CEDAR: global tuning of MC event generators

JetWeb, HepData, HepML and HepForge

Andy Buckley

Institute for Particle Physics Phenomenology
Durham University, UK

Particle Physics 2006, Warwick, 2006-04-12

Durham University

Outline

Durham University

Introduction to CEDAR

Durham University

CEDAR: Collaborative e-Science Data Analysis Resource

CEDAR is an e-Science project with several sub-projects:

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

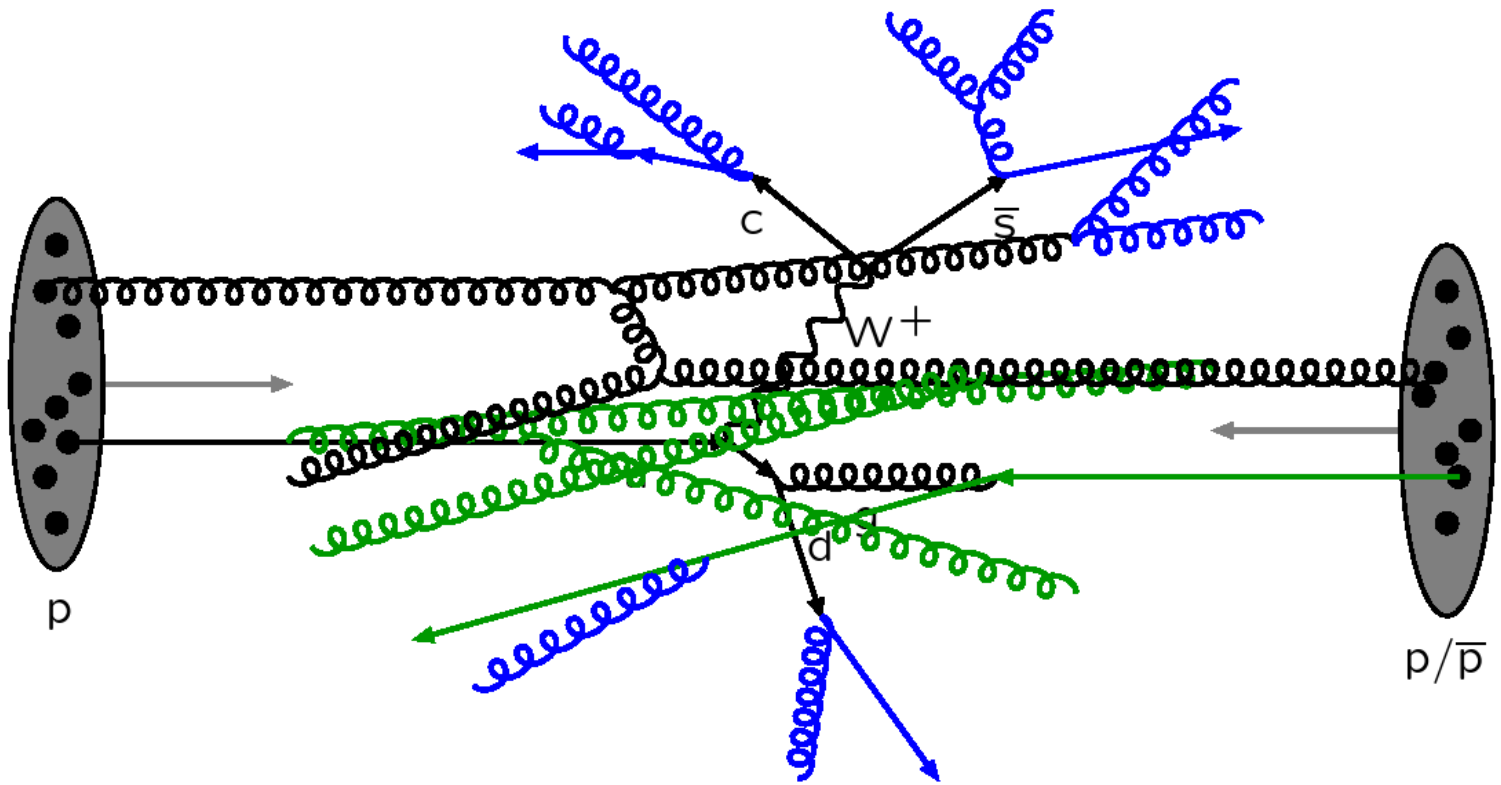
www.cedar.ac.uk

Durham University

JetWeb — a system for global MC tuning

Durham University

Components of MC event generators



ME, ISR & FSR parton showers, underlying event, hadronisation, decays. . .

Event generators: tuneable parameters

- Parton density functions → MRST/CTEQ...
- Matrix elements (kinematics, phase space, coupling/scale) → order, match to PDF?
- Parton showers and ME matching → which shower?
- Threshold effects in heavy flavour production
- Fragmentation functions
- Hadronisation → string/cluster params, strangeness...
- Underlying event

Large number of params: nonsensical tunings dominate parameter space

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

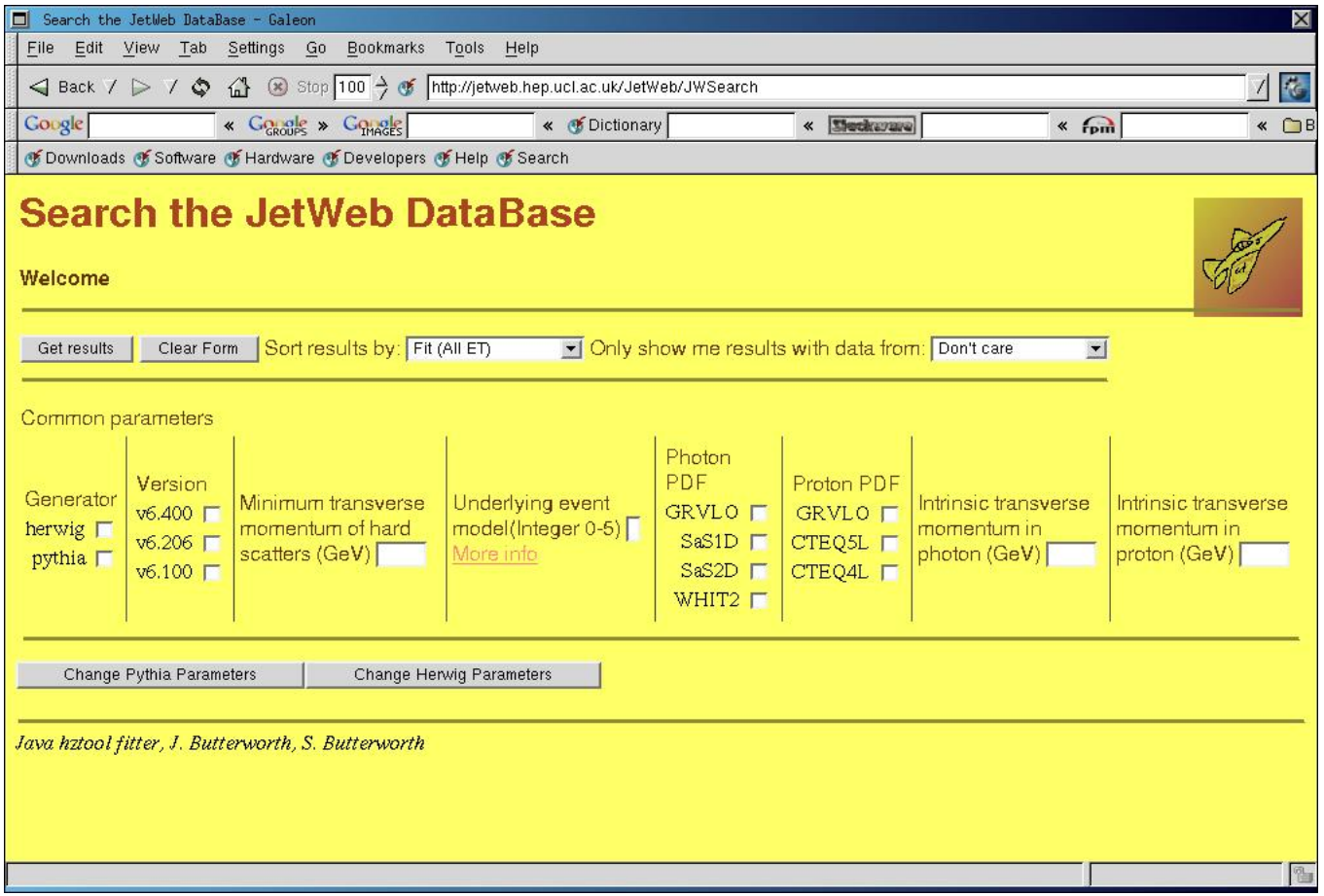
- Written in Java, run on Apache + Tomcat framework
- Database of global data fit qualities for MC models (generator & params)
- Web interface to view fits and plots
- Users can request generation of a particular model if not (sufficient) in db
- Comparisons via a set of routines, each corresponding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

Durham University

JetWeb in action (1)

Searching

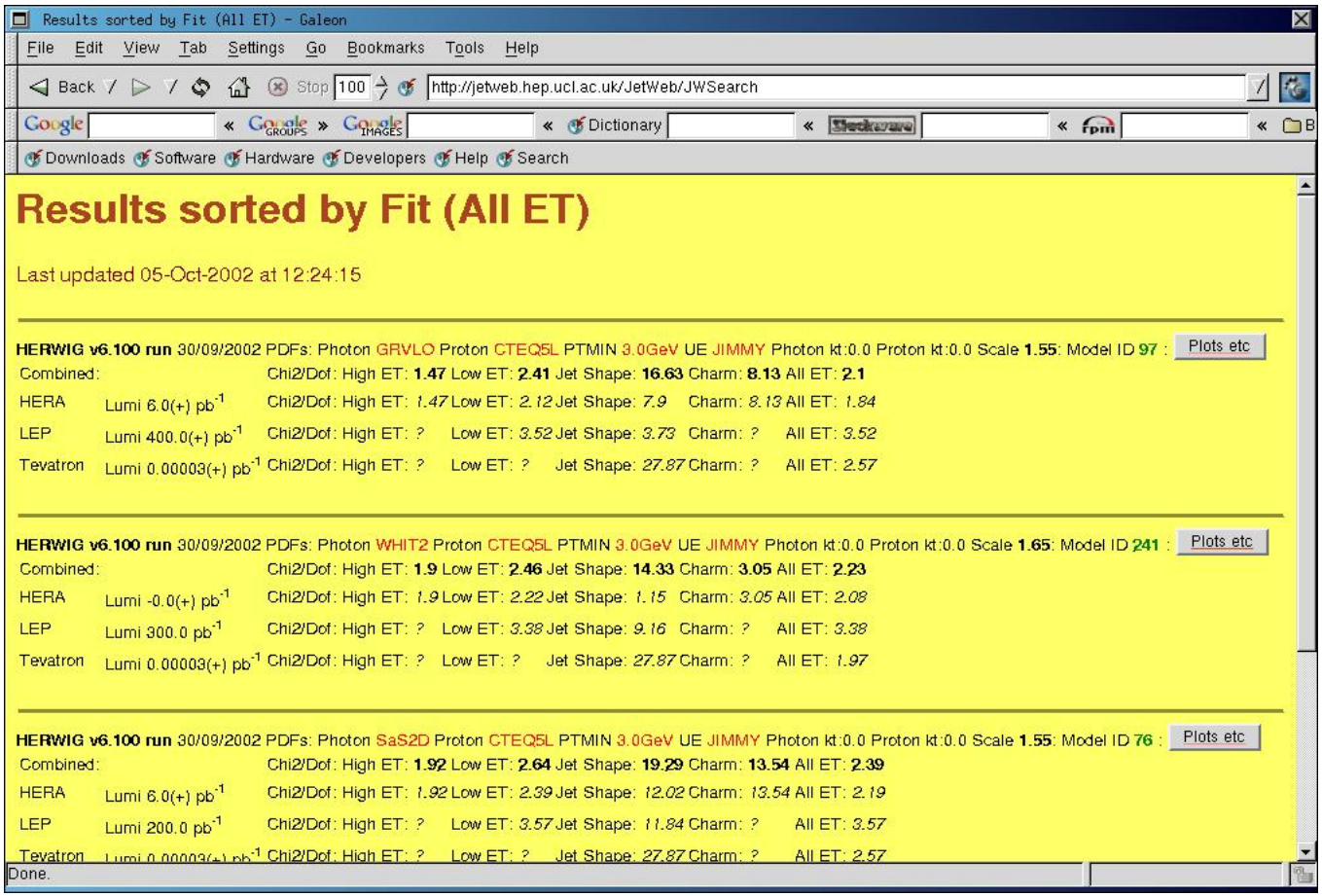
NB. This is the *frozen* copy of JetWeb. . .



