CEDAR, part II

HepData, HepML and HepForge

Andy Buckley

Institute for Particle Physics Phenomenology
Durham University, UK

HERA-LHC Workshop 2006, CERN, 2006-06-08

Durham University

Outline

Durham University

Re-introduction to CEDAR

CEDAR: Collaborative e-Science Data Analysis Resource

I expect Jon Butterworth will have already said this, but...
CEDAR is an e-Science project with several sub-projects:

- JetWeb: Monte Carlo generator tuning
- HepData: archival of published experimental data
- HepML: XML 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

HepData — the HEP reaction database

Durham University

HepData's history

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

Durham University

HepData in CEDAR (1)

HepData is undergoing active development:

- Migration to relational MySQL database
- Database sanitising e.g. uniform units…
- Database improvements e.g. axis-level properties, more searchable quantities…
- Using HepML for I/O: XSLT transformations to data formats like HTML and AIDA

Durham University

HepData in CEDAR (2)

And more...

- Java object model for data abstraction: interface for JetWeb
- Java/JSP-based Web front-end using Apache Tomcat
- JBoss Hibernate used to abstract the object-relational binding
- Similar system planned for XML persistency
- Grid authentication for expts. to submit data directly (modulo sanity checking)

Durham University

HepData HepML

HepML is a set of XML-based data format defns. for HEP

- XML representation for generator configs and HepData data sets (and more to come?)
- (I'll just mention the HepML data schema here)
- To be backed up with Python, Java (and C++?) interfaces
- Version for comment has been released. Please check it out: **`http://hepforge.cedar.ac.uk/hepml/`**
- Experiments will submit data to HepData: we need your comments on what HD and HepML should allow you to do. Abstracted interfaces to HepML (e.g. ROOT routine)?

Durham University

HepML fragment (1)

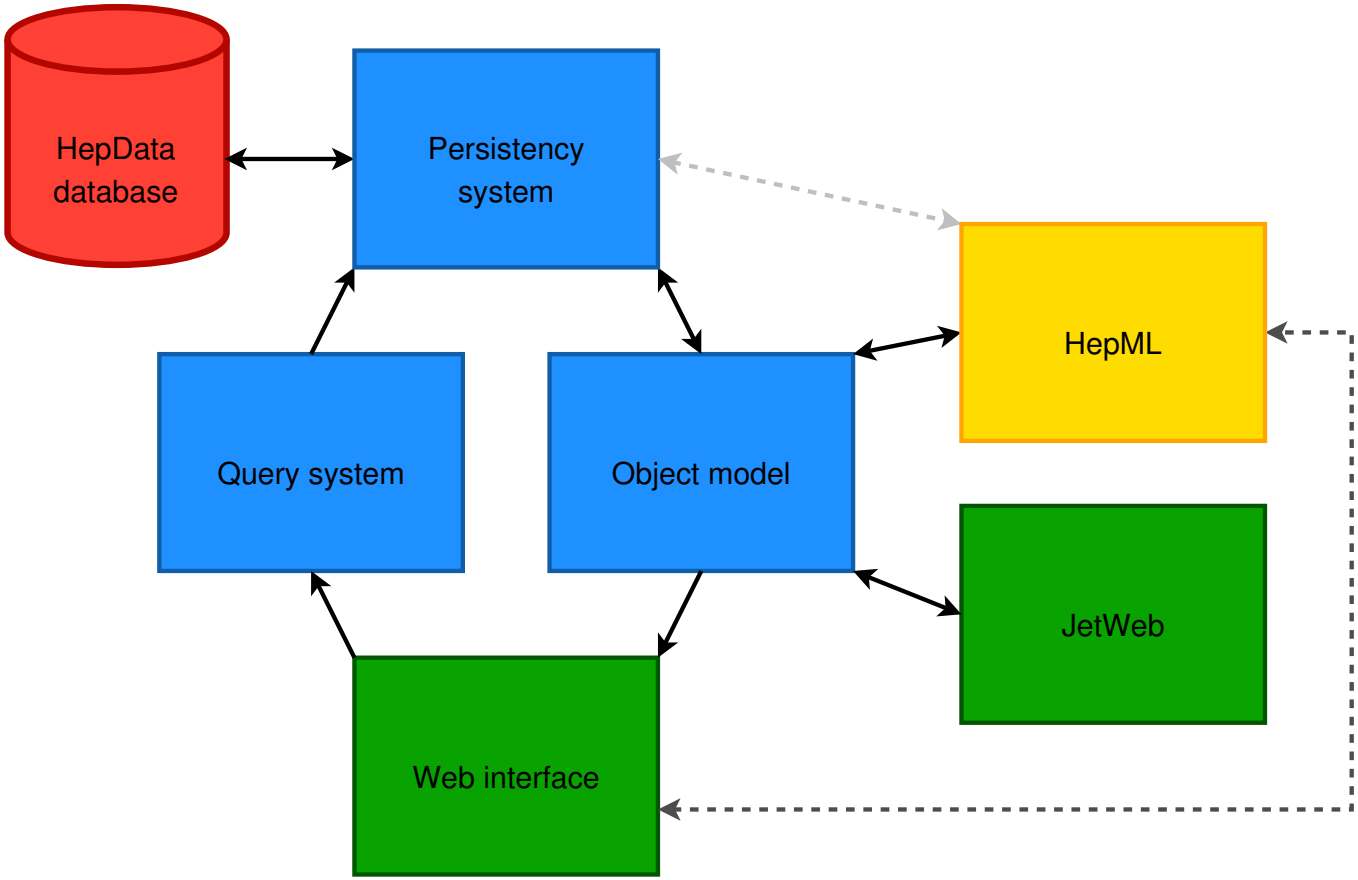
```
<?xml version="1.0" encoding="UTF-8"?>
<hepml xmlns="http://www.cedar.ac.uk/hepml/hepdata/0.1/">
  <data timestamp="2006-04-07 13.09.27">
    <paper irn="3326047" paperId="3552">
      <dataset datasetId="1">
        <comment>
          FITTED, BACKGROUND SUBTRACTED, PEAK OMEGA CROSS SECTION,
          ...
        </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>
```

