JetWeb and HepData

Global tuning of MC event generators

Andy Buckley

Institute for Particle Physics Phenomenology
Durham University, UK

CHEP06, Mumbai 2006-02-13

Durham University

Outline

Durham University

CEDAR: Collaborative e-Science Data Analysis Resource

JetWeb and HepData are components of CEDAR:

- JetWeb: global tuning of Monte Carlo generator parameters
- HepData: archival of published experimental data
- HepForge: development environment for HEP software
- HepCode: centralised repository of pheno code/programs
- HepML: set of XML data formats for data sets and MC config

www.cedar.ac.uk

Durham University

Event generators: motivations

Physics motivations of event generators:

- Provide bg estimates for new physics searches
- Simulating new physics processes
- Detector calibration
- Designing new experimental facilities

Event generators: necessary inputs
(no pain, no gain)

But some inputs are needed:

- Matrix elements for hard process
- Perturbative parton showers (matched to MEs)
- Phenomenological models for non-perturbative physics

These last two contain parameters which must be tuned to experimental data.

Durham University

Event generators: tuneable parameters

- Parton density functions (PDFs)
- Matrix elements (kinematics, phase space, coupling/scale)
- Parton showers (and matching to MEs)
- Threshold effects in heavy flavour production
- Fragmentation functions
- Hadronisation (string params., multiplicities, strangeness production. . .)

Large number of params: nonsensical tunings dominate parameter space

Durham University

Data for event generator tuning

Useful constraints from $\sim$ all data:

- Tevatron: jets, photons, EW, charm, beauty
- HERA: DIS inclusive, photo-production, jets, c, b
- Fixed target DIS
- LEP: annihilation, $\gamma\ \mathrm{e}^-$, $\gamma\ \gamma$
- SPS: min bias, inclusive hadrons

Durham University

Caveats

There are always caveats, especially in high dimensional problems!

- *Global* tuning to data required — selective tunings are dangerous
- Expect *sets* of reasonable configurations, rather than one "ultimate" tuning
- Any tuning depends on your definition of "best". Both in terms of what phenomenological features a given expt. considers important and in terms of the fitting measure used.

Durham University

Introducing JetWeb

JetWeb is a centralised system for global MC tunings

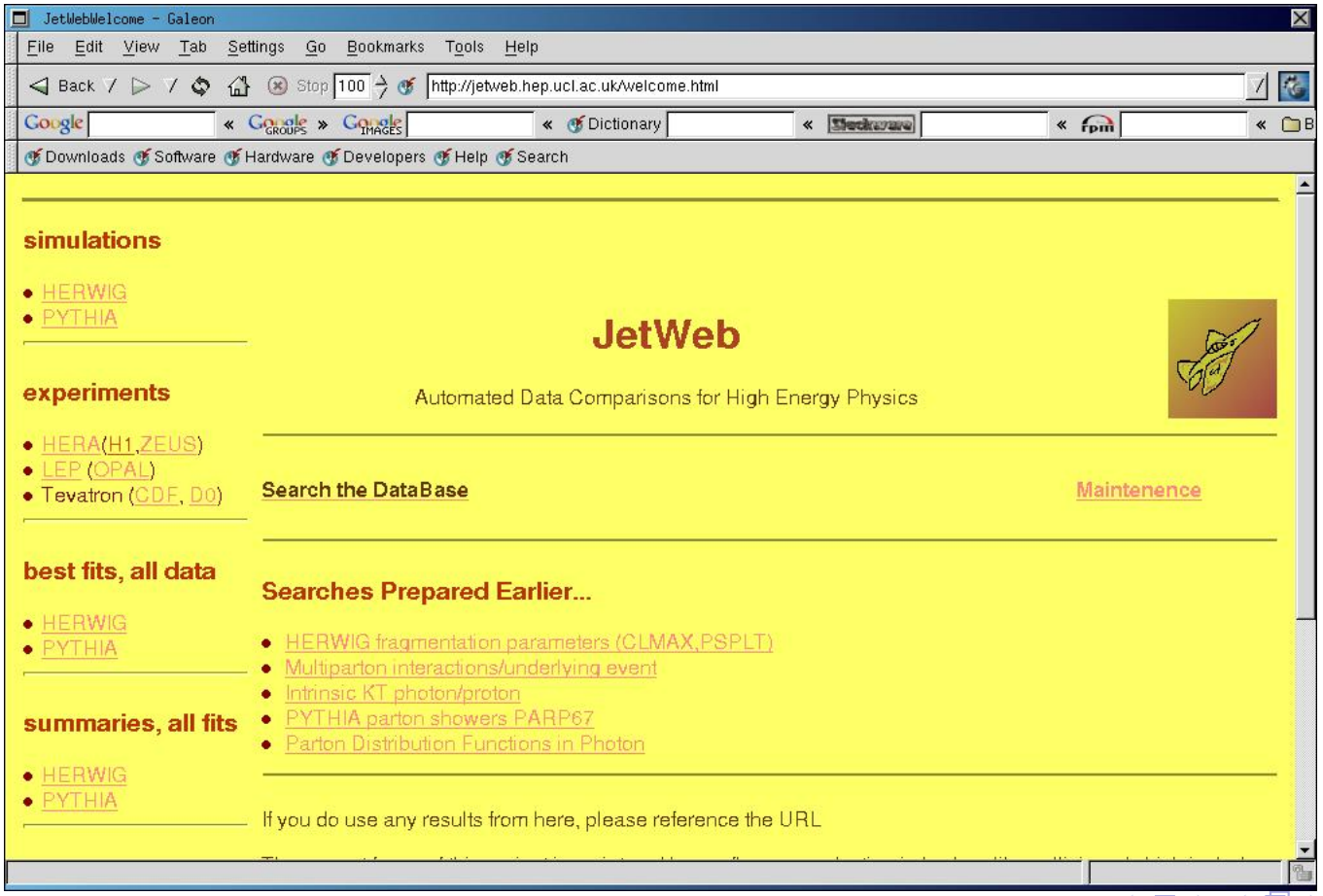
- Written in Java, run on Apache+Tomcat framework
- Database of global data fit qualities for MC models (generator & params)
- Web interface to view fit qualities and plots
- Web interface to request generation of a particular model if not (sufficient) in db
- Comparisons via a set of routines, each corresponding to a published paper