JetWeb in action (2)

Fit list

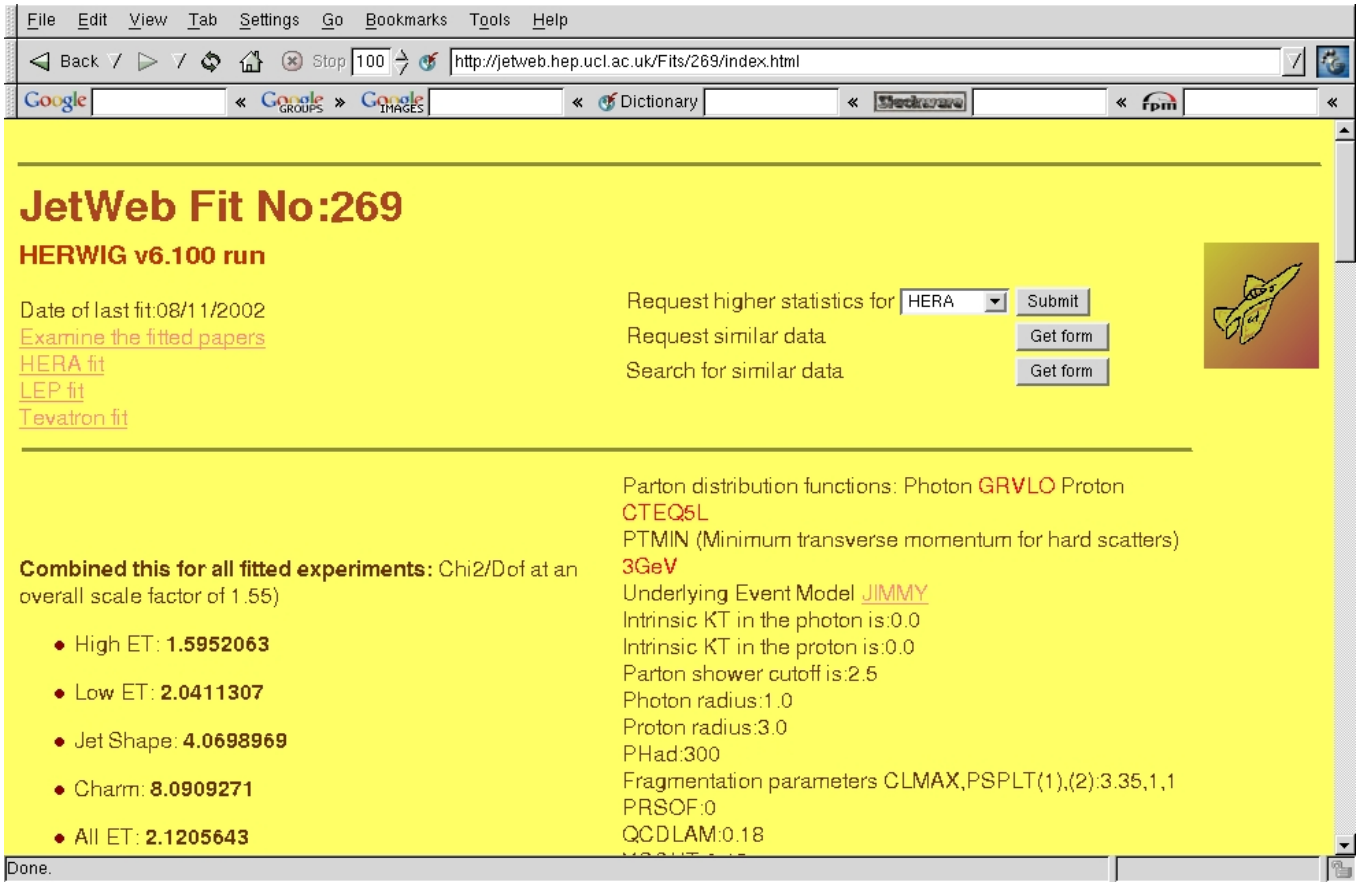
NB. This is the *frozen* copy of JetWeb. . .



JetWeb in action (3)

Fit details

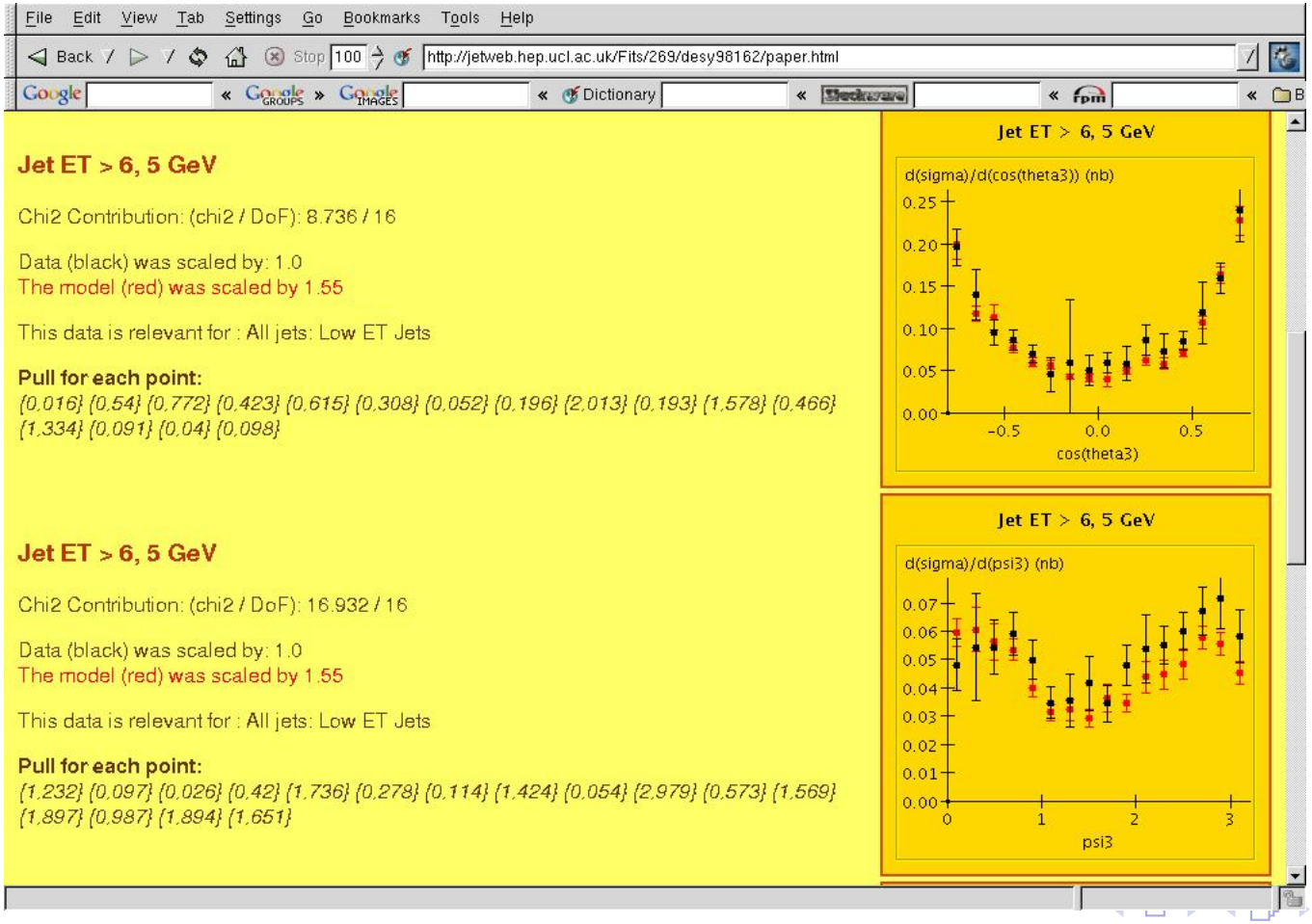
NB. This is the *frozen* copy of JetWeb. . .



JetWeb in action (4)

Paper view

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



HzTool: `http://hepforge.cedar.ac.uk/hztool/`

JetWeb is only the front-end, distribution and comparison system. “Experimental analyses” are 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: generator specifics in HzSteer package
- Rivet: C++ replacement being written by CEDAR: **`http://hepforge.cedar.ac.uk/rivet/`**

Durham University

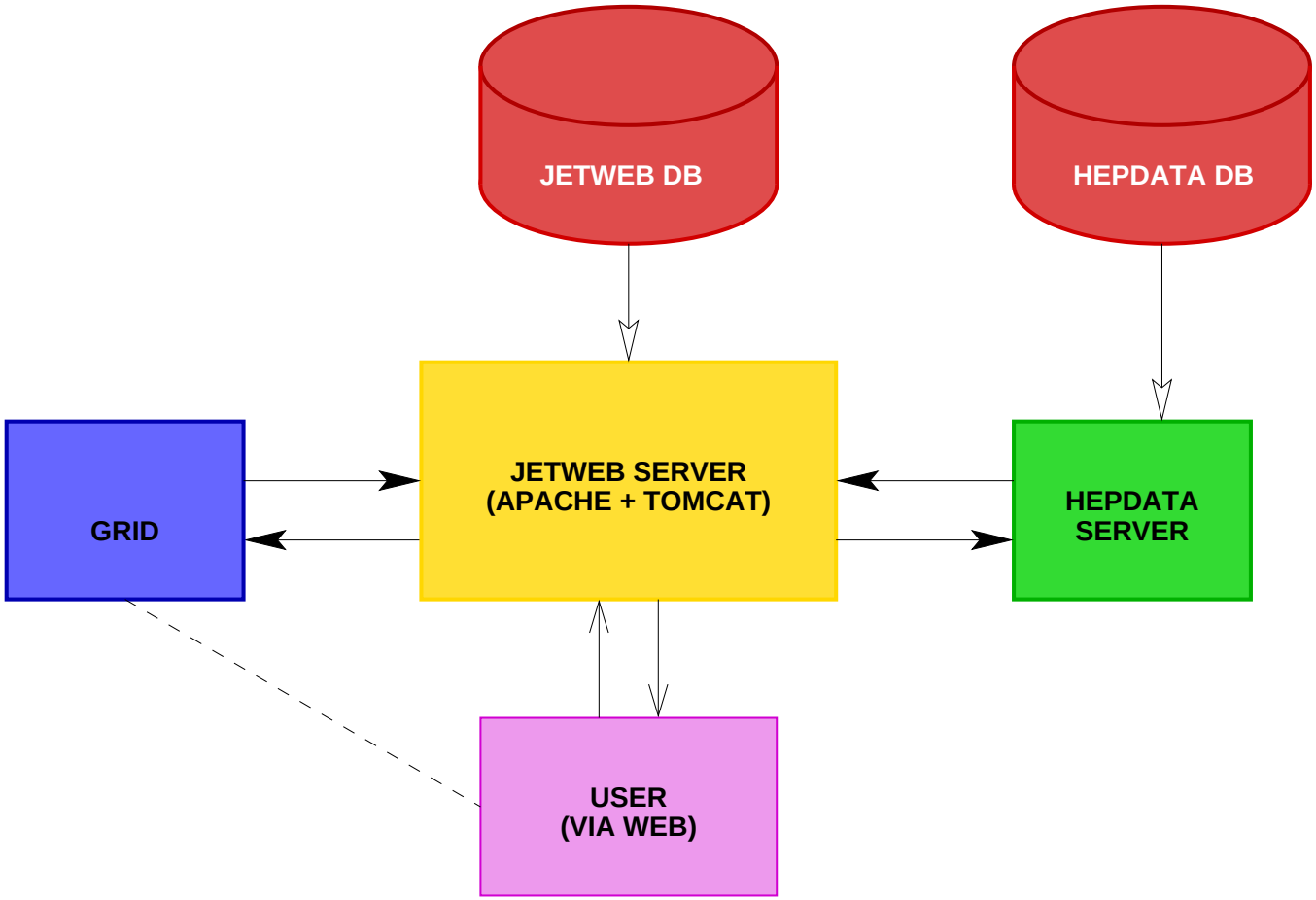
JetWeb enhancements

CEDAR is enhancing JetWeb in several ways:

- Experimental data to be taken from (upgraded) HepData
- Framework re-factor to make model extensions easier
- Plans to use Grid authentication and job distribution
- New histogramming modules using the AIDA interfaces
- (Scanning parameter space, re-weighted fits...)
- First CEDAR version of JetWeb to be available soon