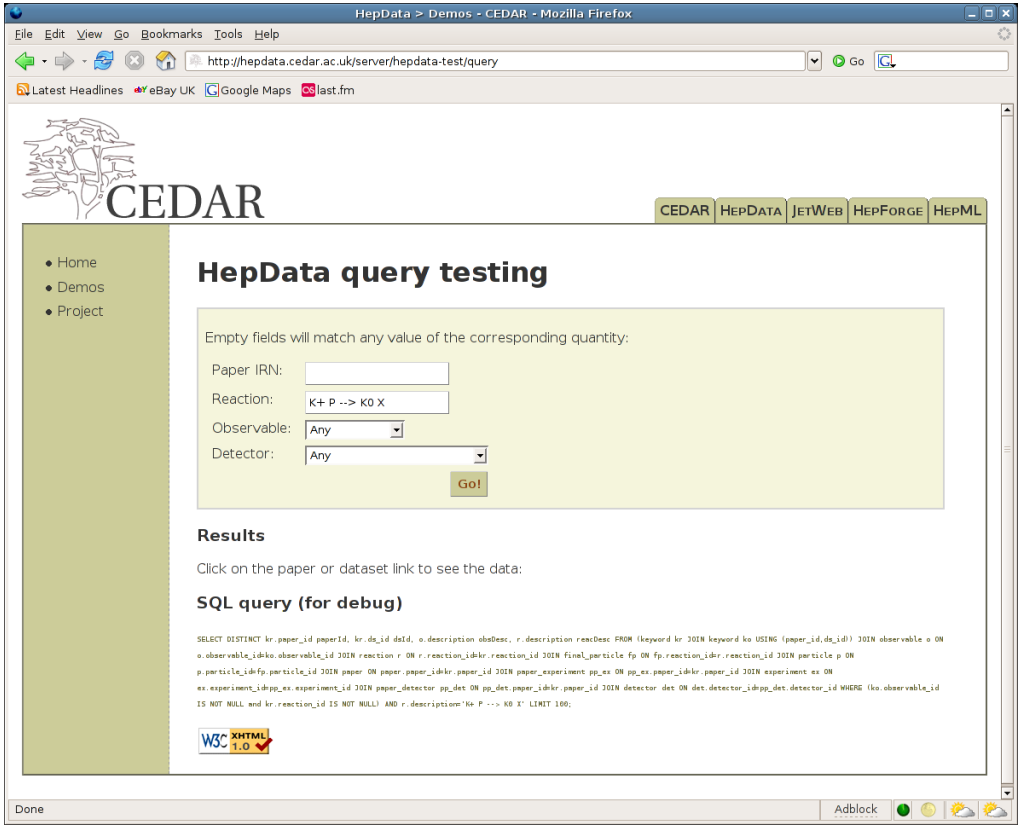
Durham University

HepData behind the scenes



HepData demos (1)

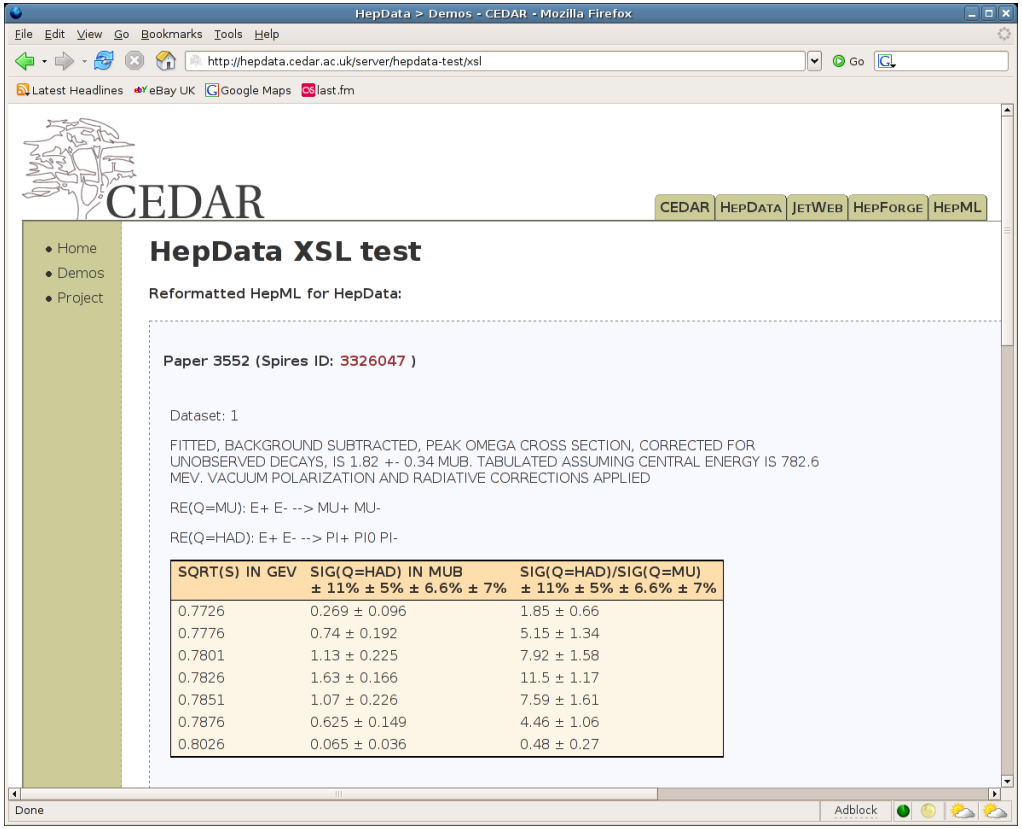
Query interface (breaking the rules — spot the SQL!)