Durham University

JetWeb in action (1)

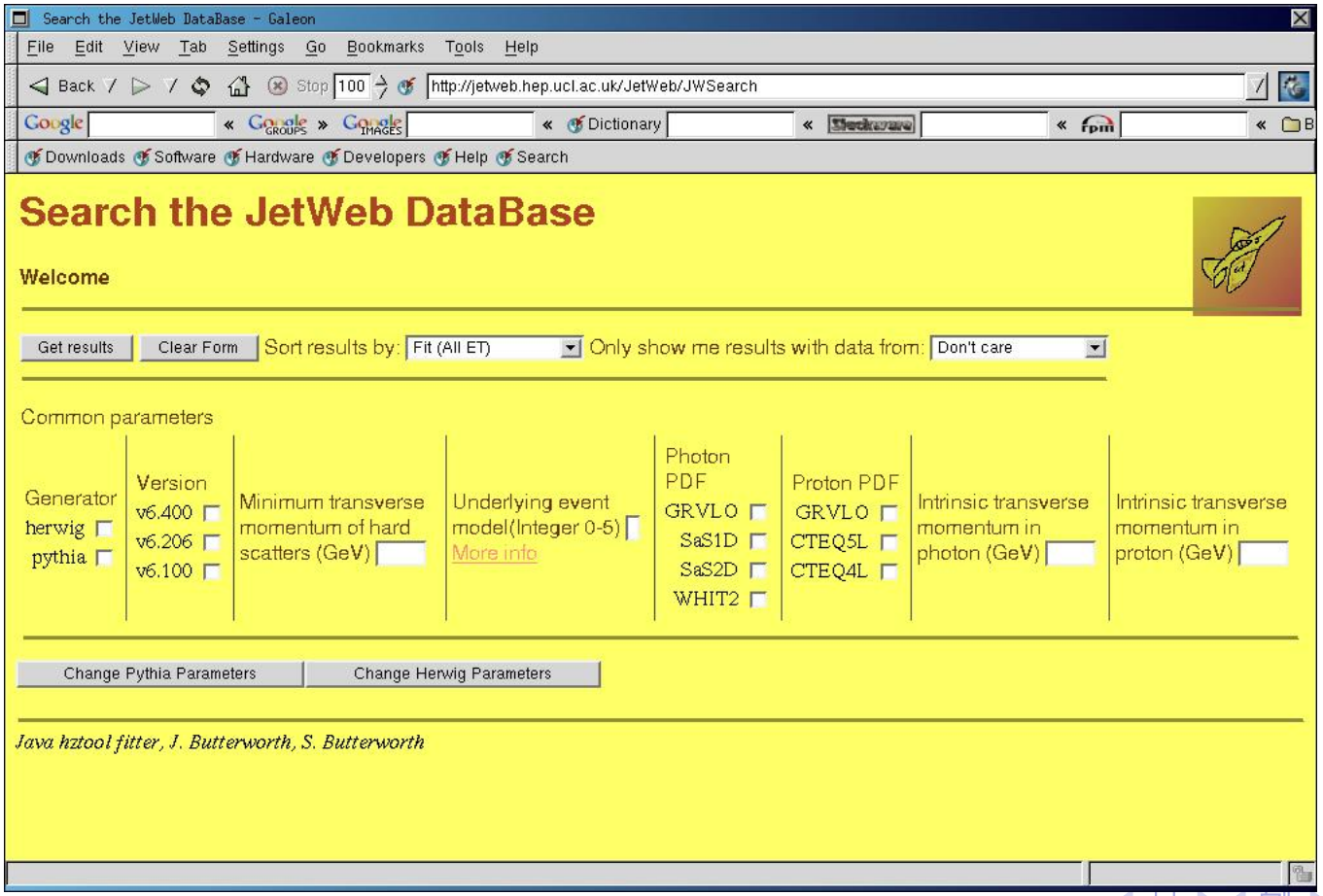
Welcome screen

NB. This is the *frozen* copy of JetWeb. . .