Durham University

(New) JetWeb behind the scenes



HepData — the HEP reaction database

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

A brief aside on “HepML”

- XML representation for generator configs and HepData data sets (and more to come?)
- Defined by a set of XML schema docs under the `http://www.cedar.ac.uk/hepml/` 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)
- Will be a common MC generator log/config format (MCnet)
- HepData XML schema version 0.1 (release for comment) very soon: please check it out!
  `http://hepforge.cedar.ac.uk/hepml/`

Durham University

HepML fragment (1)

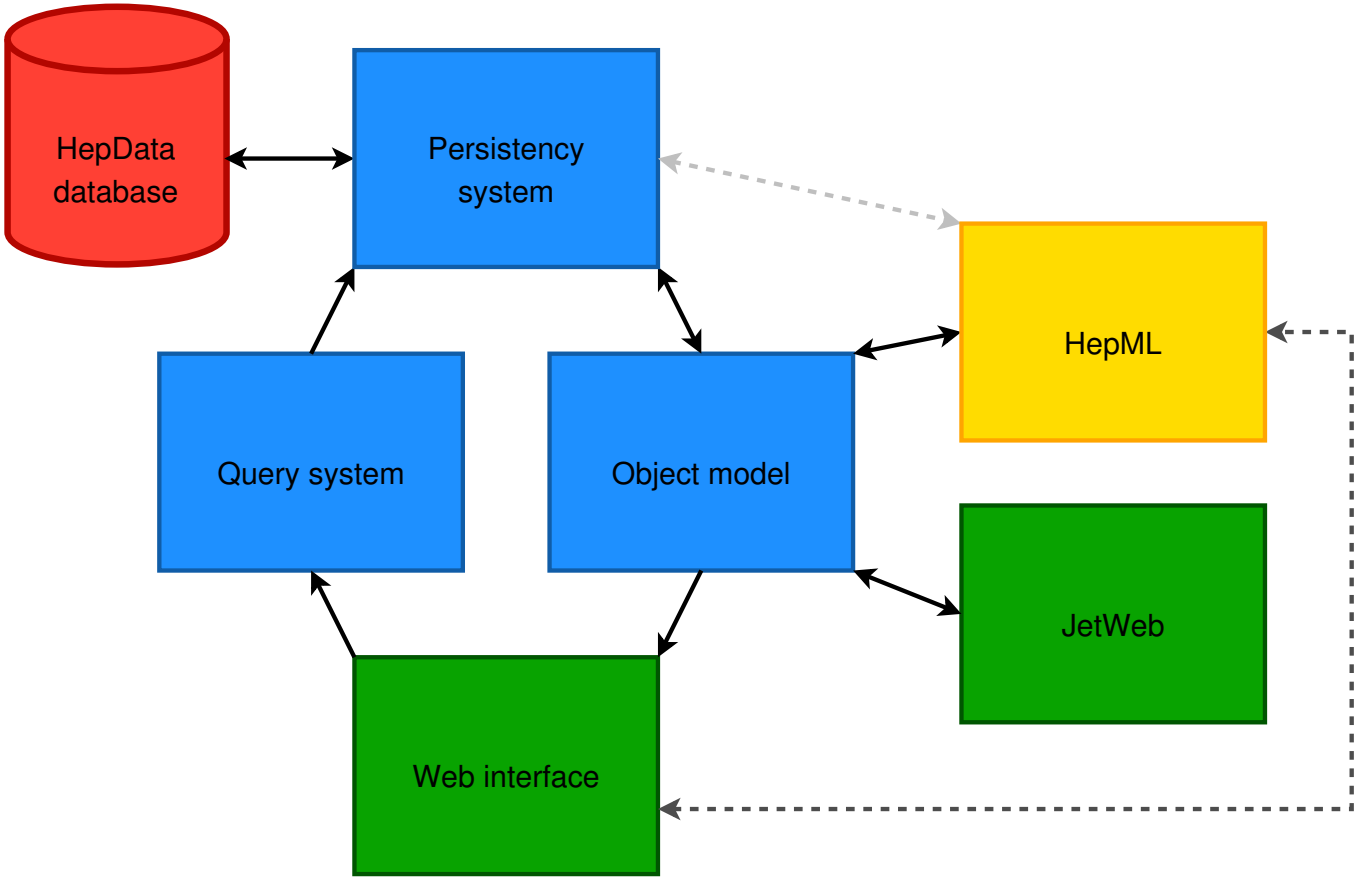
```
<?xml version="1.0" encoding="UTF-8"?>
<hepml xmlns="http://www.cedar.ac.uk/hepml/hepdata/0.1/">
  <data timestamp="2006-04-07 13.09.27">
    <paper irn="3326047" paperId="3552">
      <dataset datasetId="1">
        <comment>
          FITTED, BACKGROUND SUBTRACTED, PEAK OMEGA CROSS SECTION,
          ...
        </comment>
        <property name="RE(Q=MU)" value="E+ E- --&gt; MU+ MU-" />
        <property name="RE(Q=HAD)" value="E+ E- --&gt; PI+ PI0 PI-" />
        <xaxis header="SQRT(S) IN GEV" xaxisId="1">
          <bins>
            <bin pointId="1" value="0.7726" />
            <bin pointId="2" value="0.7776" />
            ...
          </bins>
        </xaxis>

...
```

HepML fragment (2)

```
                                ...
<yaxis header="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 norm="percent" source="sys" plus="11.0" minus="11
     description="FROM NUCLEAR ABSORPTION CORRECTION" />
  ...
  <pointerrors norm="abs" source="stat">
    <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>
```

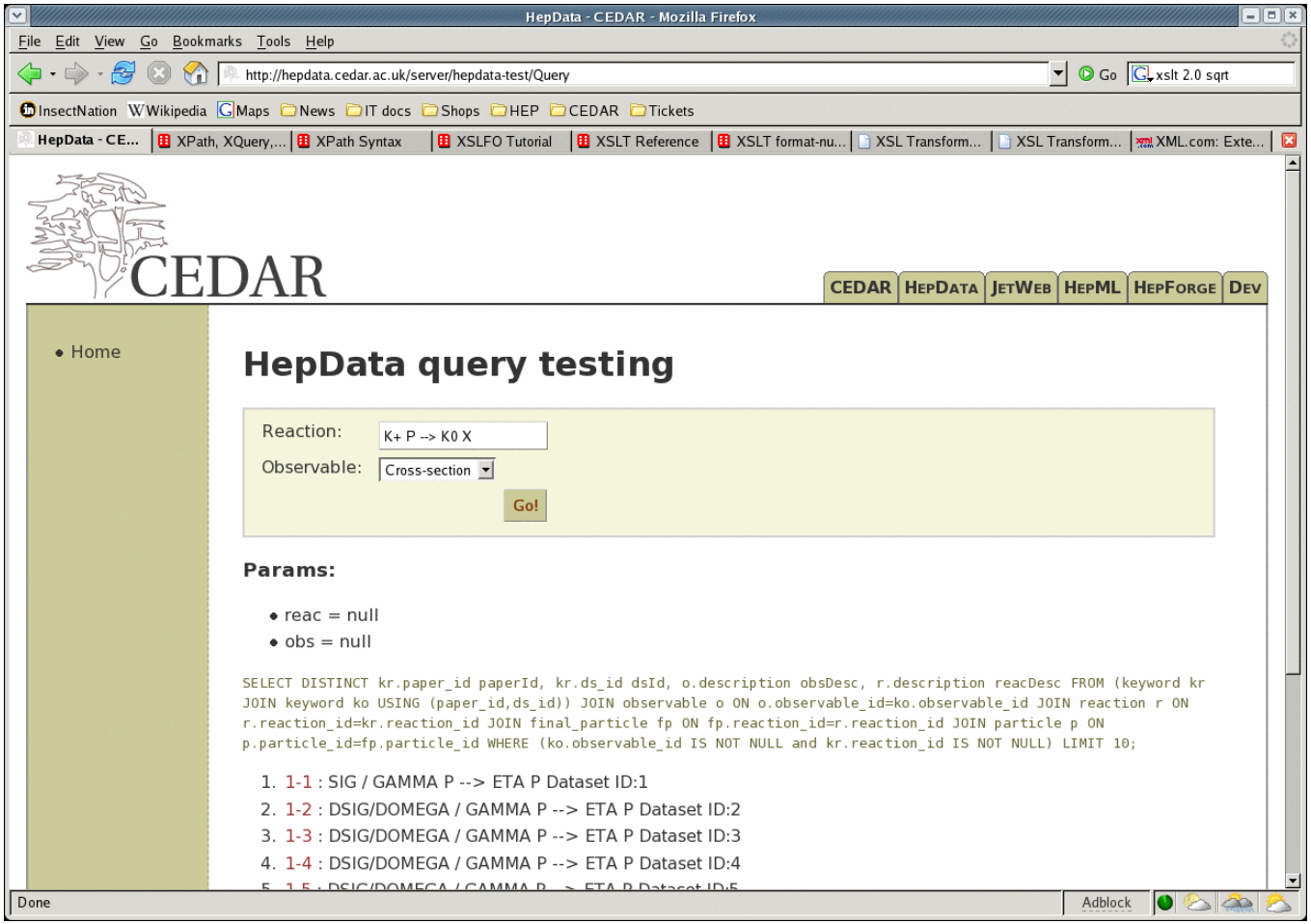
HepData behind the scenes



Durham University

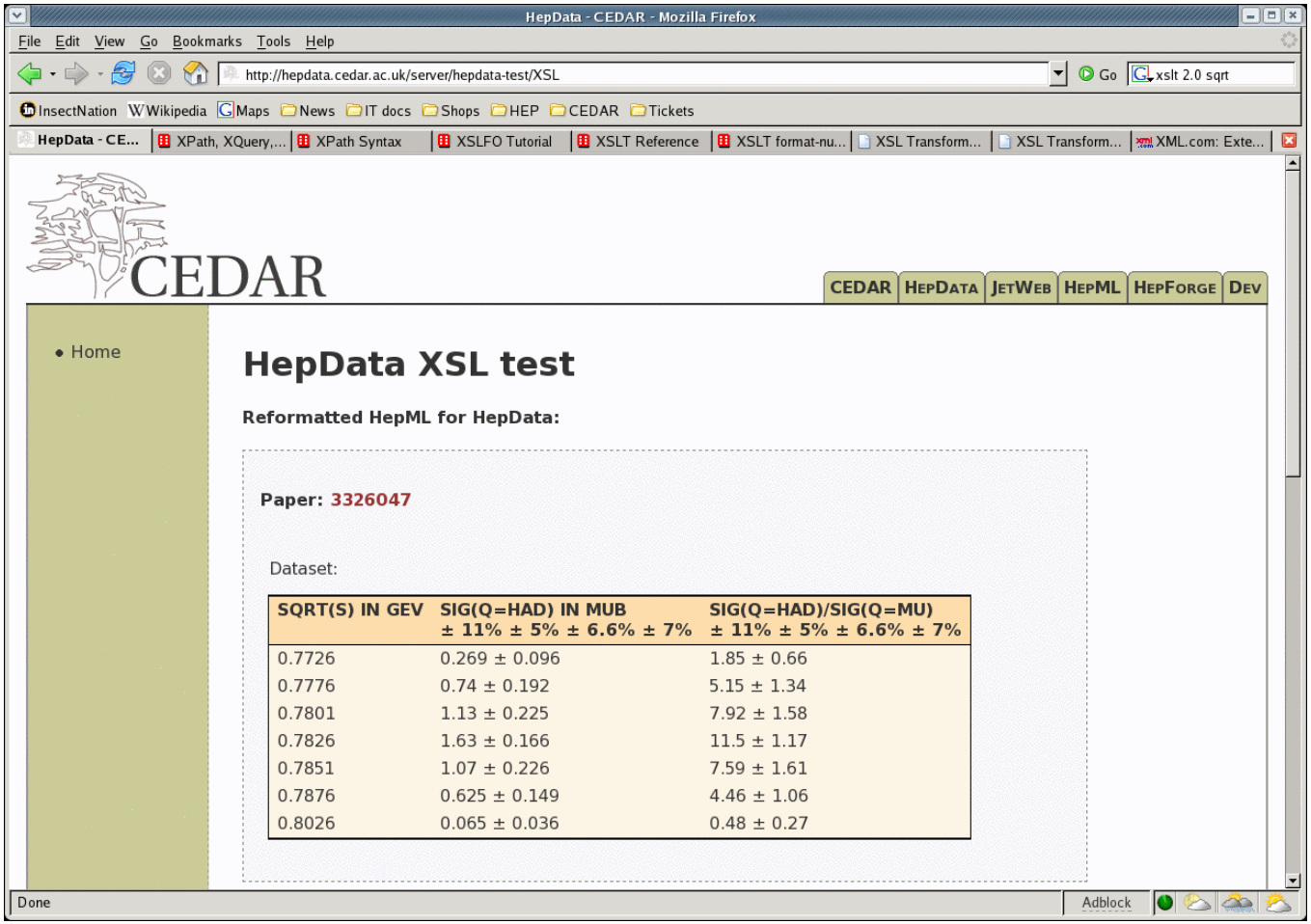
First HepData demos (1)

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



First HepData demos (2)

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



HepForge — a lightweight development environment for HEP

Durham University

Software engineering for small HEP projects

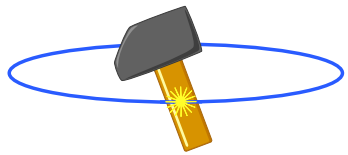
- Everyone has written code that might be re-used. . .
- But hard to get managed development started: need to find/install software. . .
- Small projects don't have the resources to do nice things, like:
  - Using standard build/installation systems e.g. autotools, libtool
  - Being independent of execution/build environment e.g. `/cern`!
  - Formal quality control, feedback and bug tracking
  - Version control systems (CVS, Subversion)

HepForge provides many of these features in a powerful, easy-to-use environment.