Durham University

HepData demos (2)

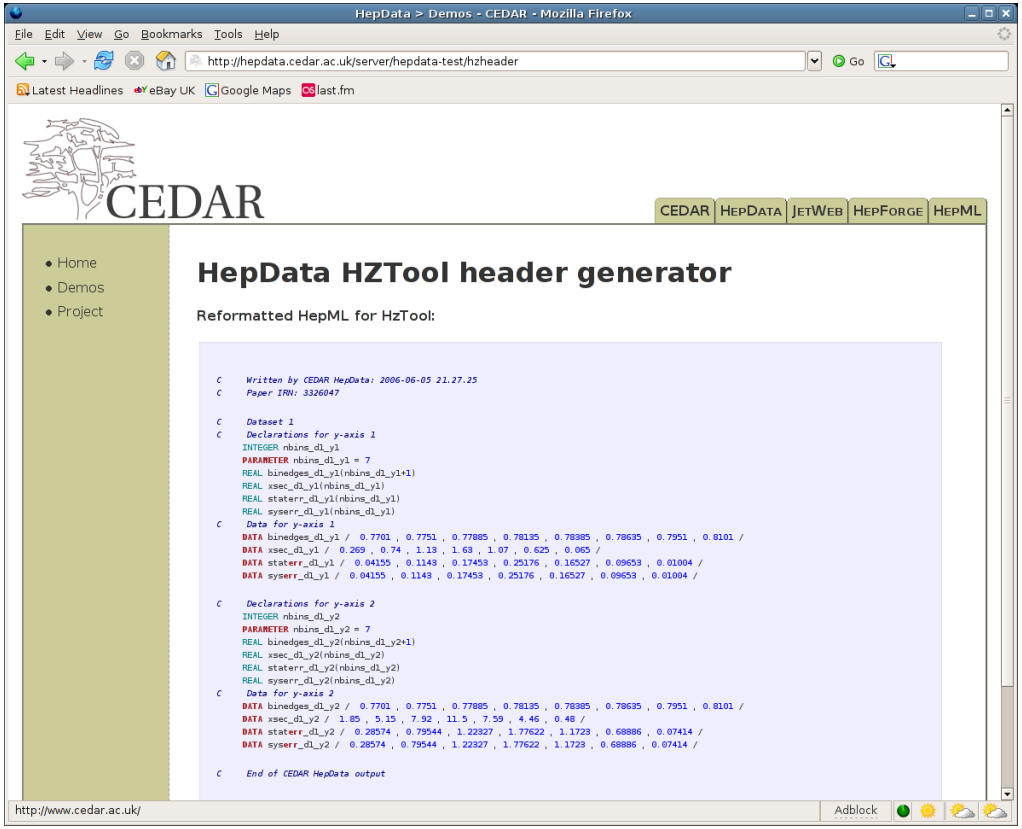
HepData → HepML $\xrightarrow{XSL}$ HTML



Durham University

HepData demos (3)

HepData → HepML $\xrightarrow{XSL}$ HZTool Fortran



Durham University

HepData summary

- HepData re-engineering in progress:demos at **`http://hepdata.cedar.ac.uk`**
- Not everything is finalised yet...
- Your chance to make sure that it can do everything you want!
- HepML is also incomplete: metadata is not yet fully handled, for example
- Do we need to provide special data set support like correlation matrices?
- Get involved!

Durham University

HepForge — a lightweight development environment for HEP

Durham University

Software engineering for small HEP projects

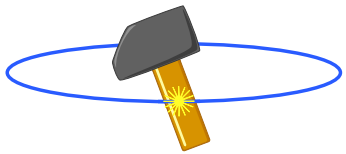
- Everyone has written code that might be re-used...
- But: "energy barrier" to tidying up / managing it as a project
- Small projects don't have resources for nice things like:
  - Use std. build/install systems e.g. autotools, libtool
  - Independence of build/run environment (e.g. `/cern`!)
  - "Formal" quality control, feedback and bug tracking
  - Version control systems (Subversion, CVS)

HepForge aims to reduce this barrier and encourage more publically released, well-developed HEP code

Durham University

HepForge

- Collaborative development system for HEP
- Online at **`http://hepforge.cedar.ac.uk`**
- Features including:
  - Shell access with full set of dev tools
  - Web space (with several convenient features)
  - Subversion/CVS version control (+ Web viewers)
  - Wiki and bug tracker
  - Mailing lists for developers and users
  - Downloads manager
- We've done the "boring bits"!
- Many users already: Herwig++, Pythia6, LHAPDF... and ∼25 others



Durham University

Who should be interested in HepForge?

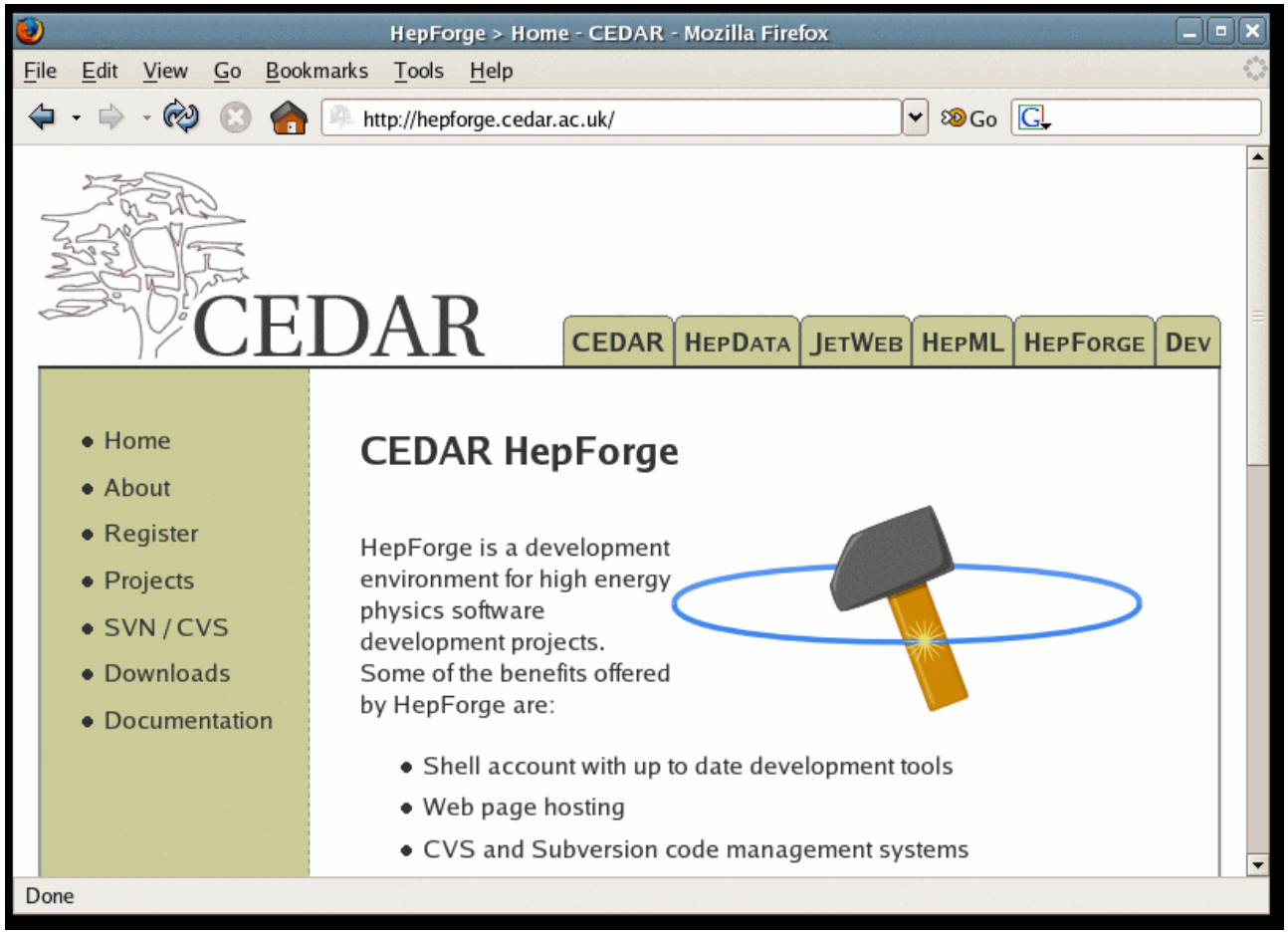
- Who's it for? Small–medium size projects
- *Probably* not experiments! But not necessarily...
- *You* should:
  - Stand-alone code preferred, since it aids re-use
  - Use standard methods, e.g. support `make install`
  - Intention to document properly
  - Support users
  - (General responsible development, essentially!)