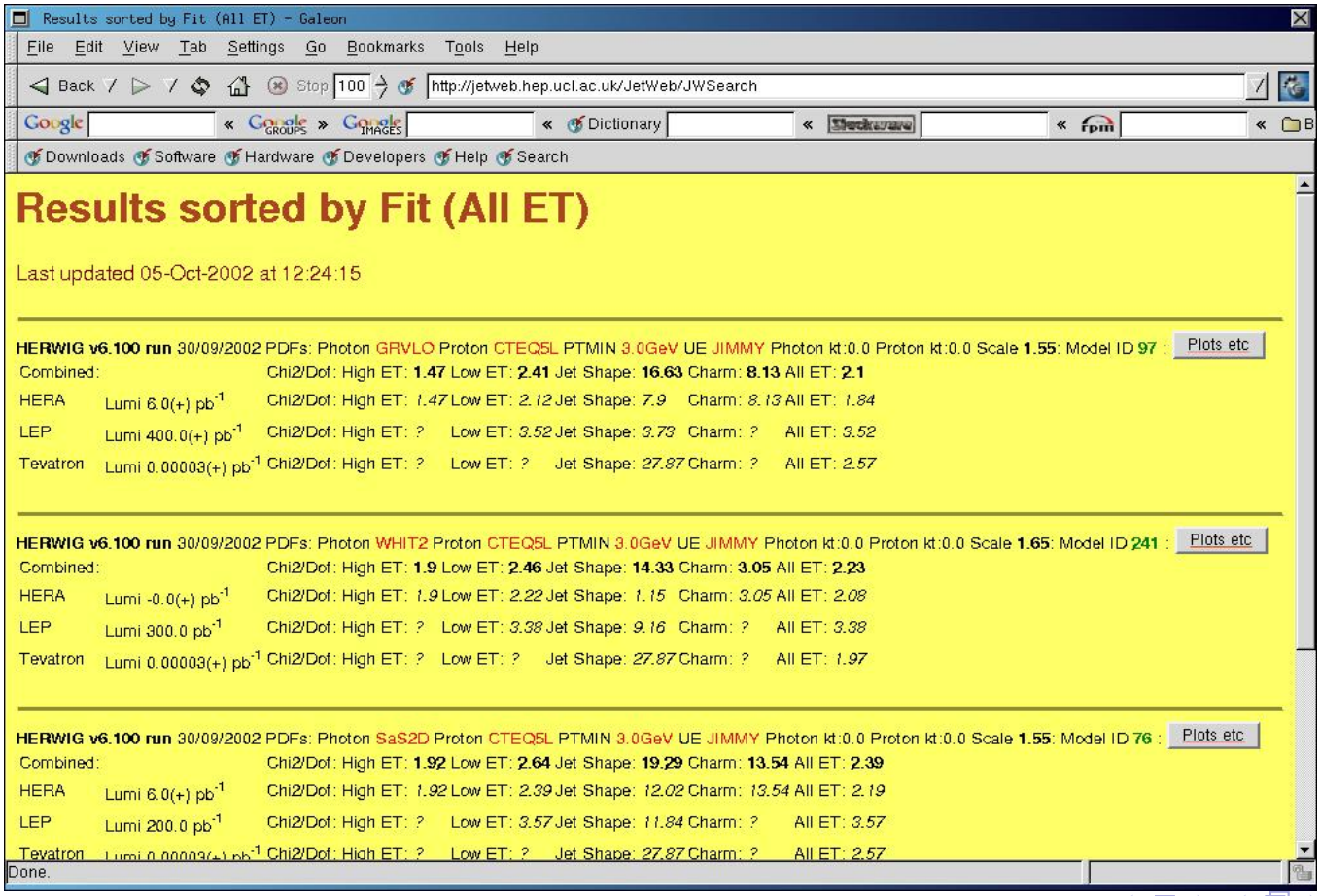
JetWeb in action (2)
Searching

NB. This is the *frozen* copy of JetWeb. . .



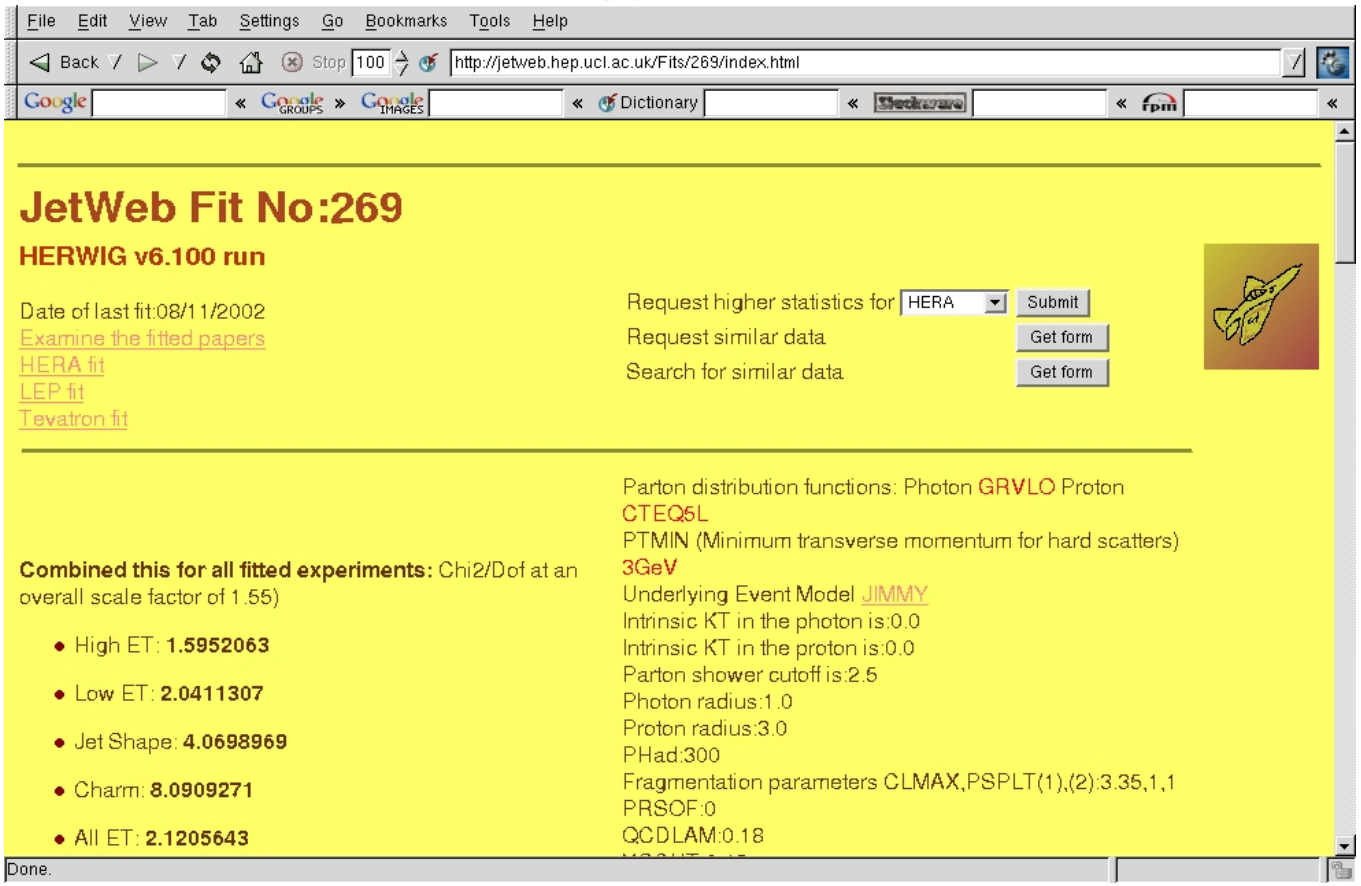
JetWeb in action (3)
Fit list

NB. This is the *frozen* copy of JetWeb. . .