Durham University

HepForge

- New collaborative development system
- Online at **http://hepforge.cedar.ac.uk**
- Features including:
  - Shell access with full set of dev tools
  - Web space (with several convenient features)
  - CVS and Subversion version control (+ public viewers)
  - Wiki and bug tracker
  - Mailing lists for developers and users
  - Downloads manager
- We've done the "boring bits"!
- Quite a few users: Herwig++, Pythia6, LHAPDF...



Durham University

Who should be interested in HepForge?

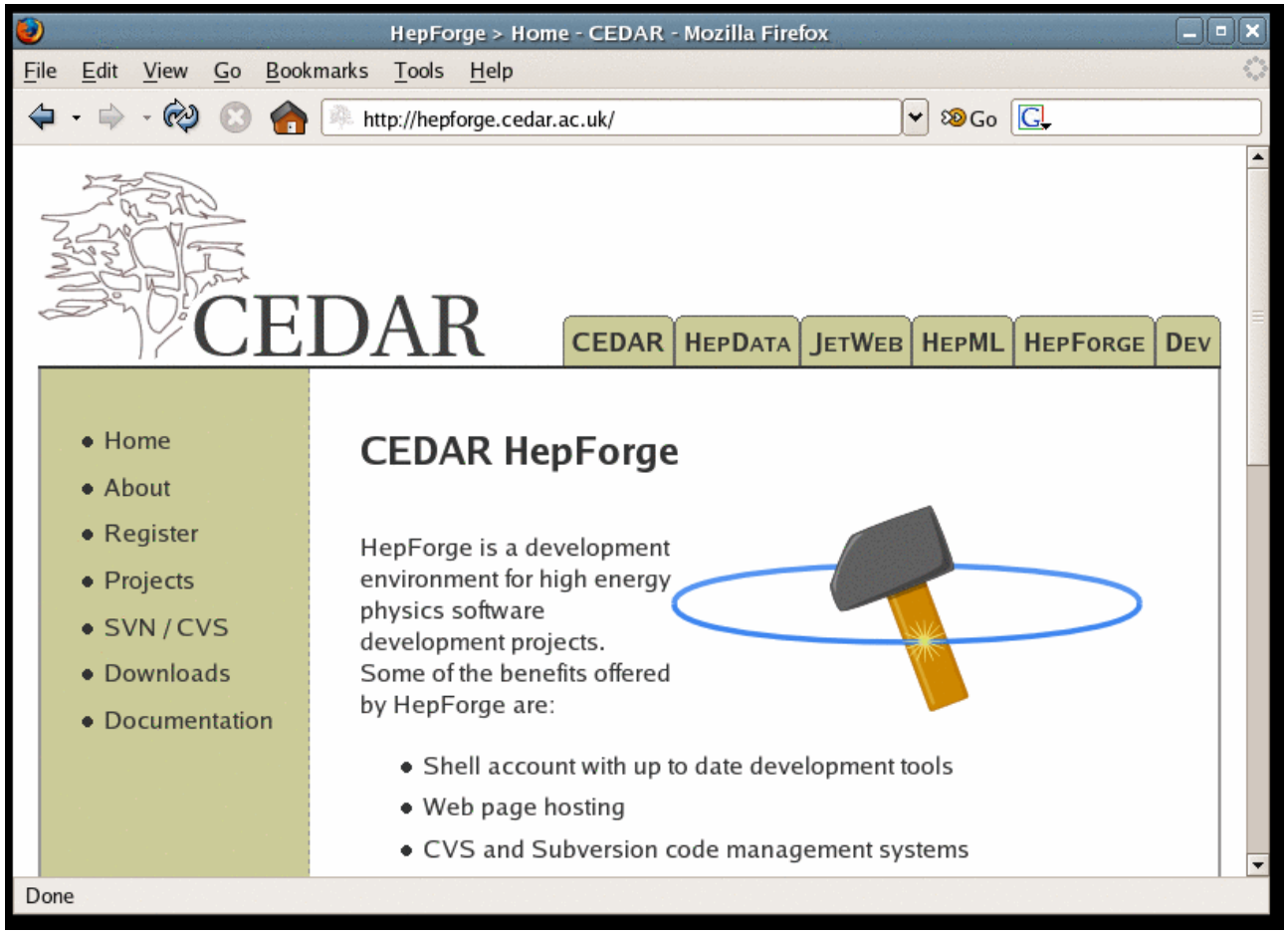
- Who's it for? Small–medium size projects.
- Stand-alone preferred but not reqd. Aids re-use.
- Intentions to use standard external systems
  Don't re-invent or *break expected behaviour* (UI design)
- Intention to document properly
- General responsible development, essentially

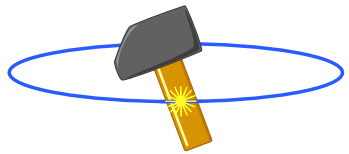
In return, HF provides powerful software management infrastructure with minimum entry level and learning curve

Durham University

HepForge tour

Home page

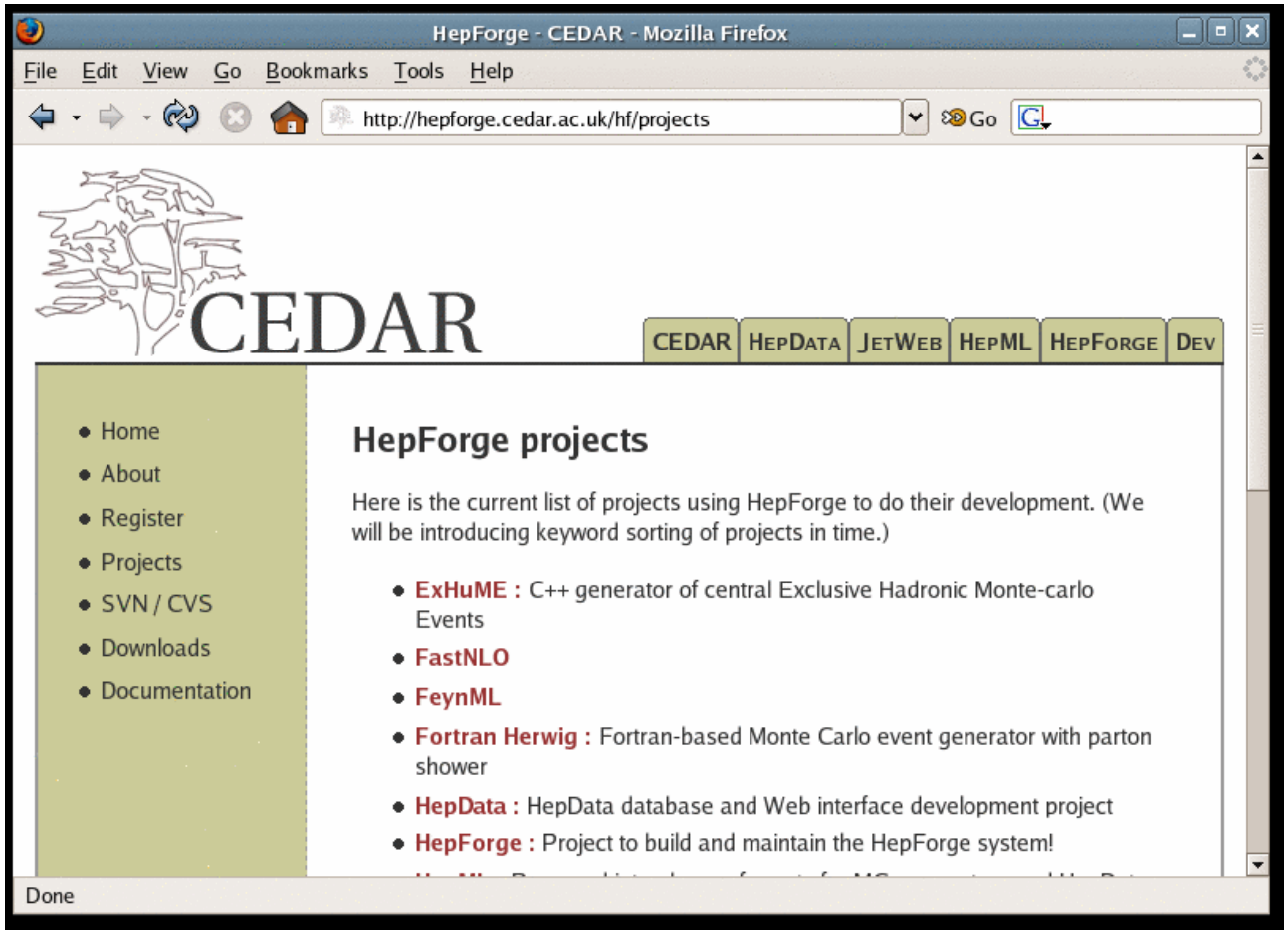


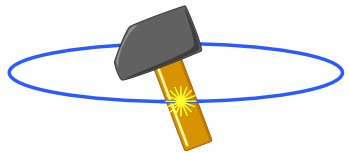


Durham University

HepForge tour

Project list (1)