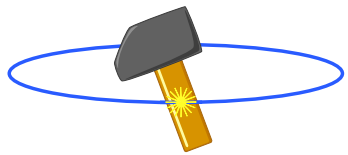
In return, HF provides powerful software development tools with a minimal learning curve

Durham University

HepForge tour

Home page

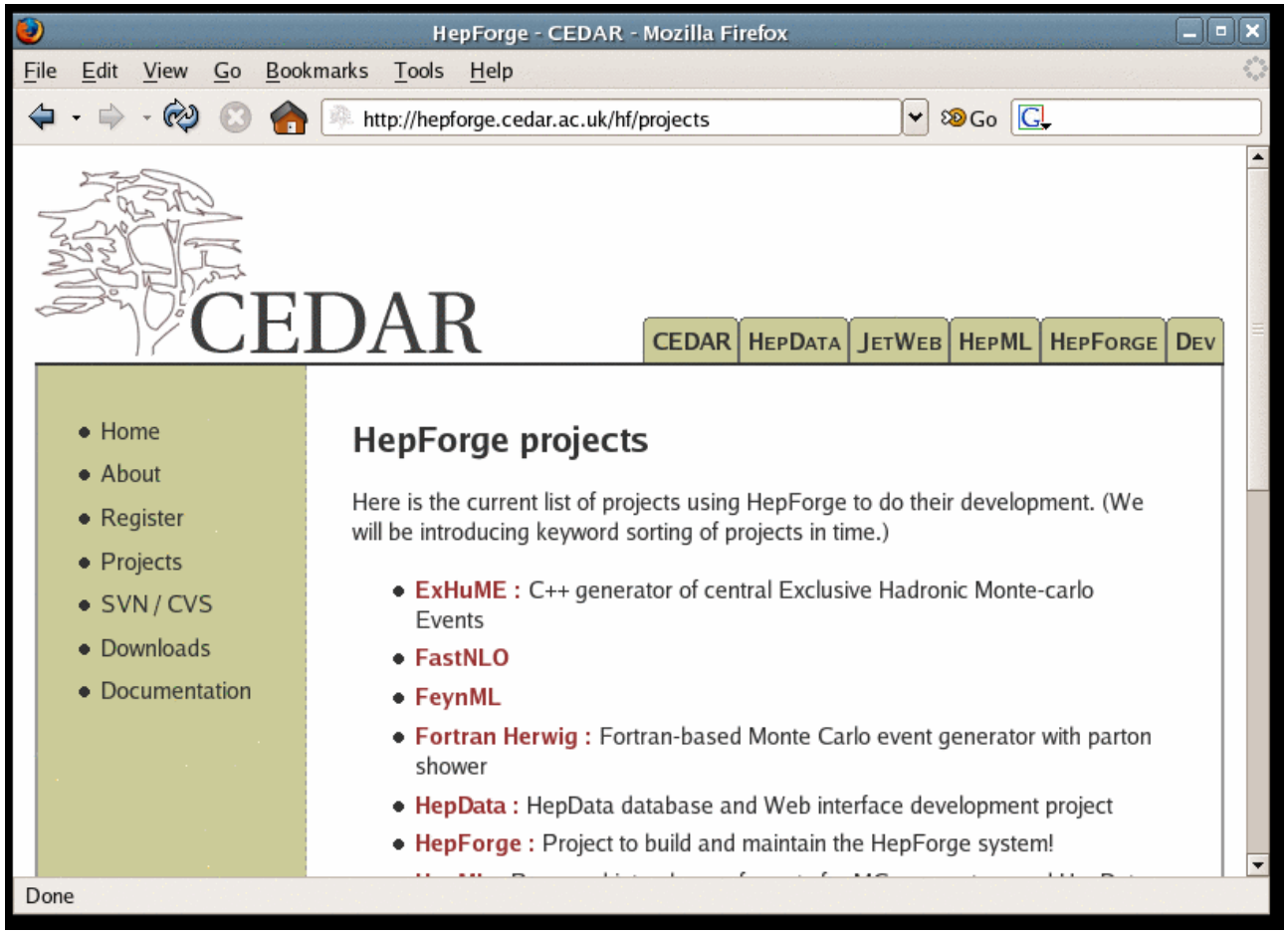


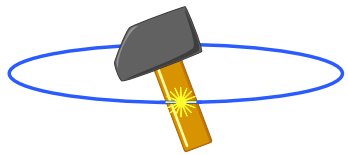


Durham University

HepForge tour

Project list (1)