JetWeb in action (4)
Fit details

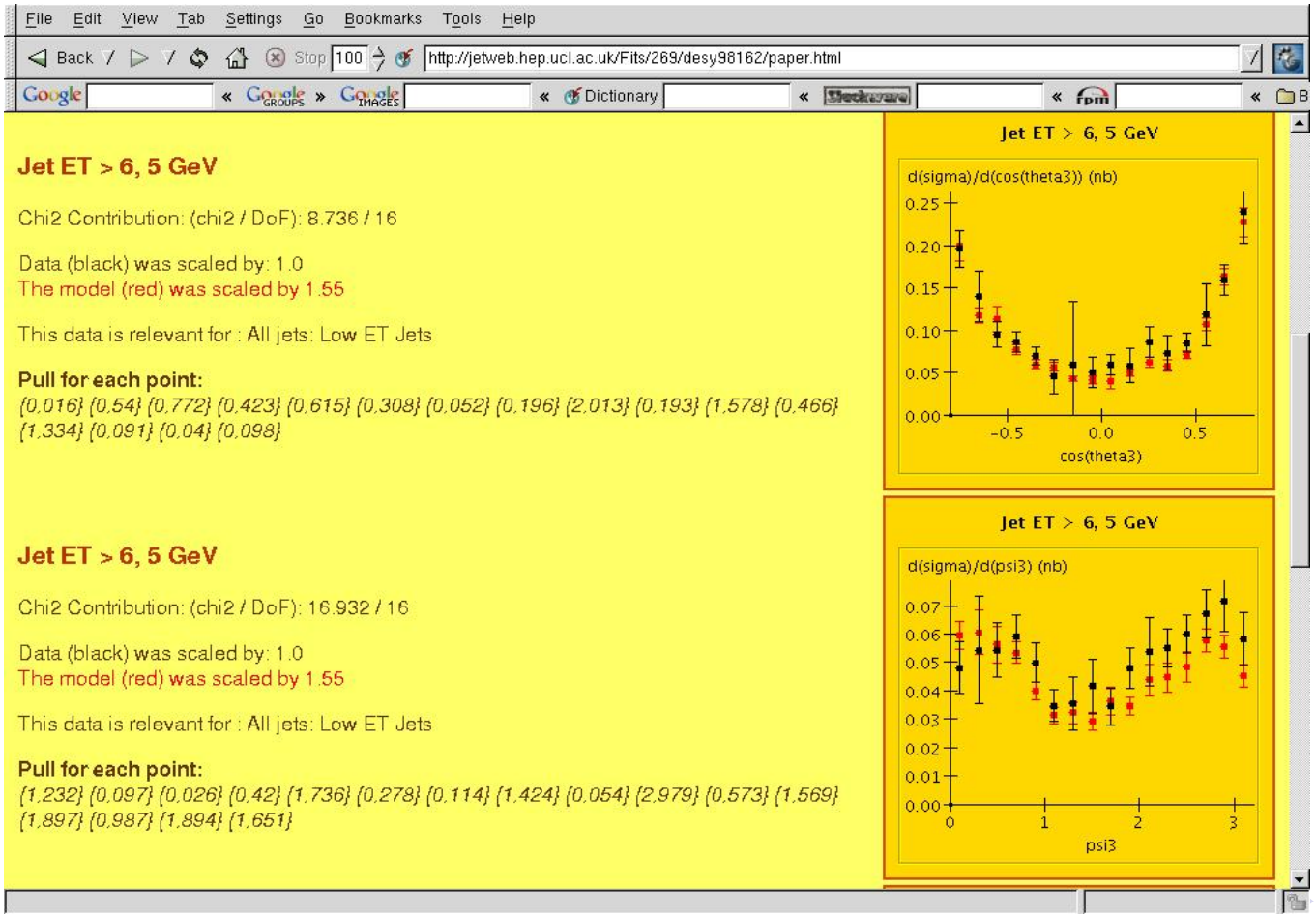
NB. This is the *frozen* copy of JetWeb. . .



JetWeb in action (5)

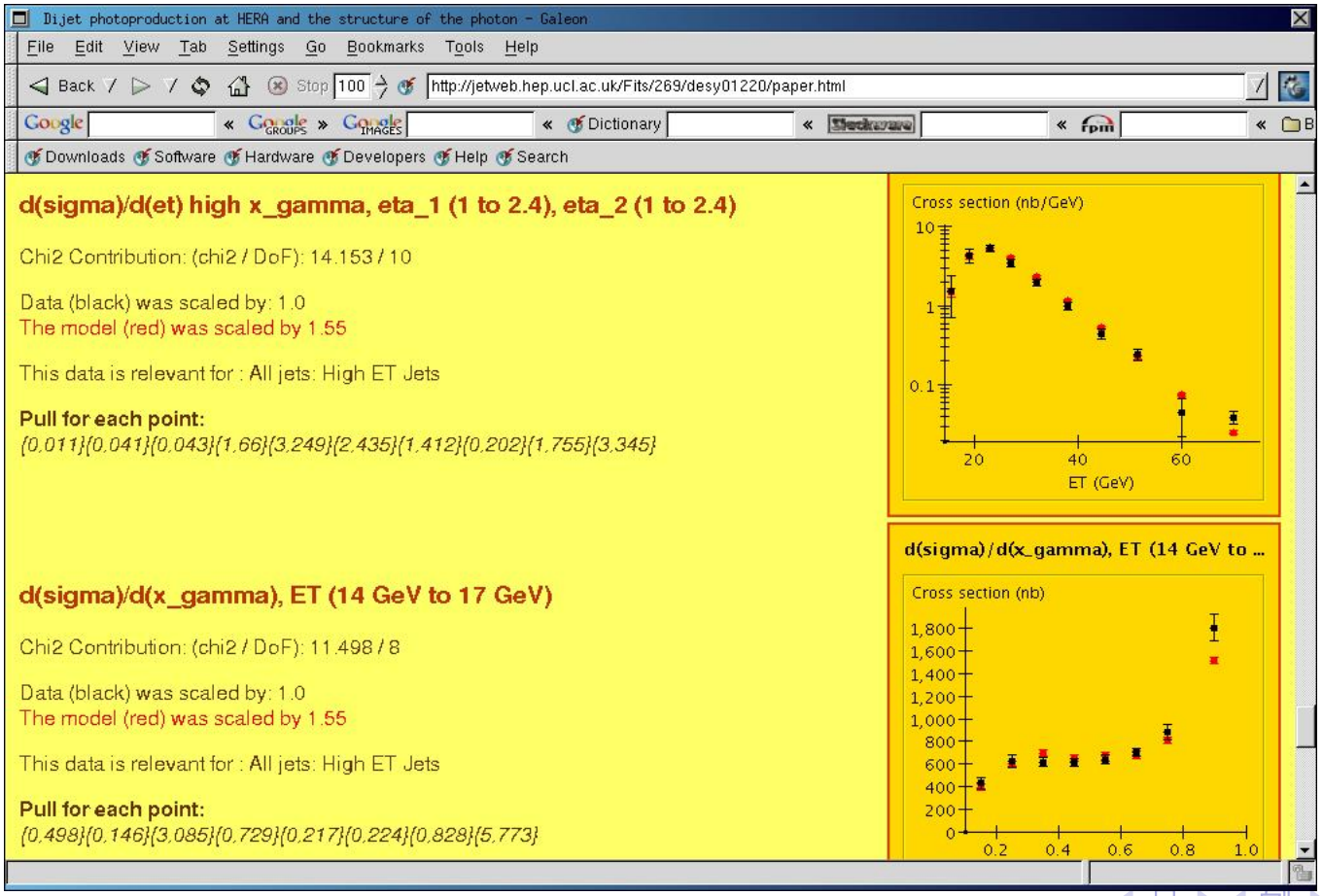
Paper view (1)