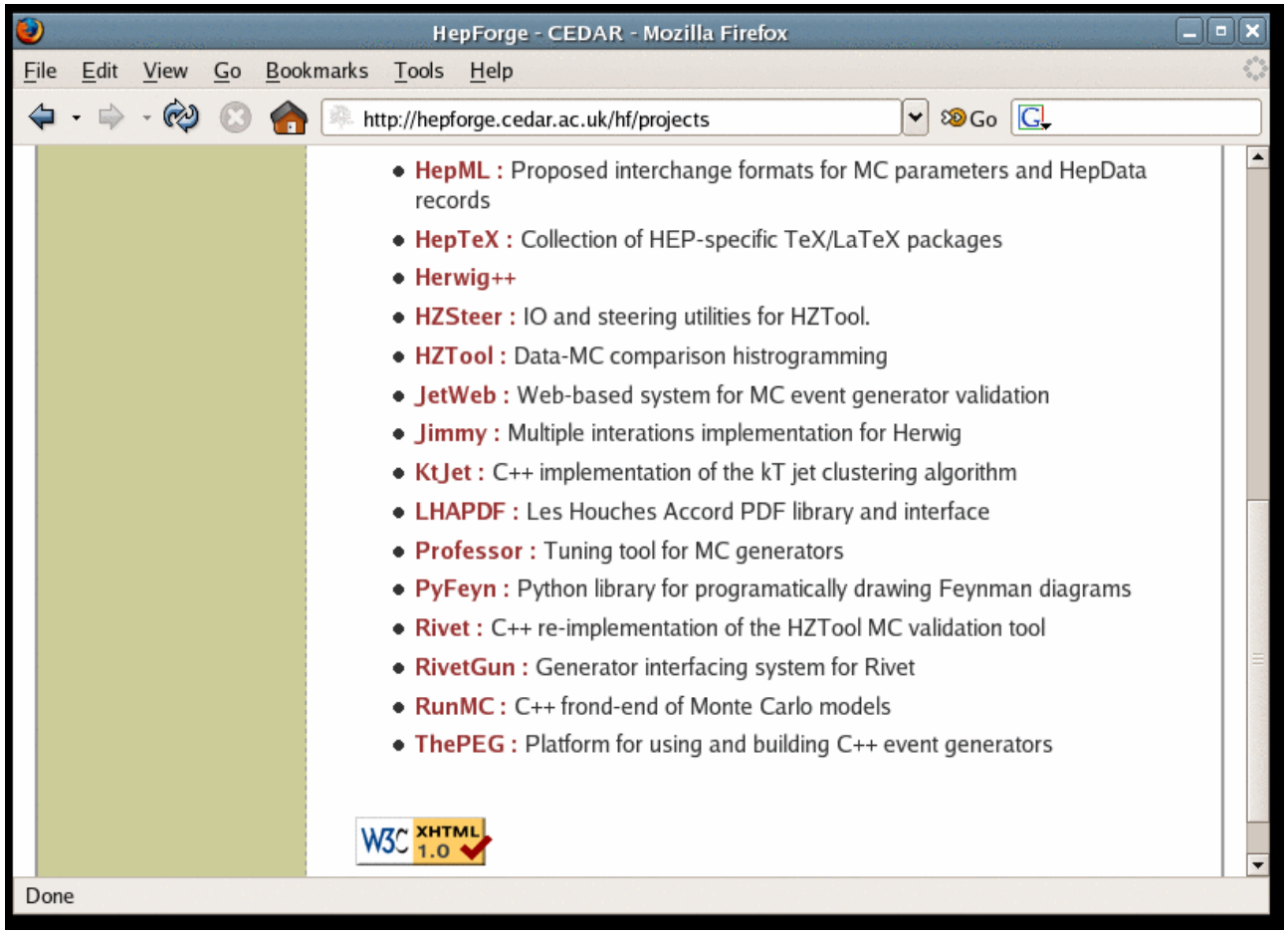


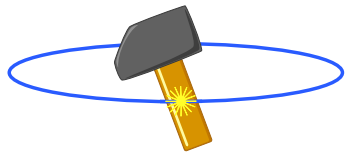


Durham University

HepForge tour

Project list (2)

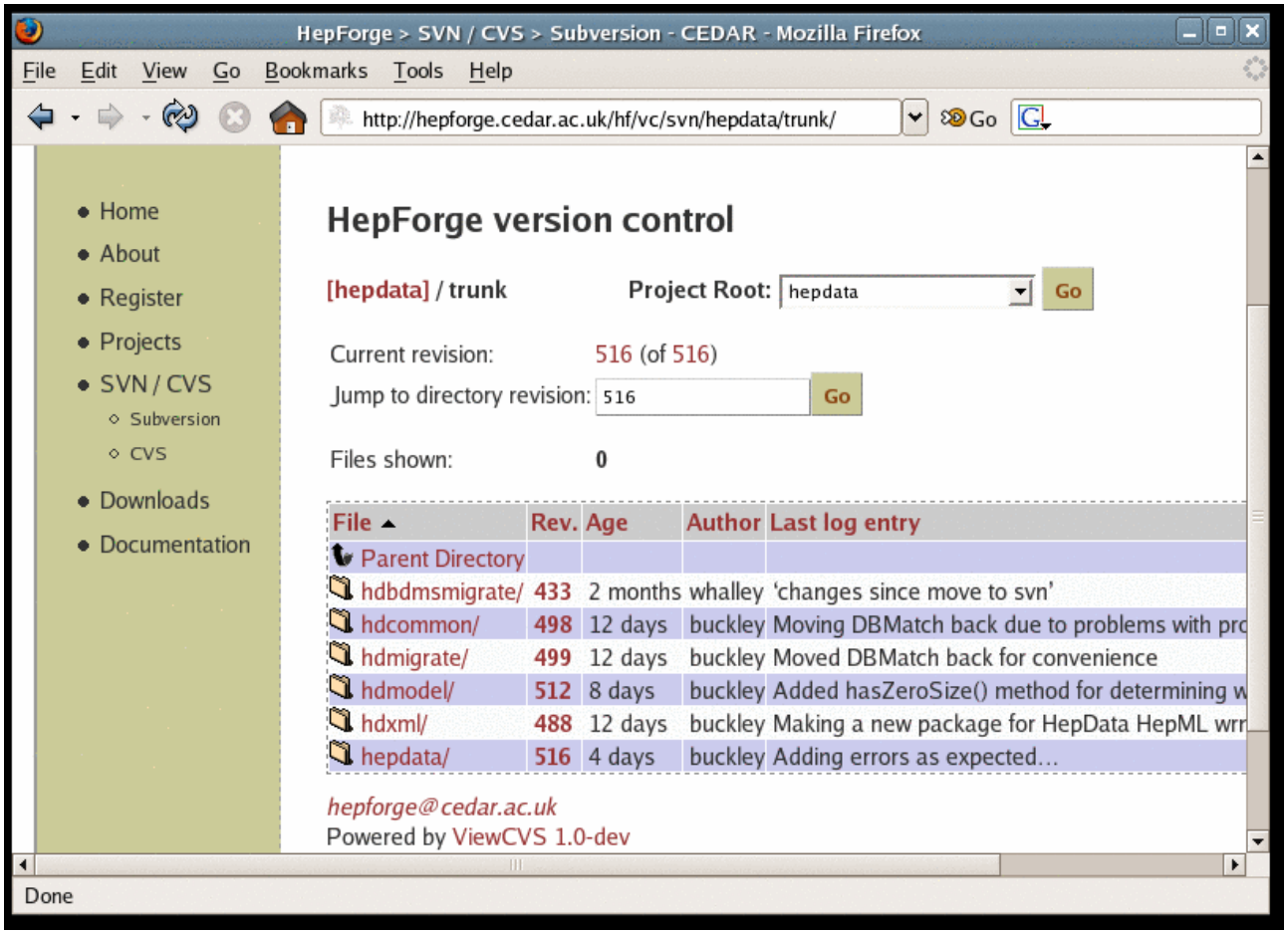


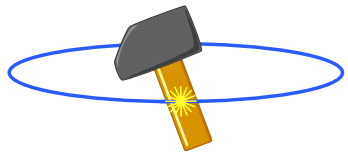


Durham University

HepForge tour

Project VC listing

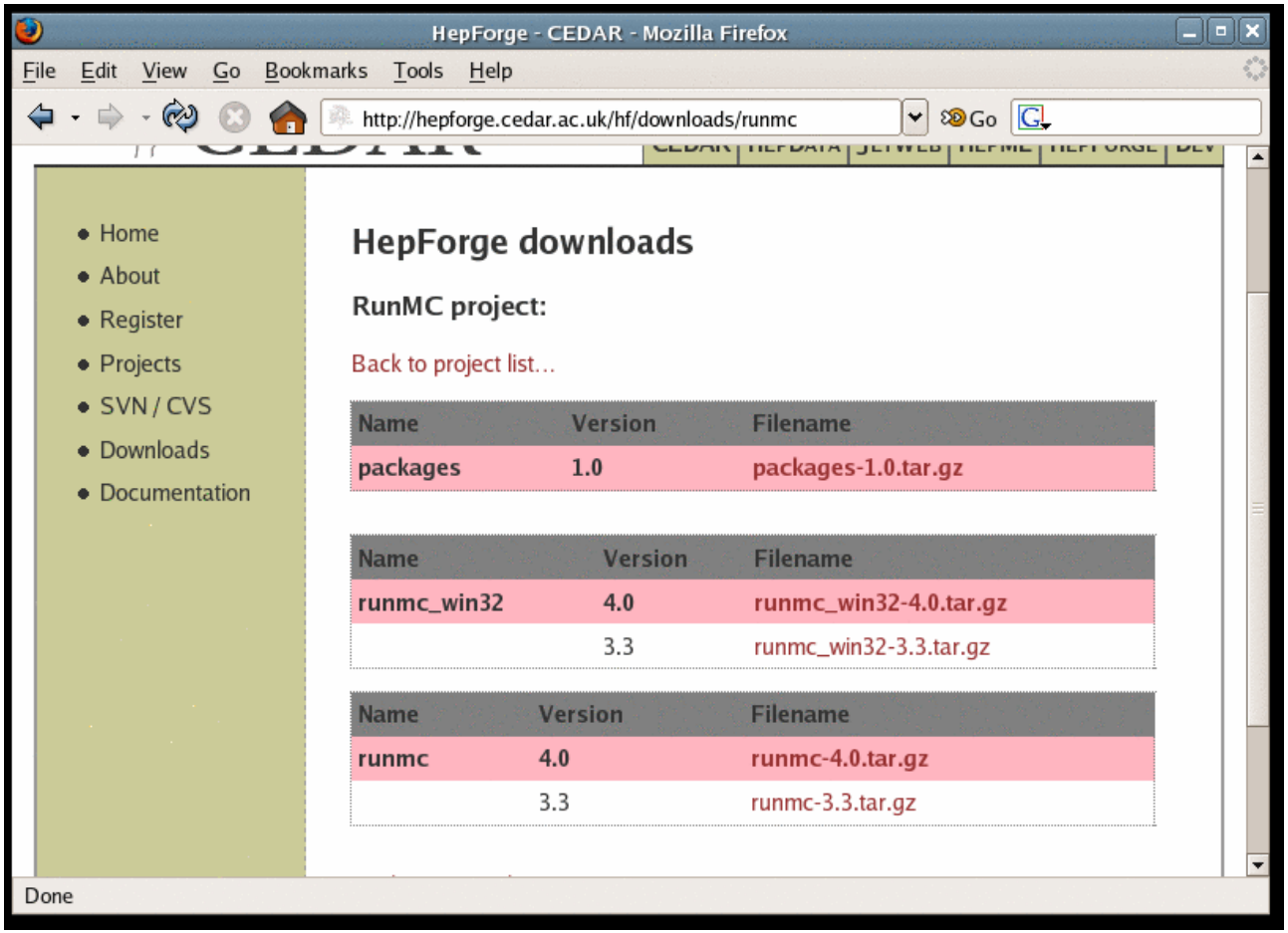


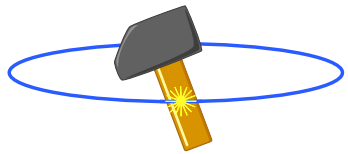


Durham University

HepForge tour

Project downloads listing (also personal copy via SSI)