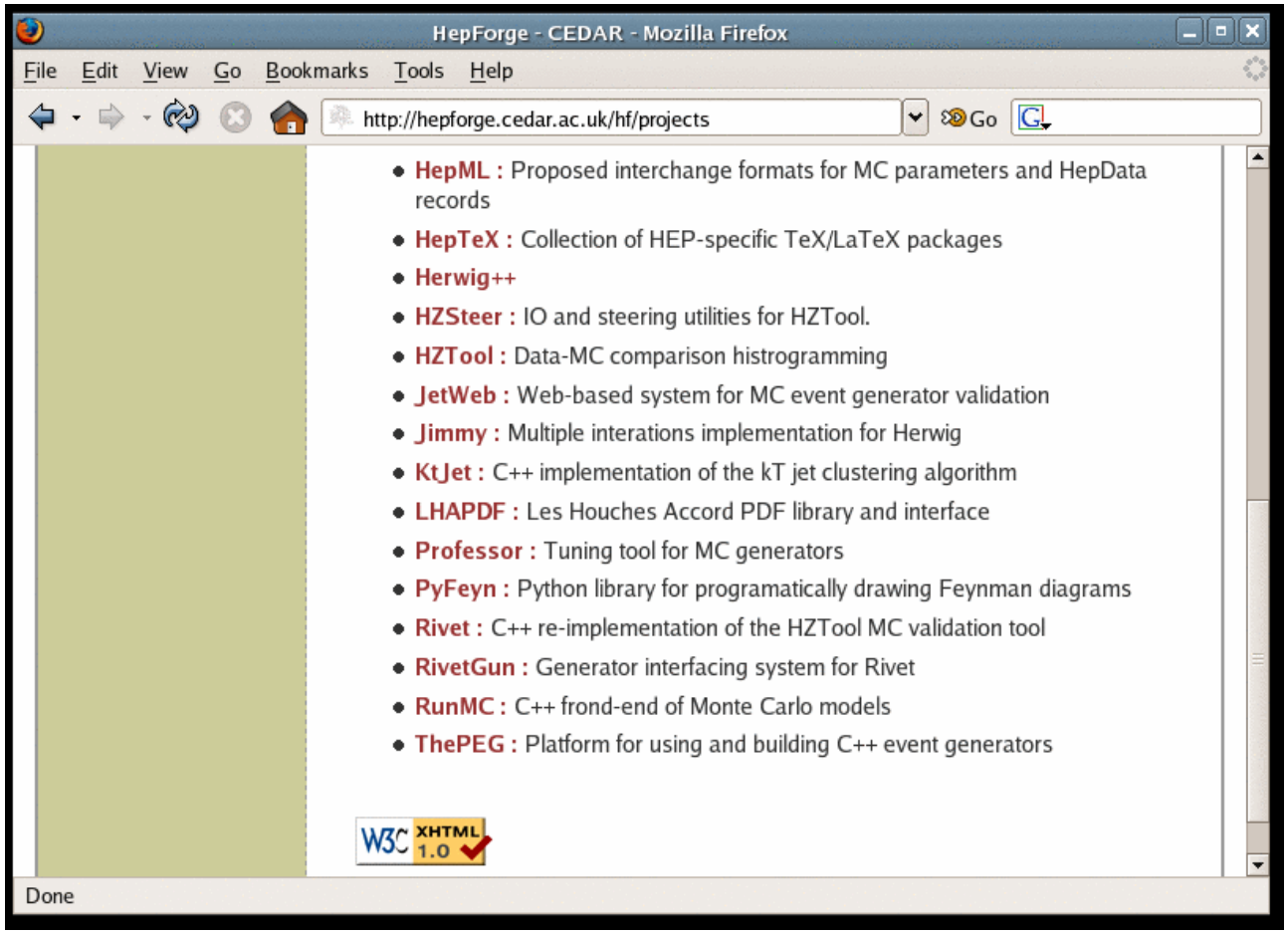


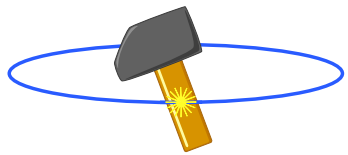


Durham University

HepForge tour

Project list (2)

