At present, the size of the JetWeb database is restricted by the relatively small proportion of experimental data distributions which have HZTOOL routines to generate the predicted distributions from Monte Carlos generators.

CEDAR will do the following for JetWeb:-

- Re-design HZTOOL in OO structured C++ for long term maintenance and development. This is vitally important as new HEP code, and in particular new Monte Carlos, are being written using C++ in such

tion stored within that one unit (a data record). While this DBMS has proved very resilient and stable over the years, it is clearly not suitable for the present purpose of directly interfacing with JetWeb, or with any other resource over the Internet/Grid.

CEDAR will do the following for HEPDATA:-

- Migrate the data from the HEPDATA BDMS hierarchical database to a MySQL relational database. Not only will this provide the necessary means to interface with JetWeb, and other resources if necessary, but it will also address the long term future needs of HEPDATA by using a DBMS which is more standard and maintainable than its present one.
- Integrate the new database into JetWeb as the source of its “real” data. This will expand the number of available data sets to which JetWeb has access.
- Make the new database available on the Internet/Grid as a networked database for general use. This will involve implementing and expanding on the existing functionality of the HEPDATA web search and display methods. Expansions envisaged include direct access to the data through the conventional networks and also via the “Grid”.
- Develop new methods of direct entry and validation of data by the experiments, thus making them in control of their own data. At present the entry and verification of data in the database is generally instigated and controlled by the HEPDATA personnel. In future CEDAR will seek methods and formats (e.g. XML) in which experiments can enter and maintain their own data in the database potentially using Grid technologies and access validation methods. It should also be noted that the development of an XML document format for transfer of particle physics data will be useful also for data output to the Grid etc. as well as input to HEPDATA.

more... HEPCODE

As well as the improvements to JetWeb and HEPDATA and their integration as discussed in the previous sections, it is also the intention of CEDAR to provide access to current and validated versions of various HEP theory and experiment software used at present, and in the future LHC era. A partial list of such software includes JETRAD, DYRAD, EXCALIBUR, (DI)PHOX, ZFITTER, RACONWW and MADGRAPH. No central repository of such codes currently exists in an organised and consistent way at present and as such it will be of great benefit to the community.

A prototype version, HEPCODE, has been set up on the IPPP website (<http://www.ipp.dur.ac.uk/hepcode/>) at Durham which includes, as well as a table of available programs and links, a web form for submission of new codes to the repository.

SUMMARY

In summary we list the various tasks the CEDAR project will be engaged upon:-

- Convert HEPDATA to a relational MySQL database.
- Modify JetWeb to access data direct from HEPDATA.
- Design and implement an OO replacement for HZTOOL.
- Extend the number of HZTOOL routines/models available in JetWeb.
- Include new generation C++ Monte Carlos.
- Develop CEDAR grid tools for automatic validation and data access.
- Develop standard document formats (e.g. XML) for experimental physics data and results, which can be used for example as methods for reaction data export/import from experiments, as well as being useful in a more general context.
- Develop a code repository for a wide range of HEP codes and integrate with the validation centre.
- Incorporate LHC and, eventually, ILC data.

Once established and running with all the latest codes and data sets, it is envisaged that CEDAR will become an important tool used by LHC experimentalists in their analyses. Other branches of science which have need to compare data and complex model situations may also find the techniques and tools used in CEDAR of benefit to their work.

THE CEDAR TEAM

At UCL:-

Jon Butterworth - J.Butterworth@ucl.ac.uk
Susanna Butterworth - sb@hep.ucl.ac.uk
Ben Waugh - waugh@hep.ucl.ac.uk

At Durham:-

James Stirling - W.J.Stirling@durham.ac.uk
Mike Whalley - M.R.Whalley@durham.ac.uk
A N Other - to be appointed from 1st April 2005

Web Links:-

CEDAR - <http://cedar.ac.uk/>
JetWeb - <http://jetweb.hep.ucl.ac.uk/>
HEPDATA - <http://durpdg.dur.ac.uk/hepdata/>
HZTOOL - <http://hztool.hep.ucl.ac.uk/>

ACKNOWLEDGEMENTS

The CEDAR project is funded by the PPARC e-Science grant PP/B5027/44/1. HEPDATA is funded by PPARC grant PP/B500590/1.

REFERENCES

- [1] <http://durpdg.dur.ac.uk/hepdata/>
- [2] J.M. Butterworth and S. Butterworth, Comp. Phys. Comm. vol 153/2 (2003) 164–178
- [3] J. Bromley et al., Hamburg 1995/96, Future physics at HERA, vol 1, 611–612