NB. This is the *frozen* copy of JetWeb. . .



JetWeb in action (6)
Paper view (2)

NB. This is the *frozen* copy of JetWeb. . .



HzTool

JetWeb is only the front-end, distribution and comparison system. Reproduction of experimental analyses done by HzTool

- HzTool is a library of Fortran routines for reproducing observables
- Typically one HzTool routine per paper
- Typically written by paper authors
- Generator independent (at least in principle!)
- Due to HzTool's background, most routines are from HERA (at present)

JetWeb now accesses HzTool via a 'HzSteer' package into which most (all?) generator dependencies have been moved.

Durham University

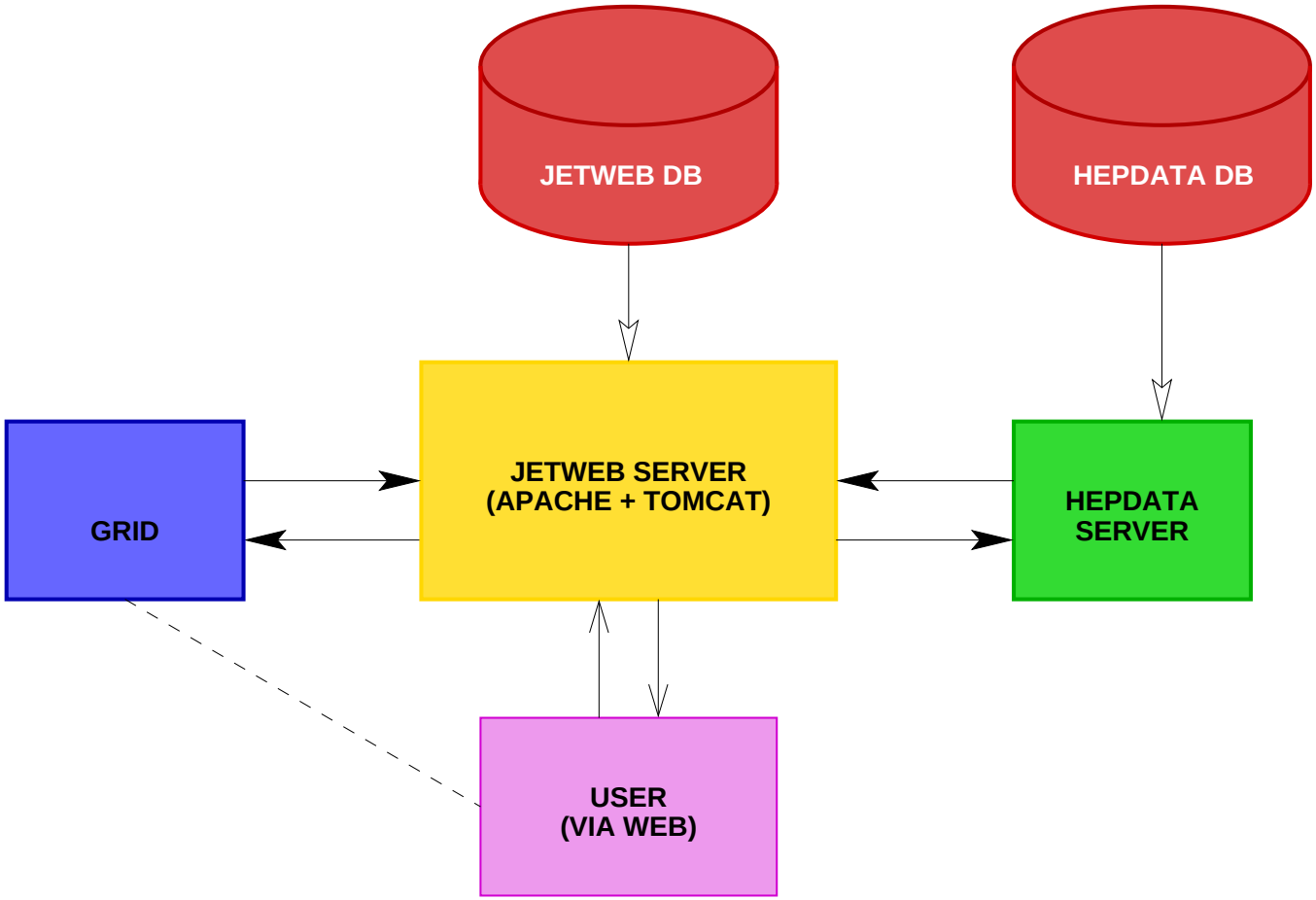
JetWeb enhancements

CEDAR is re-writing / enhancing JetWeb in several ways:

- Removal of experimental data from JetWeb database: now connecting directly to new HepData system. Single source for scattering data
- Framework re-factor to make it easier to add new generators/modules (the ‘build one to throw away — you will anyway’ rule in action!)
- Plans to use Grid authentication and job distribution
- New histogramming modules using the AIDA interfaces
- (Scanning parameter space, re-weighted fits. . .)