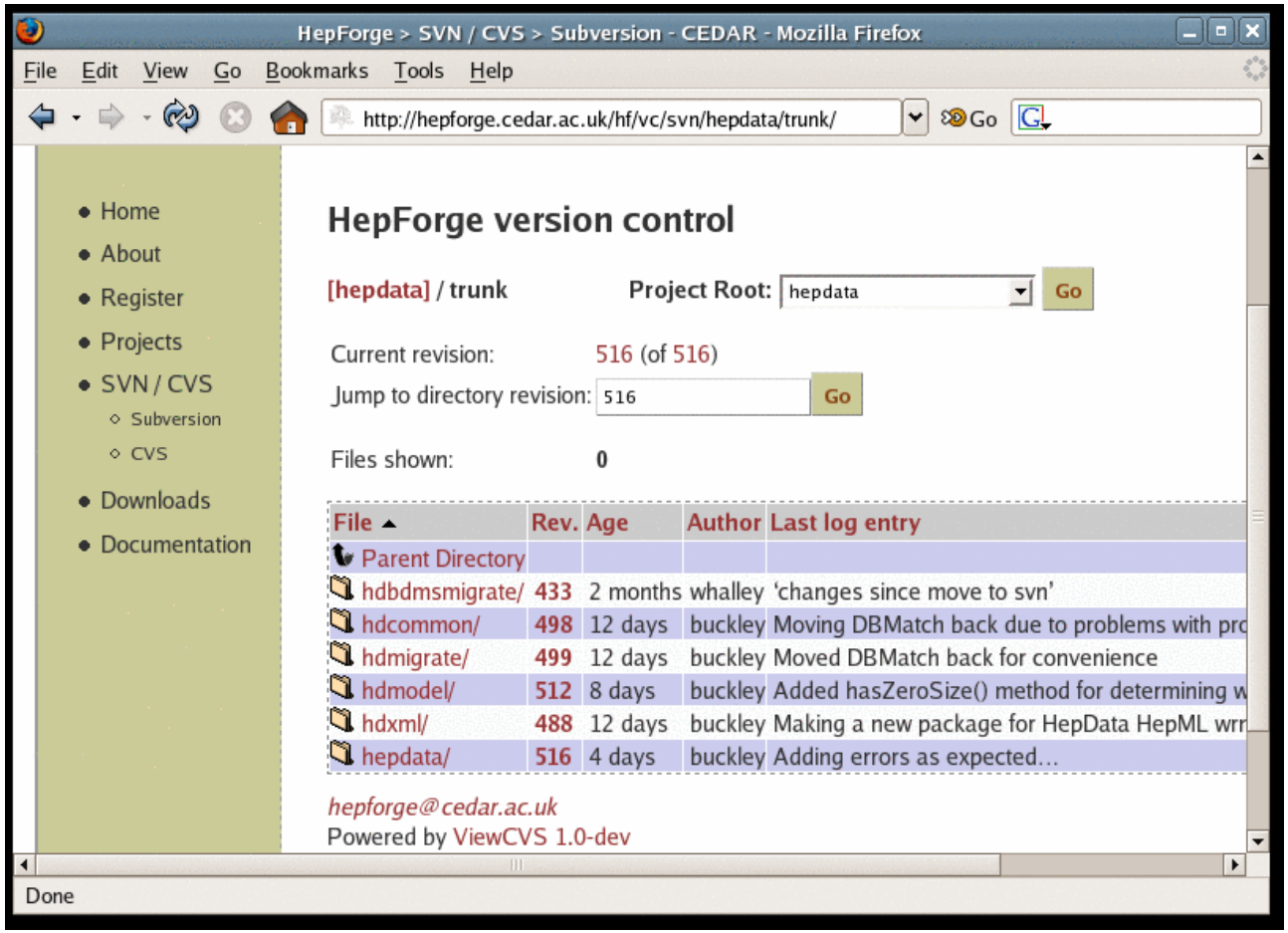


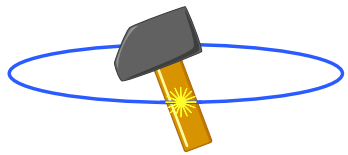


Durham University

HepForge tour

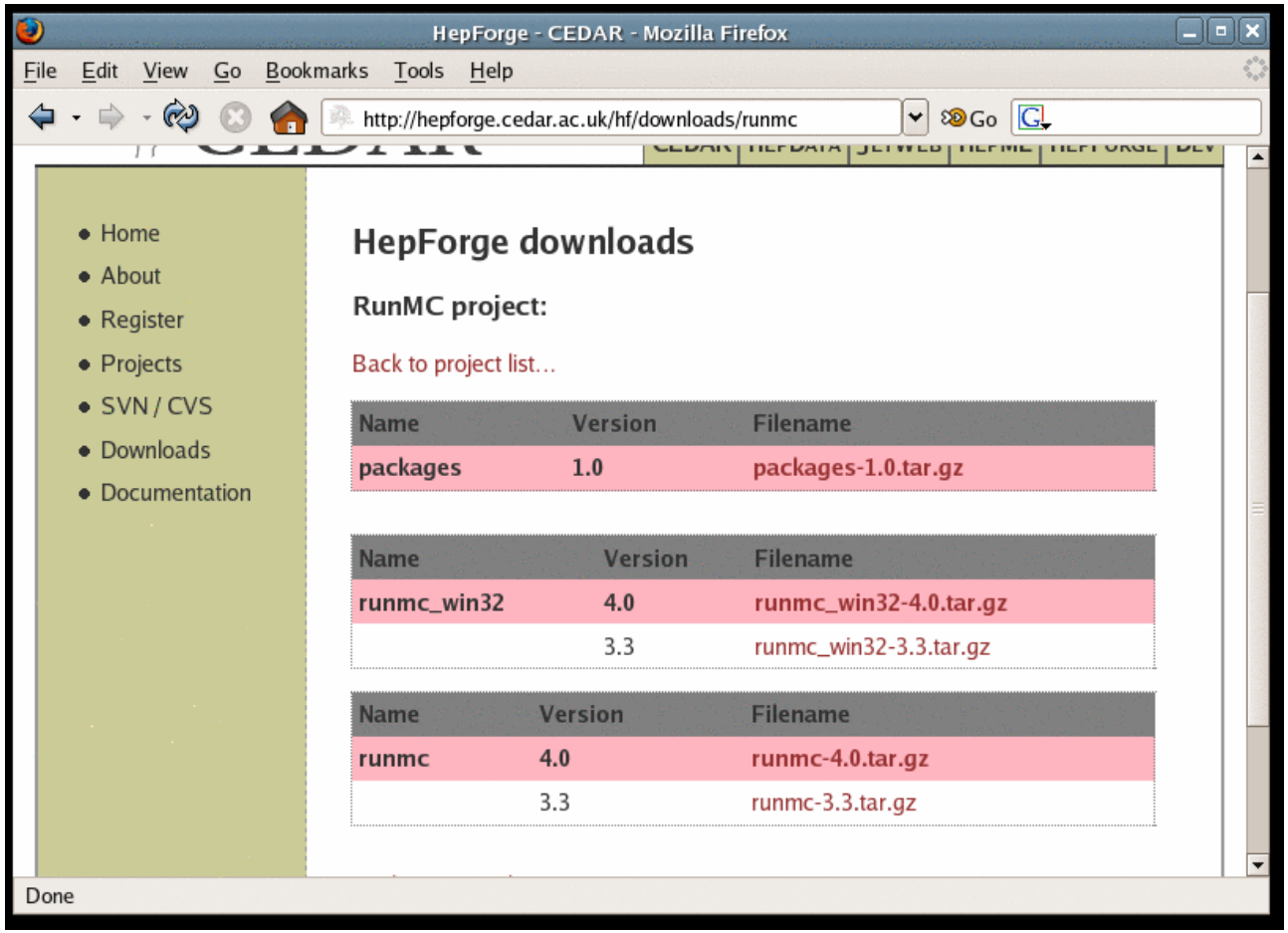
Project VC listing

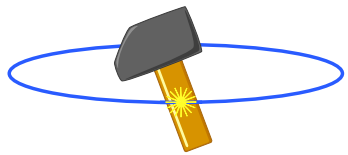




HepForge tour

Project downloads listing (also personal copy via SSI)