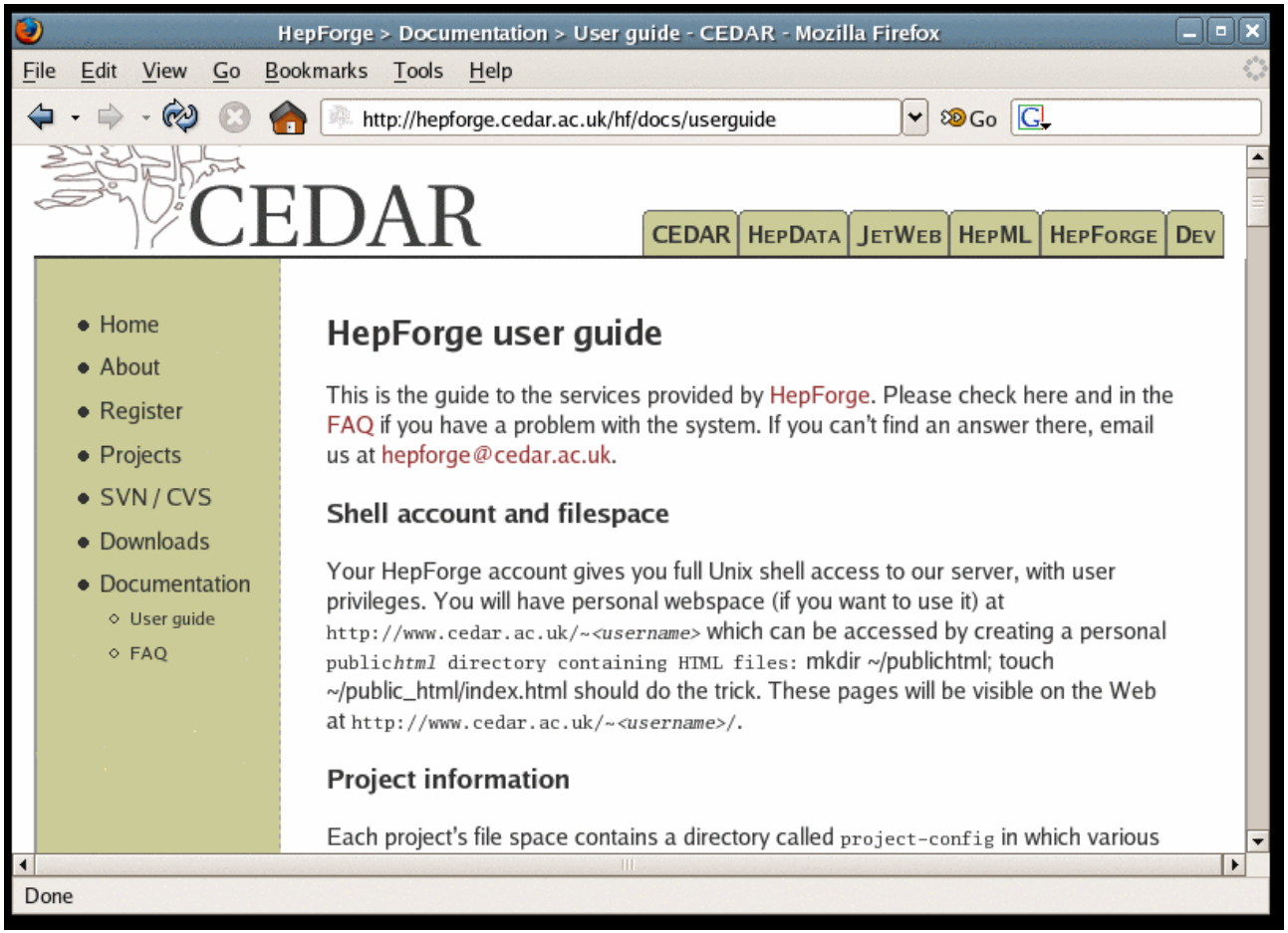


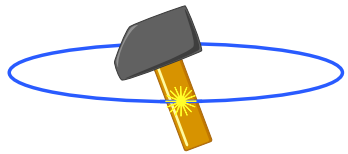


Durham University

HepForge tour

Documentation: user guide (note scrollbar!)