Durham University

(New) JetWeb behind the scenes



Rivet

- Rivet is a C++ re-implementation of HzTool
- Based on concept of ‘projections’: maps from final state event record to observables
- Both a collection of tools (jet algorithms etc.) and a library of analyses
- Like HzTool, an external steering package will be used to pass params to generators
- Ben Waugh will talk about HzTool and Rivet in detail on Thursday

Developers: AB, Jon Butterworth, Leif Lonnblad, Ben Waugh, James Monk

Durham University

Introducing (legacy) HepData

- Searchable ∼30 years' worth of scattering data, PDFs etc.
- Historically stored in hierarchical BDMS database: very inflexible and little modern support
- Data available as text files, PAW kumacs or GIF images
- Legacy db stores pretty much everything as a string
- Adding records requires writing a Fortran routine
- Mirrors to SPIRES, PDG info

Durham University

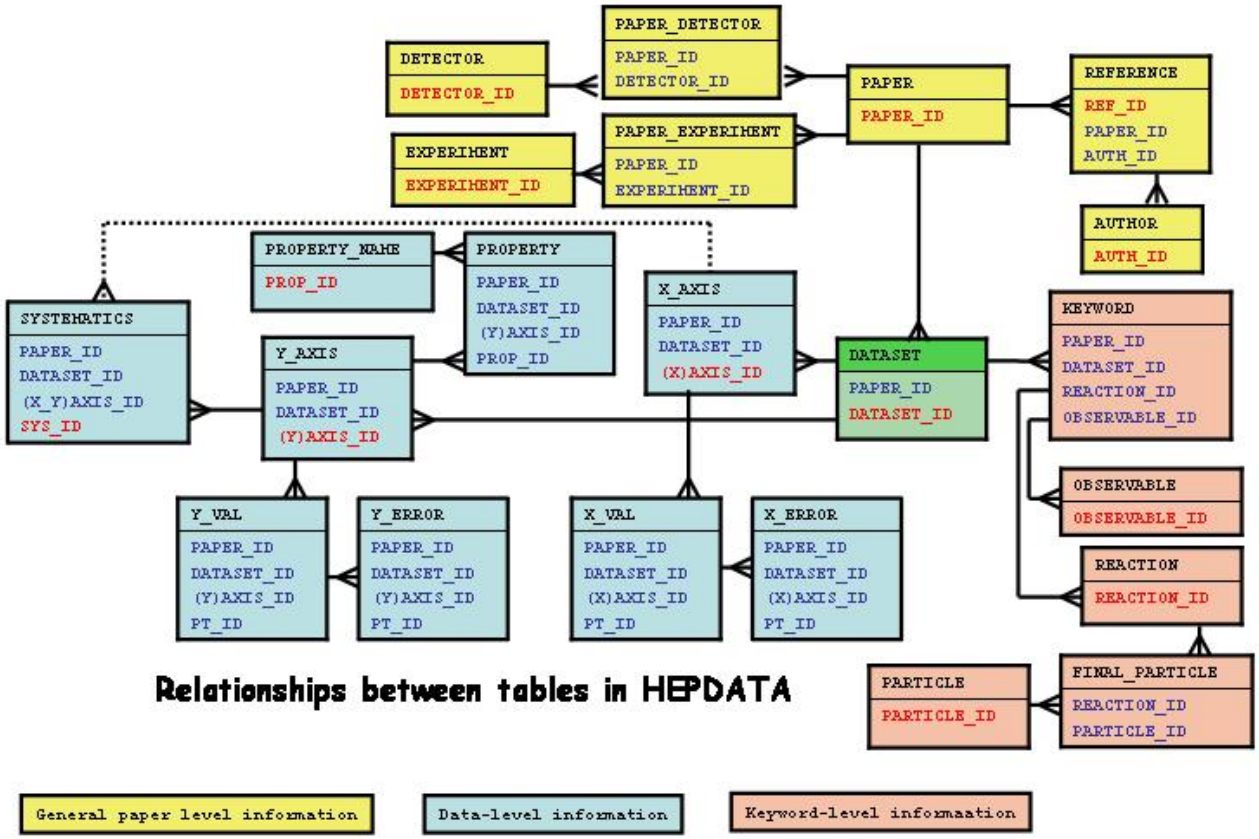
HepData enhancements

- Migration to relational SQL database with Java servlet front-end, sharing object model components with JetWeb
- Database sanitising e.g. axis-level properties, uniform units, measurement classes. . .
- Using “HepML” and AIDA interfaces: XSLT transformations to many data formats possible
- Will use Grid authentication for expts. to submit data directly (modulo sanity checking)

Durham University

New HepData db relational structure

Note this is not final — in fact users shouldn't even know about it!



A brief aside on “HepML”

- XML representation for generator configs and data sets
- Defined by a set of XML schema docs in the `http://www.cedar.ac.uk/xml/` namespace
- Will also contain Python API and Java XSLT transformer classes using HD object model
- Intended for use by JetWeb etc., also by external projects (e.g. Professor)
- Current sub-schemas for HepData data I/O and for generator configurations
- We’d like to work with generator authors (and MCDB) to make HepML a common interfacing/config format

Durham University

HepML fragment (1)