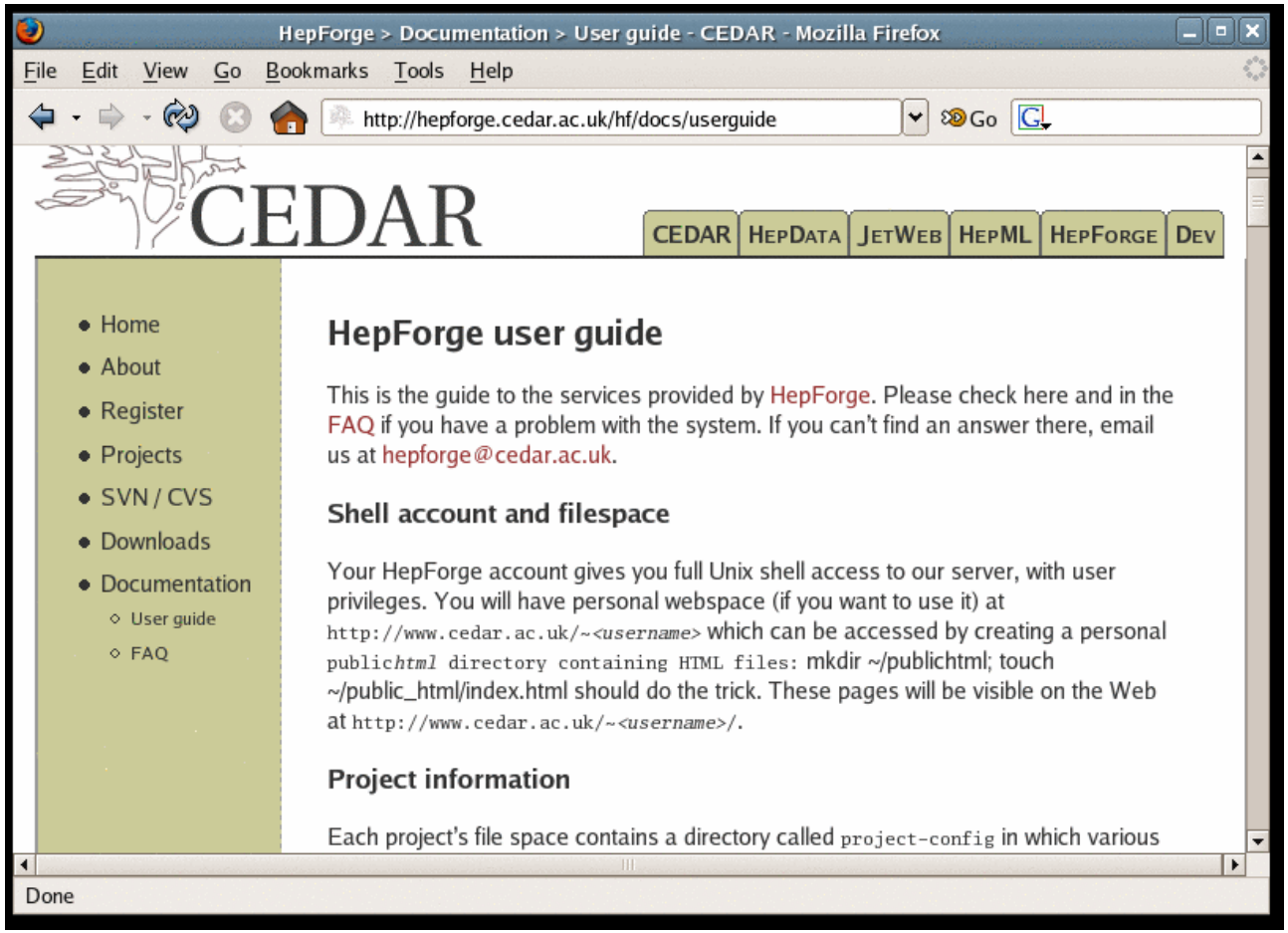


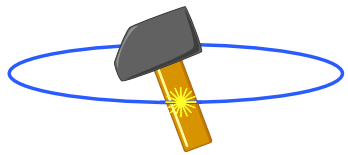


Durham University

HepForge tour

Documentation: user guide (note scrollbar!)