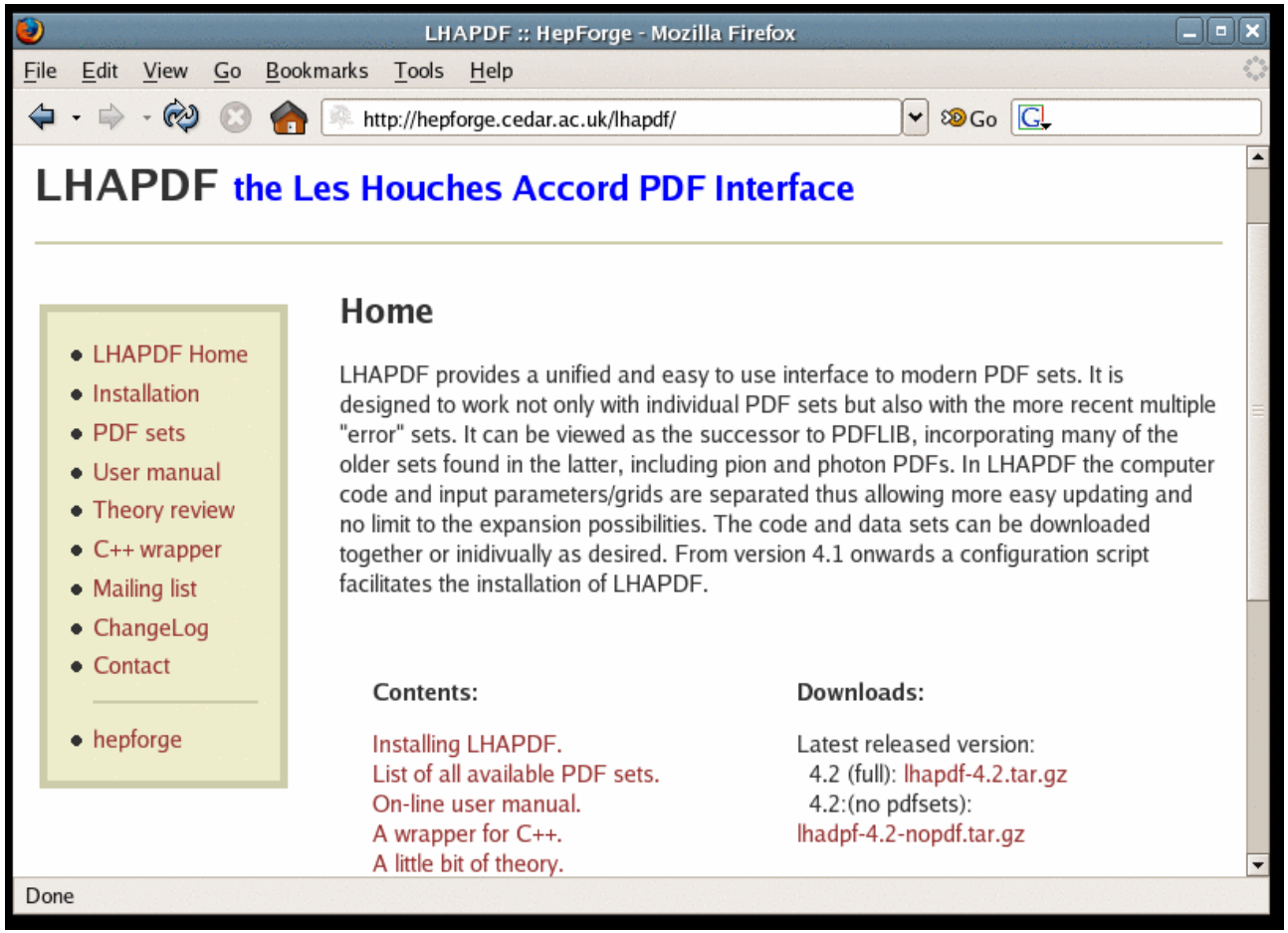


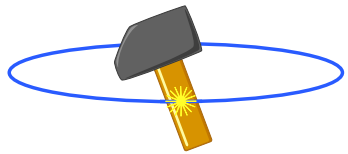


Durham University

HepForge tour

Project Web page: LHAPDF

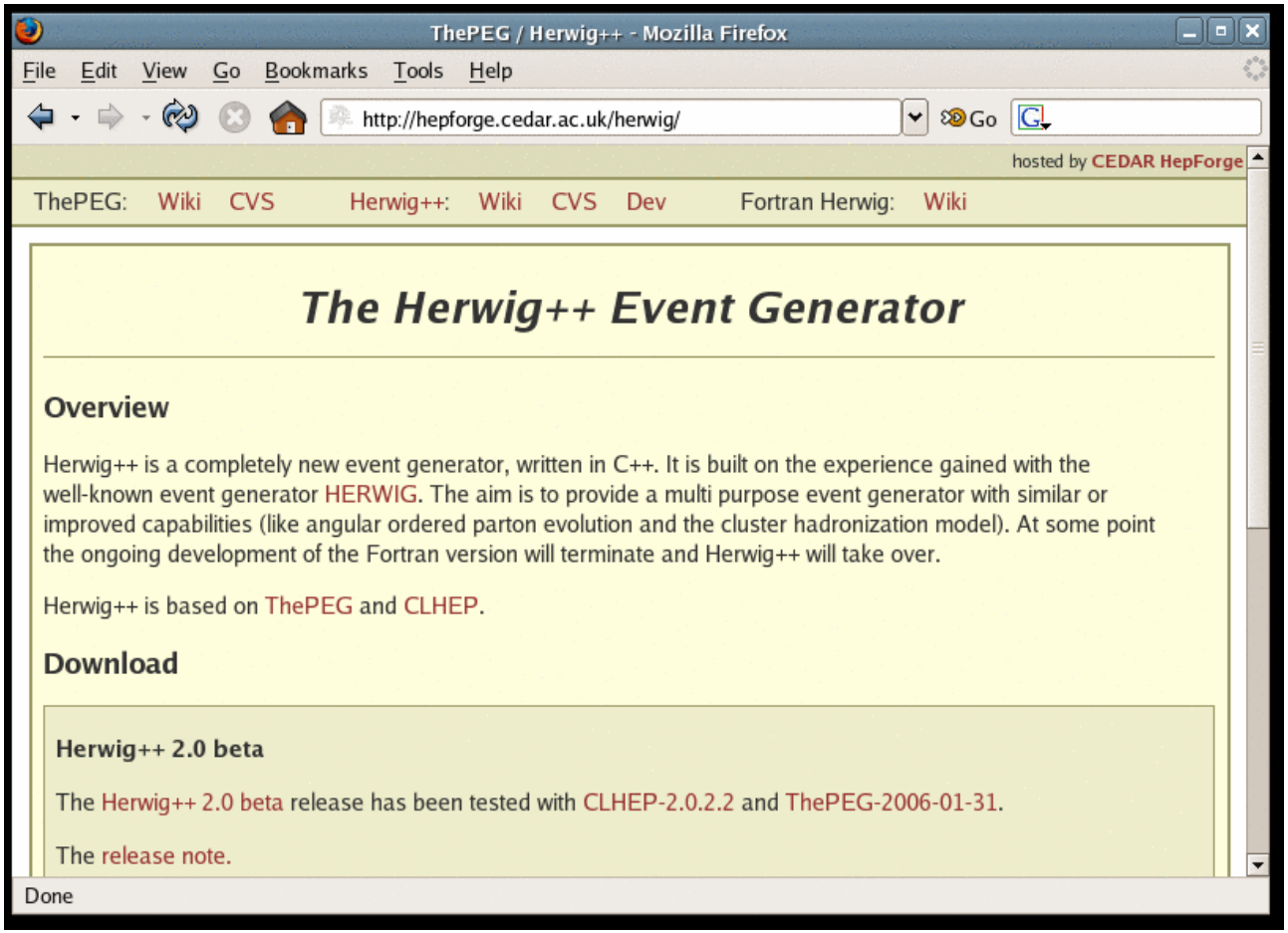


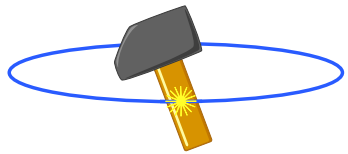


Durham University

HepForge tour

Project Web page: Herwig++

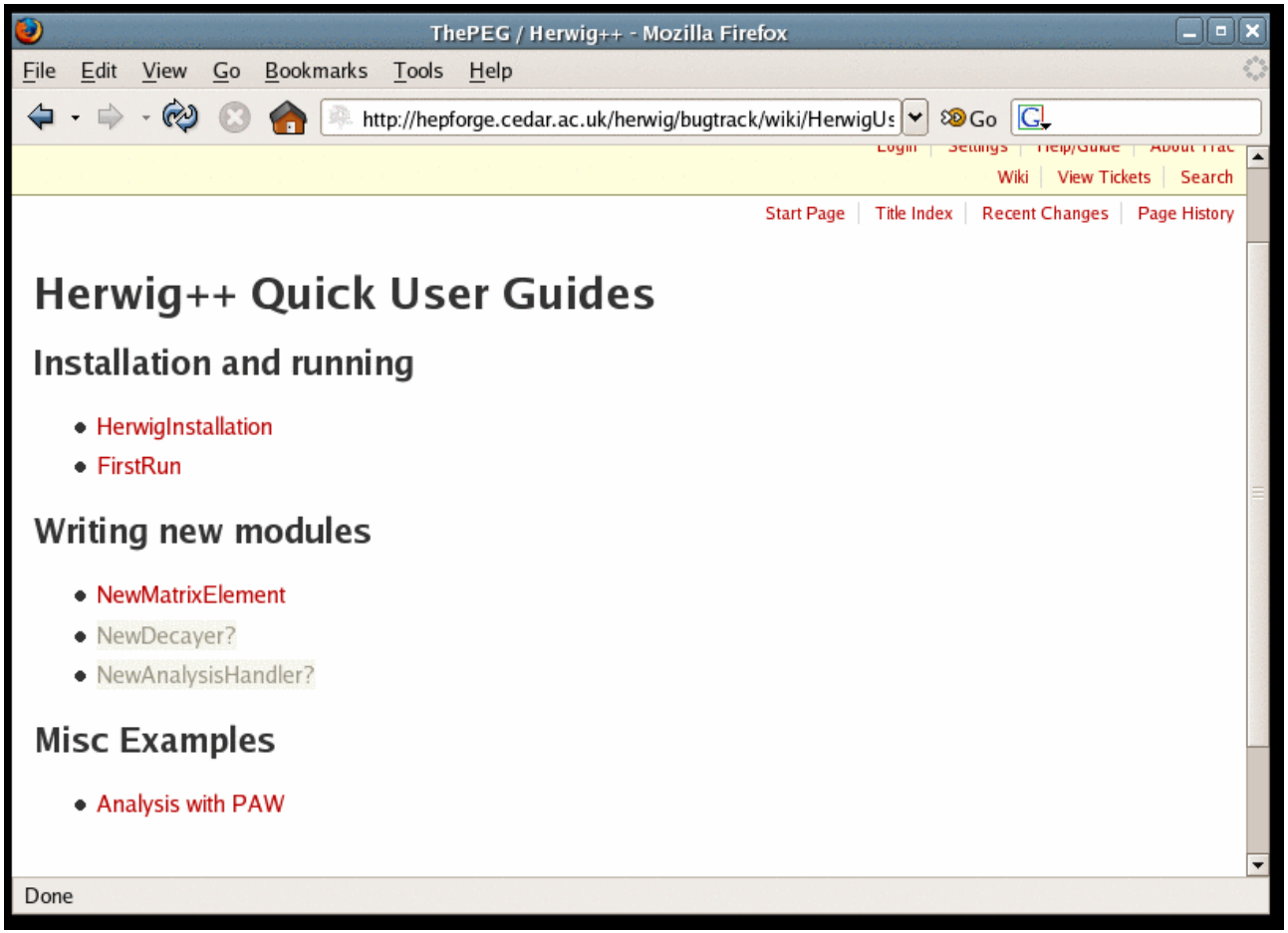


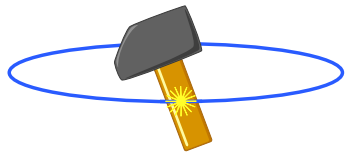


Durham University

HepForge tour

Project wiki

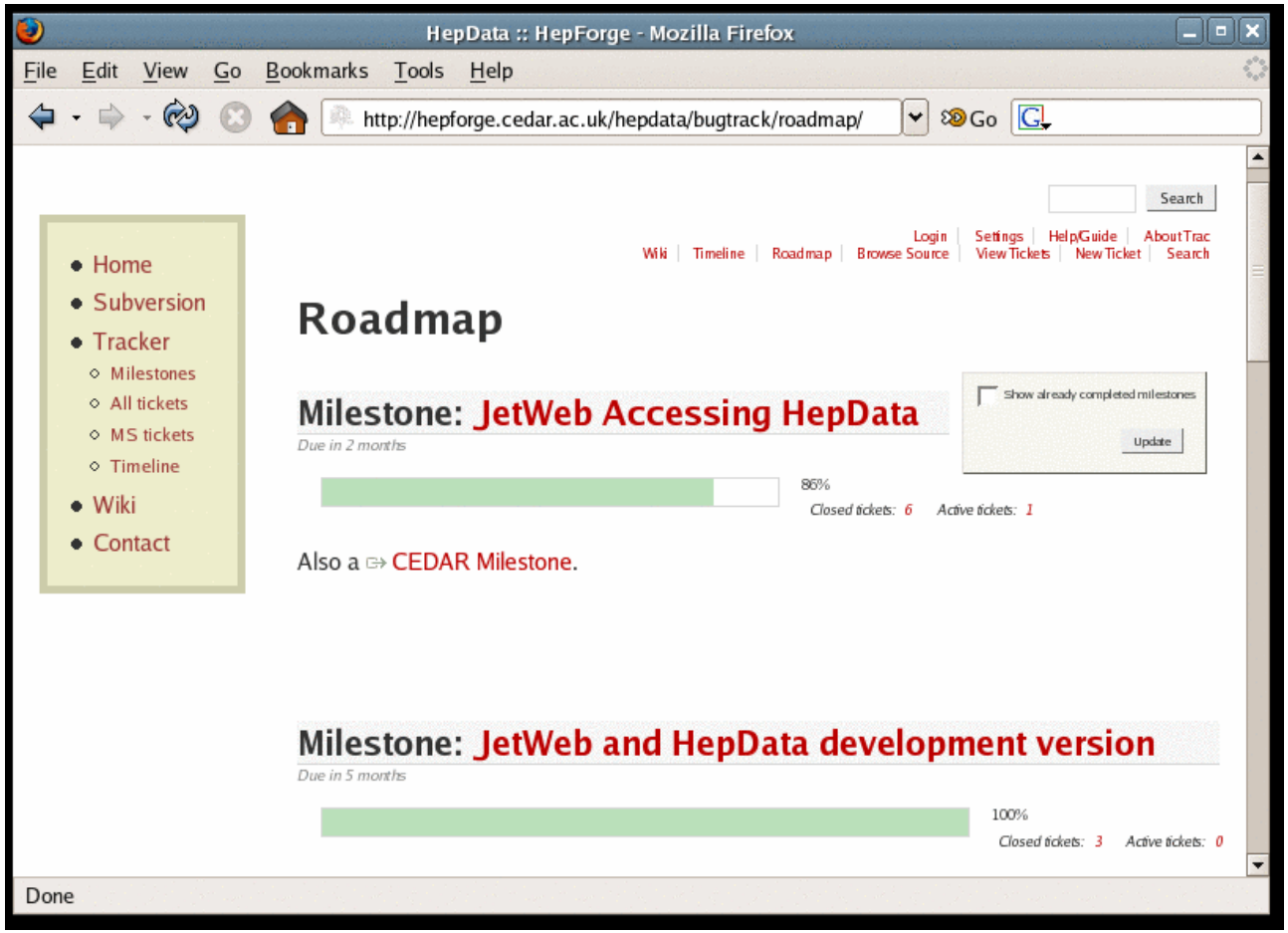


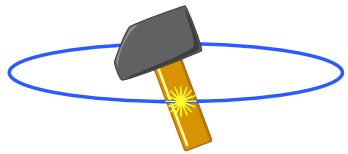


Durham University

HepForge tour

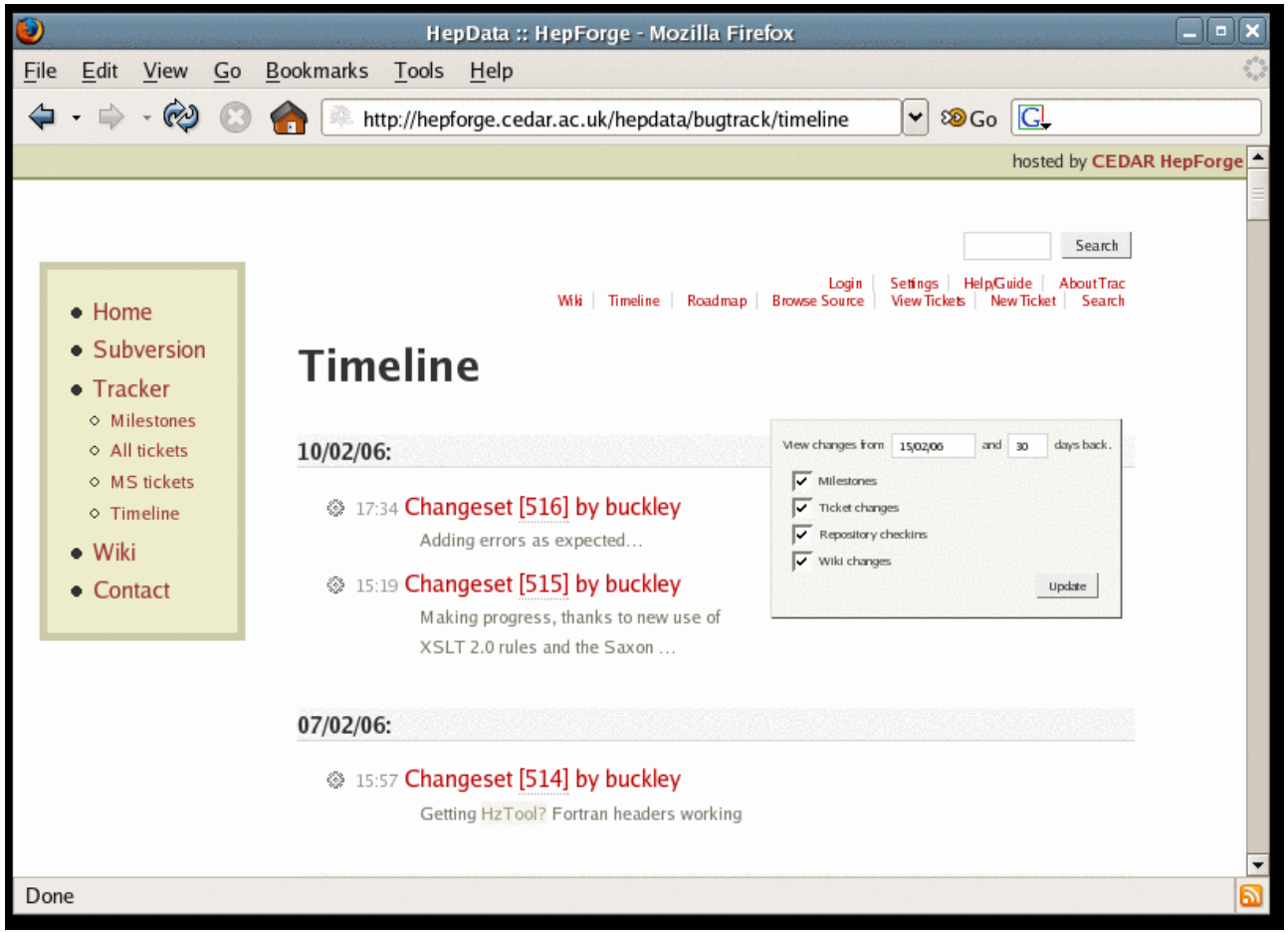
Project bug tracker: milestones

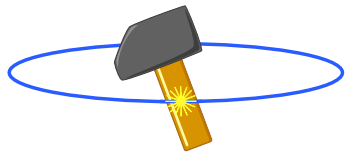




HepForge tour

Project timeline (integrated with SVN)





Durham University

HepForge tour

Project bug listing

HepData :: HepForge - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://hepforge.cedar.ac.uk/hepdata/bugtrack/report/3

Go

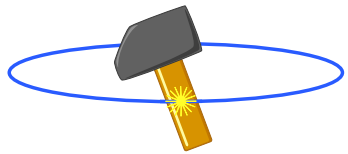
JetWeb Accessing HepData

Ticket	Summary	Component	Version	Type	Owner	Created
#2	Create XSL transformer for data HepML -> HzTool Fortran headers	hepdata	1.0	task	buckley *	01/11/05

Object model and db persistency refactoring

Ticket	Summary	Component	Version	Type	Owner	Created
#42	Refactor the object model / db interaction	model	1.0	task	buckley *	17/01/06
#54	Improve error representation in the object model	model	1.0	task	buckley *	01/02/06
#46	Consider using SQL double for data values	migration	1.0	enhancement	buckley *	21/01/06
#26	Separate HepData model from migration code	hepdata	1.0	defect	buckley *	25/11/05
#53	Use enums for error type	model	1.0	enhancement	buckley *	01/02/06
#56	Separate stat/sys and	model	1.0	enhancement	buckley *	02/02/06