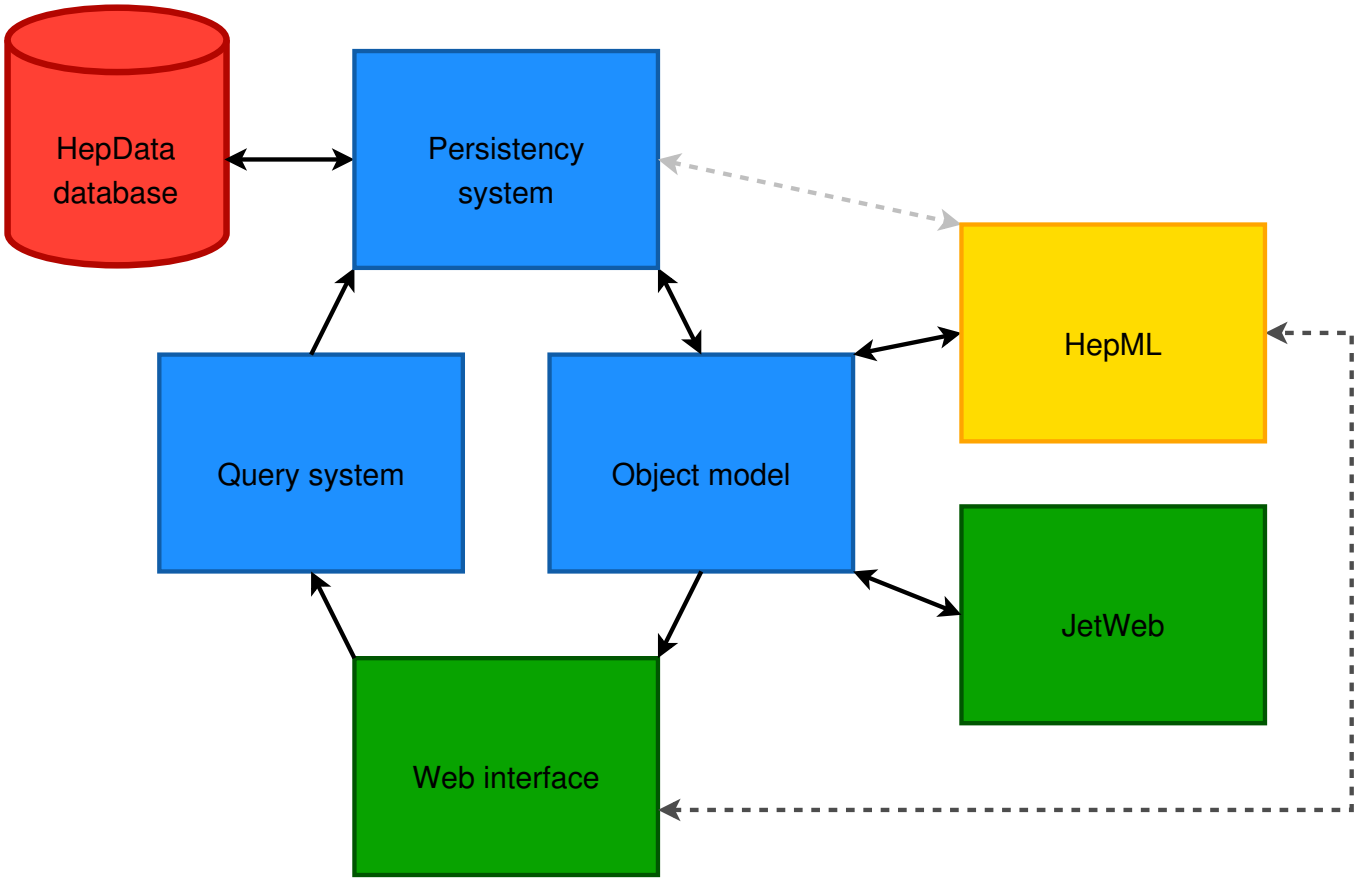
```
<?xml version="1.0" encoding="UTF-8"?>
<hepml xmlns="http://www.cedar.ac.uk/xml/"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://www.cedar.ac.uk/xml/schemas hepdata.xsd"
  <data timestamp="2006-02-12 00.14.50">
    <paper irn="3326047" paperId="3543">
      <dataset datasetId="1">
        <xaxis description="SQRT(S) IN GEV" xaxisId="1">
          <points>
            <point pointId="1" value="0.7726" />
            <point pointId="2" value="0.7776" />
            <point pointId="3" value="0.7801" />
            <point pointId="4" value="0.7826" />
            <point pointId="5" value="0.7851" />
            <point pointId="6" value="0.7876" />
            <point pointId="7" value="0.8026" />
          </points>
        </xaxis>

...
```

HepML fragment (2)

```
...
<yaxis description="SIG(Q=HAD) IN MUB" yaxisId="1">
  <points>
    <point pointId="1" value="0.269" />
    <point pointId="2" value="0.74" />
    ...
    <point pointId="7" value="0.065" />
  </points>
  <axiserror type="PCT_SYS" plus="11.0" minus="11.0"
     description="FROM NUCLEAR ABSORPTION CORRECTION" />
  ...
  <pointerrors type="NULL">
    <pointerror plus="0.096" minus="0.096" pointId="1" />
    <pointerror plus="0.192" minus="0.192" pointId="2" />
    ...
    <pointerror plus="0.036" minus="0.036" pointId="7" />
  </pointerrors>
</yaxis>
```