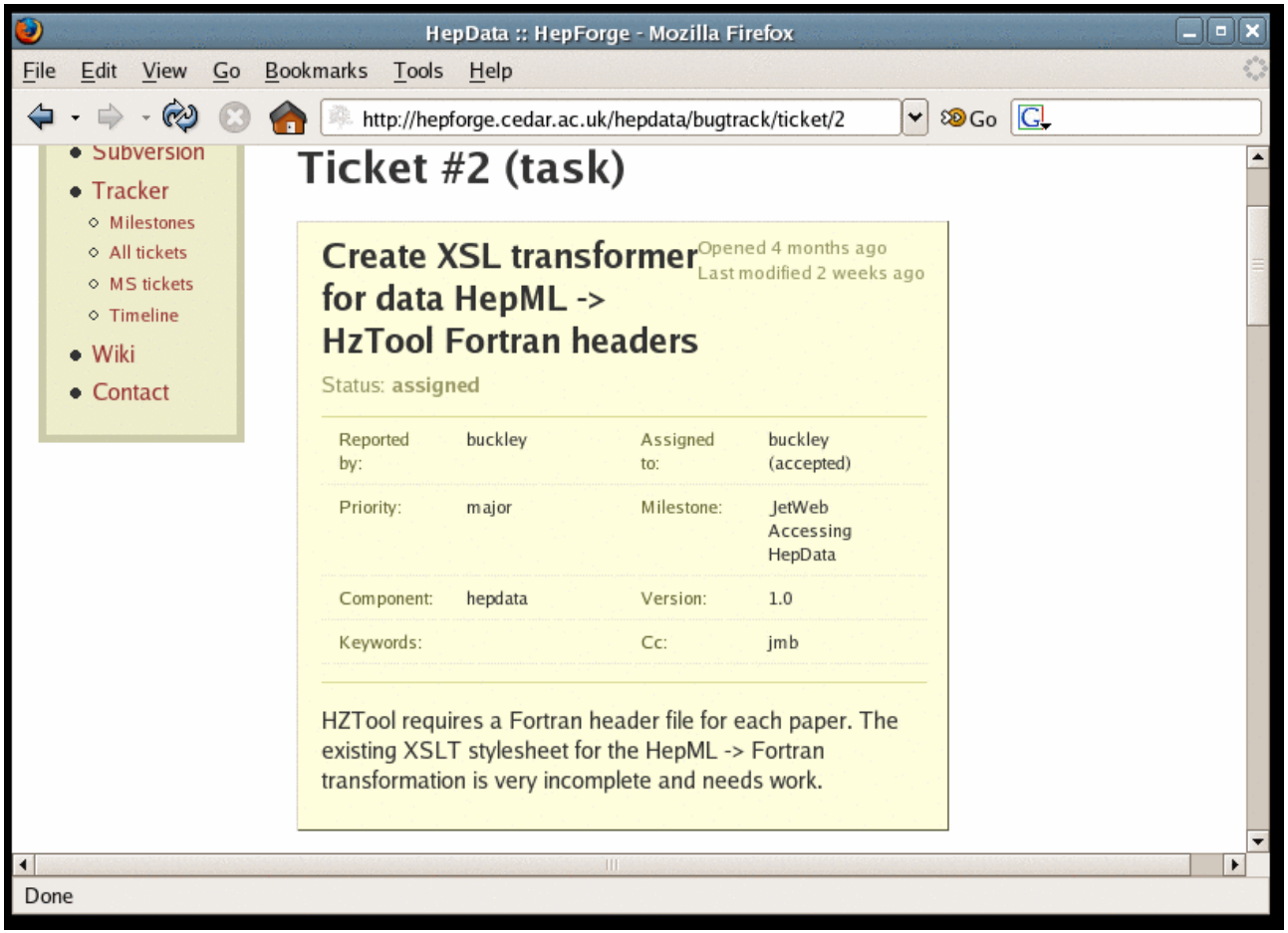
Done

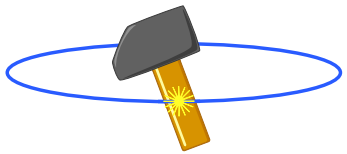


Durham University

HepForge tour

Project bug details (1)

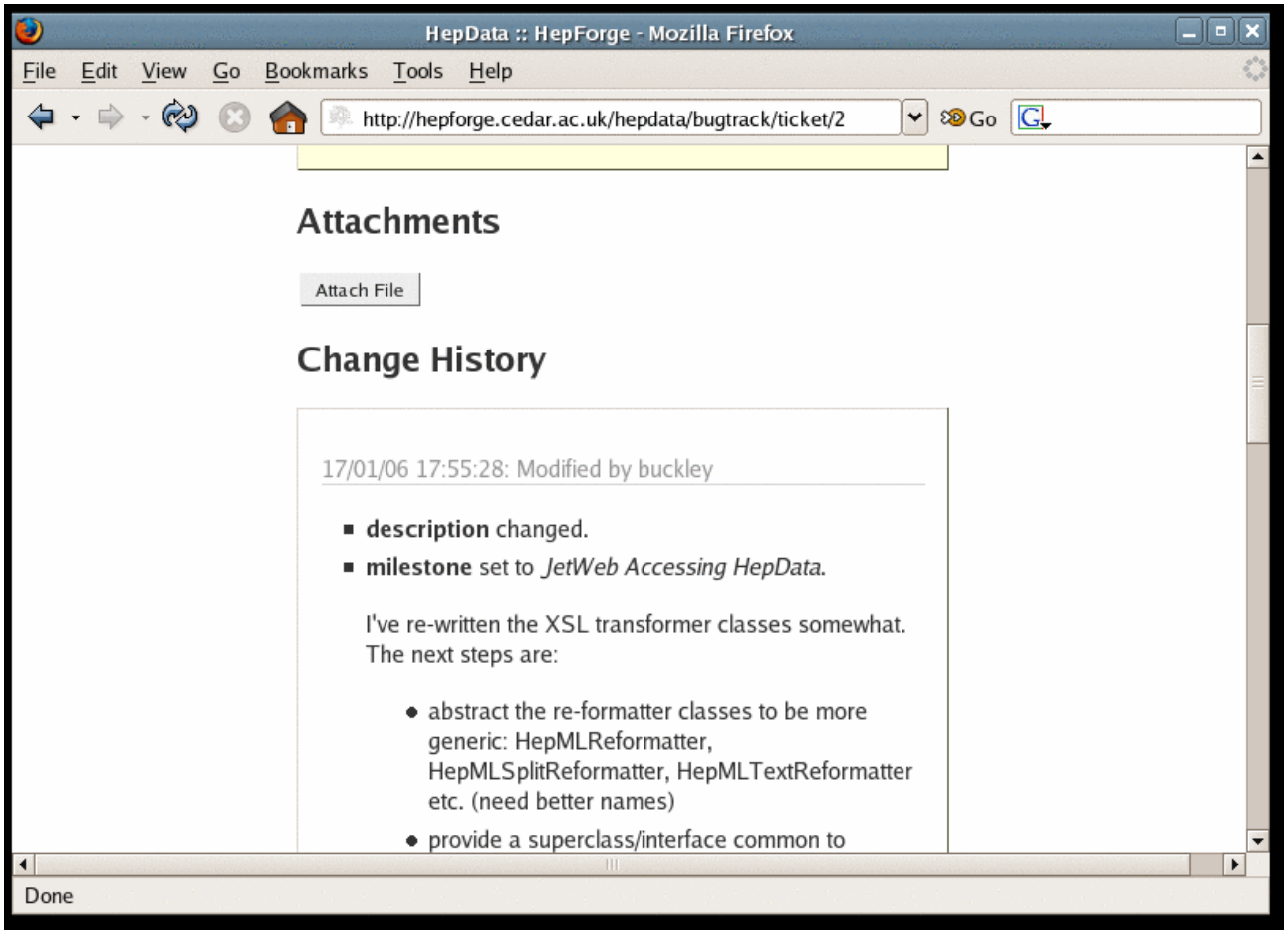


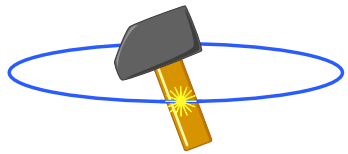


Durham University

HepForge tour

Project bug details (2)

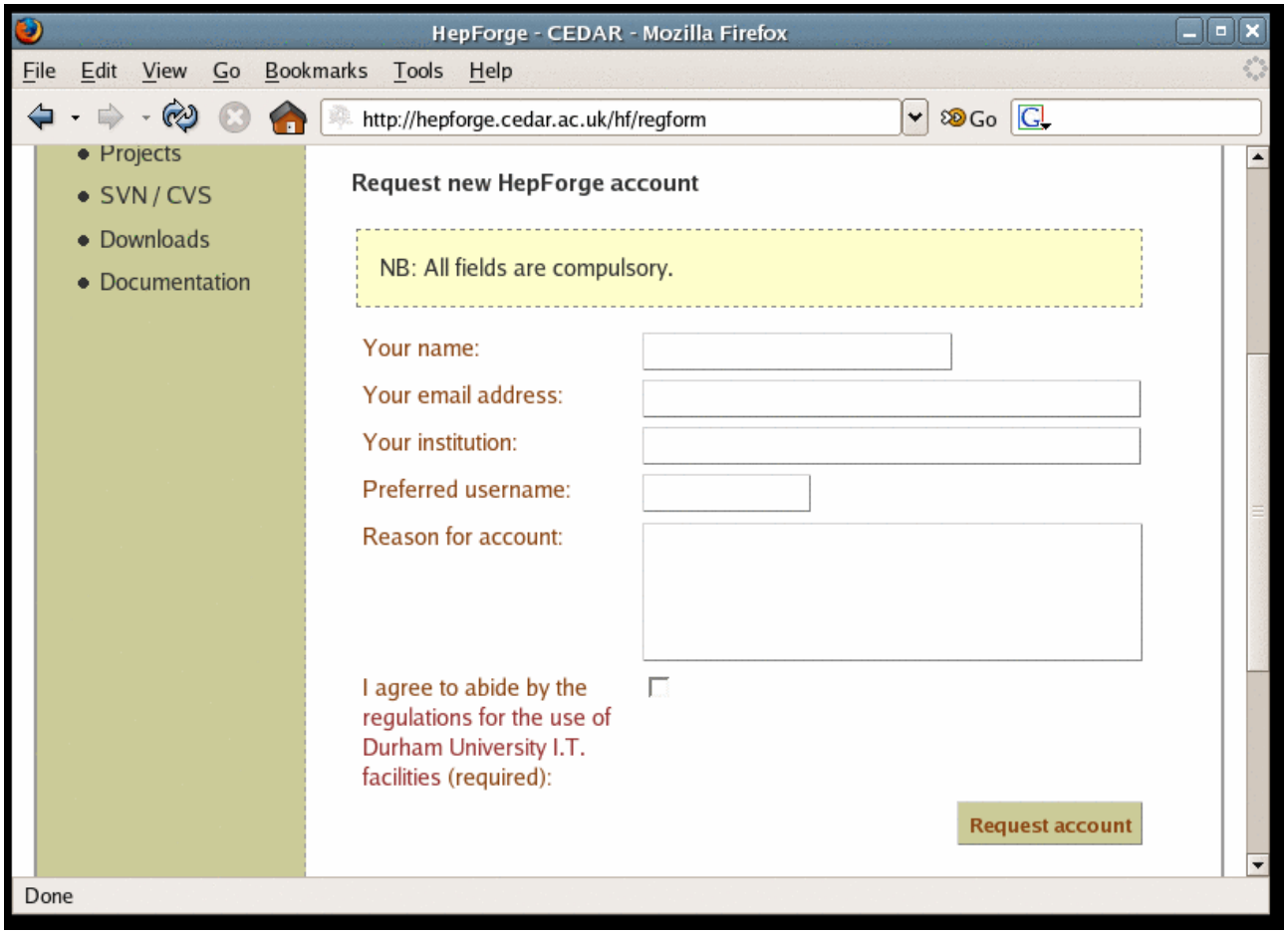


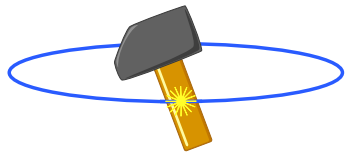


Durham University

HepForge tour

The HepForge registration form! Think about it. . .





Durham University

Registration

Please think about using HepForge!

- Current users include: Herwig++, ThePEG, Pythia, LHAPDF, RunMC, FastNLO, Jimmy, KtJet...
- Plus all the CEDAR sub-projects, of course!
- Requirements:
  - has to be for a re-useable HEP project
  - not for processor-intensive use
  - commitment to document and support your project
  - encouraged to use standard build proceduresetc. We can and will help with this!

Visit **`http://hepforge.cedar.ac.uk`** to register

Durham University

Summary

Durham University

Summary

- CEDAR is primarily a generator tuning system, combining JetWeb and HepData
- Aim is that any expt MC configuration should be "CEDAR-blessed" to be considered trustworthy
- Timescale: must be ready for LHC data

- HepForge is available for HEP software development now!
- Eventually will be used to implement the HepCode system
- Feedback has all been very positive: system is powerful but very easy to use
- Consider HepForge for *your* re-useable HEP code!

Durham University

HepForge backend (1)

For the interested...

- Python-based with shell scripts for acc. management
- Various Subversion things, e.g.
  - re-write of CVS: almost complete UI compatability
  - support for symlinks, metadata, dirs, **`mv`**/**`cp`**!
  - anon read access over HTTP; rw dev access over SSH
  - use it! Migration from CVS is easy.
- ViewVC with multi-site hack (also via SSI)

Durham University

HepForge backend (2)

For the interested...

- Trac bug tracker and wiki:
  - SQLite backend and SVN integration
  - Moin-compatible wiki
  - excellent tool!
- Web system with post-processing scripts
  - HTML Tidy
  - transparent header and footer handling
- Download manager (personalisable via SSI)

Durham University