Durham University

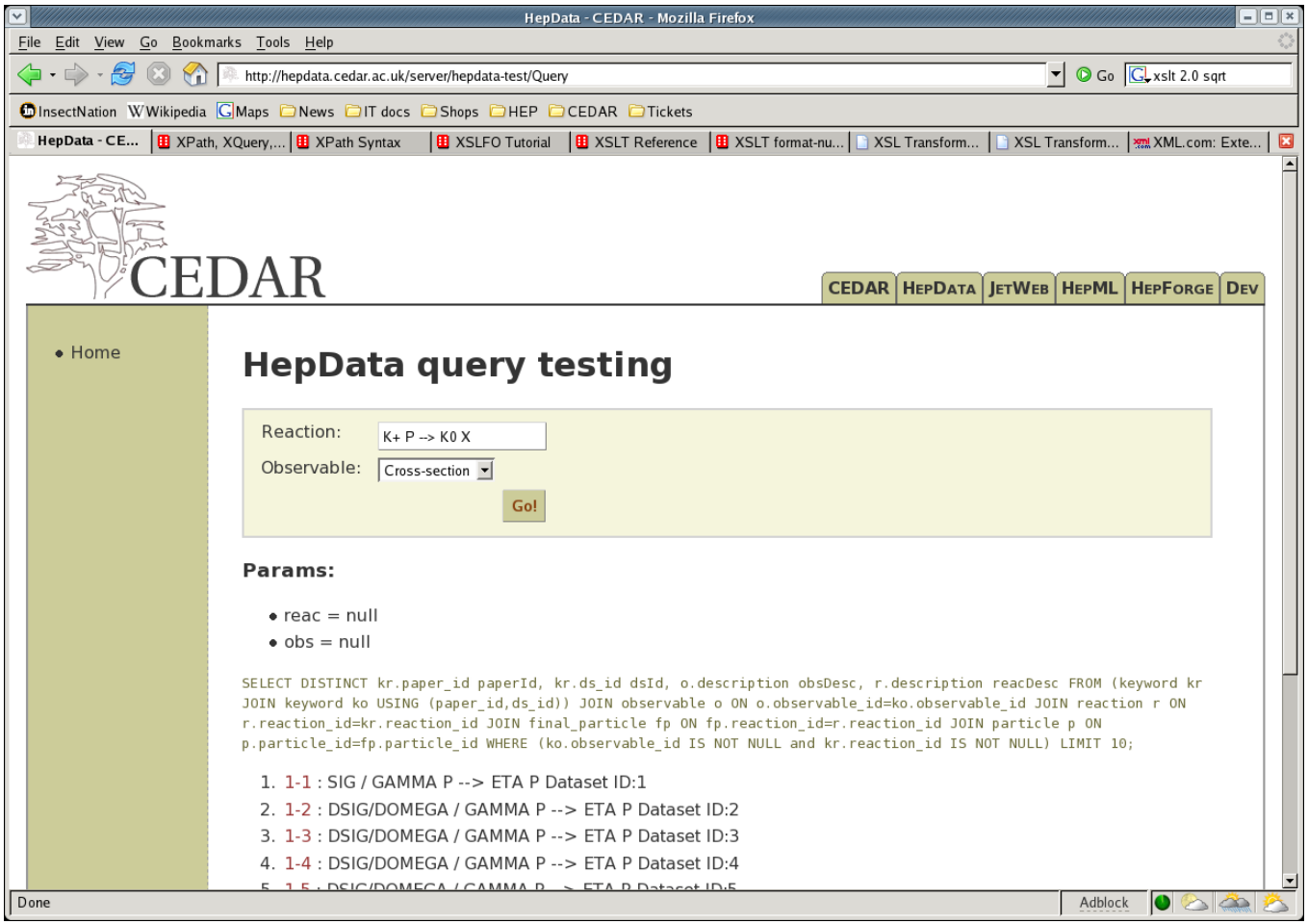
HepData behind the scenes



Durham University

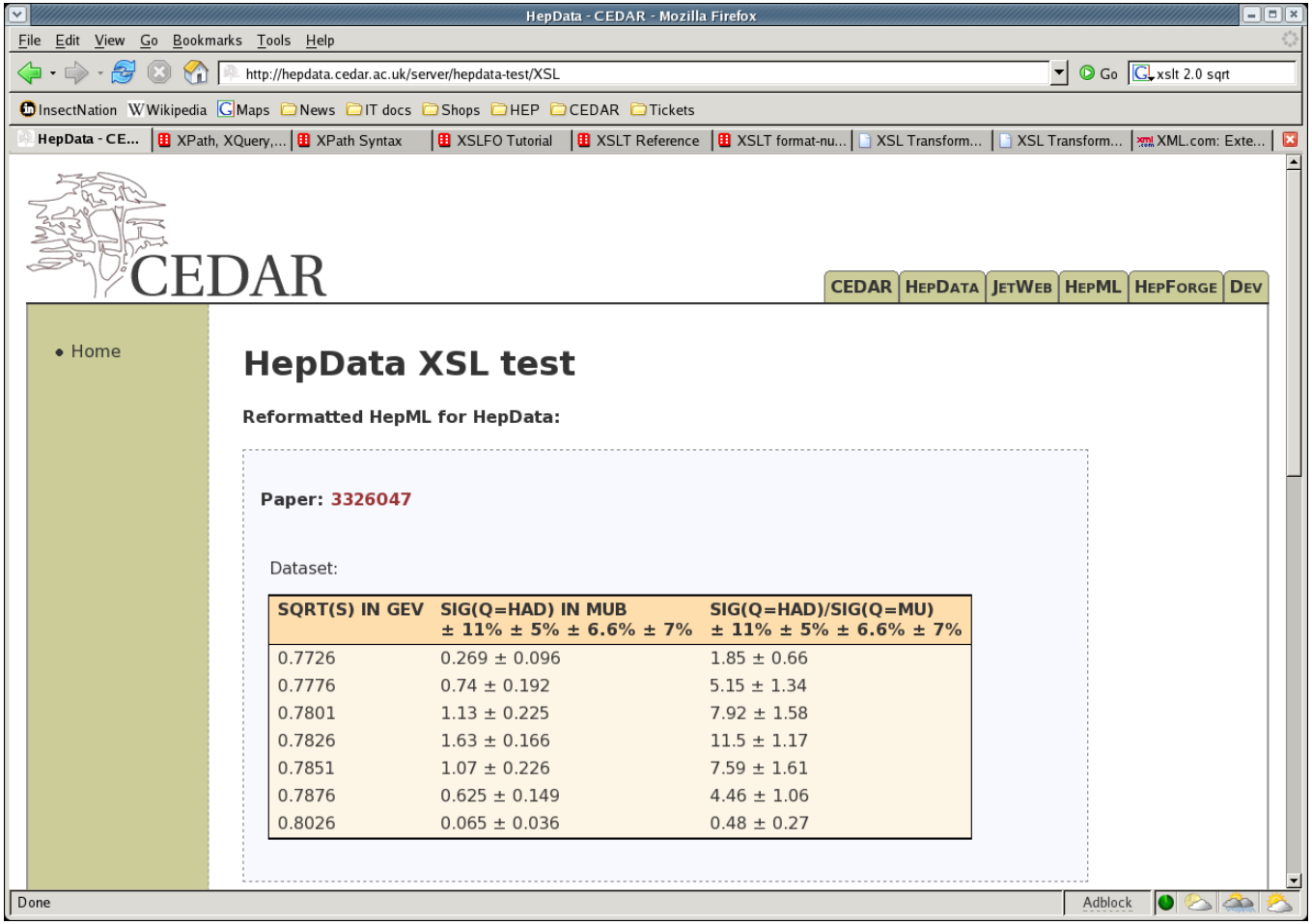
First HepData demos (1)

Query interface (breaking the rules — spot the SQL!)



First HepData demos (2)

HepData $\rightarrow$ HepML $\xrightarrow{XSL}$ HTML



Summary

- CEDAR is primarily a generator tuning system, combining JetWeb and HepData
- Hopefully greater than the sum of its parts!
- Aim is that any expt MC configuration should be "CEDAR-blessed" to be considered trustworthy
- Modular design and many spin-offs (see also HepForge talk on Thursday)
- CEDAR also hopes to encourage code re-use and uptake of common interfaces in HEP software
- Timescale: must be ready for LHC data. See `http://dev.cedar.ac.uk` and HepForge sites for roadmaps

Durham University