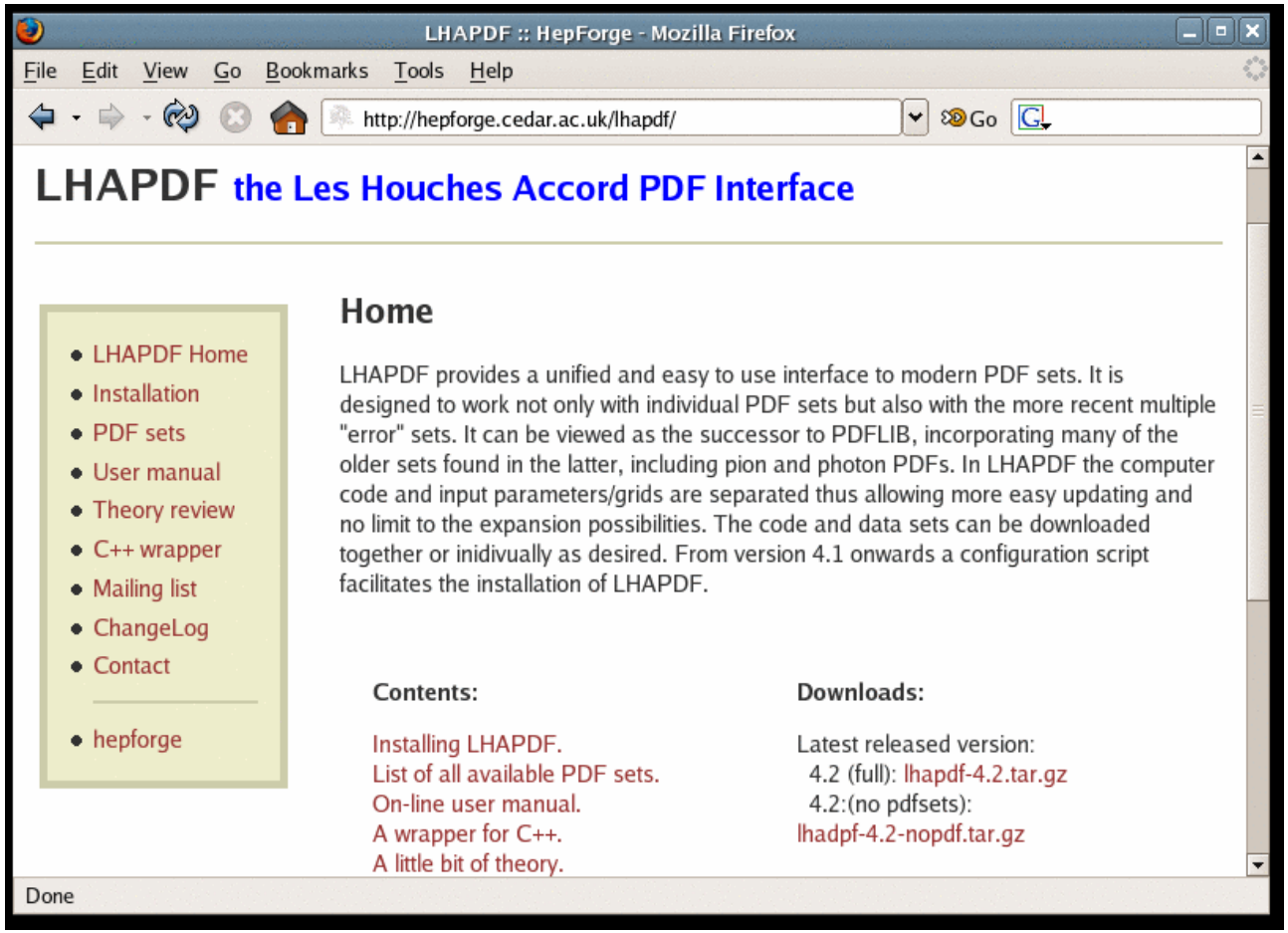


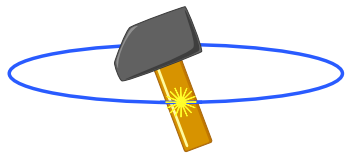


Durham University

HepForge tour

Project Web page: LHAPDF

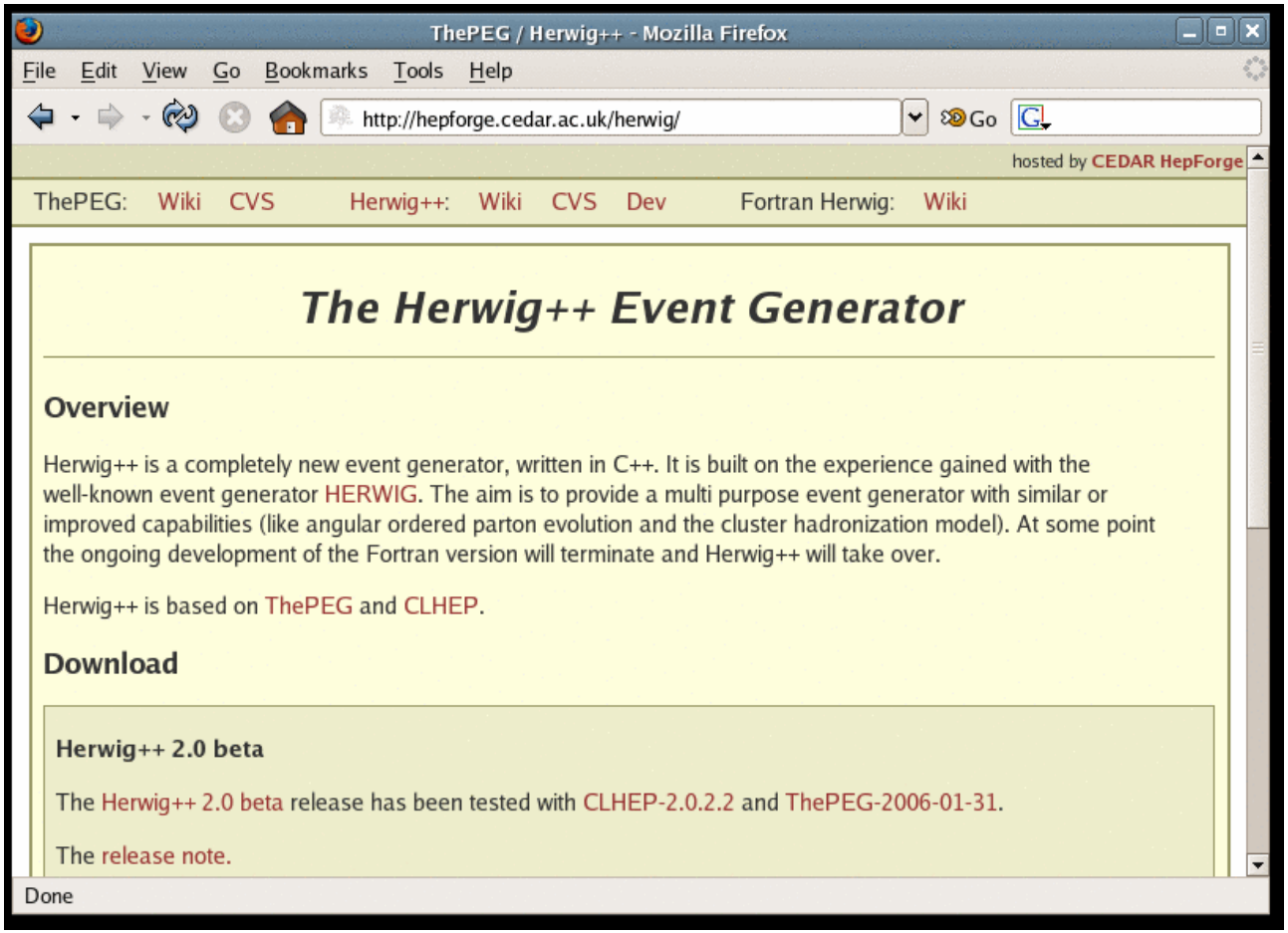


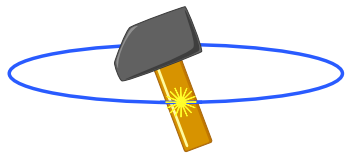


Durham University

HepForge tour

Project Web page: Herwig++

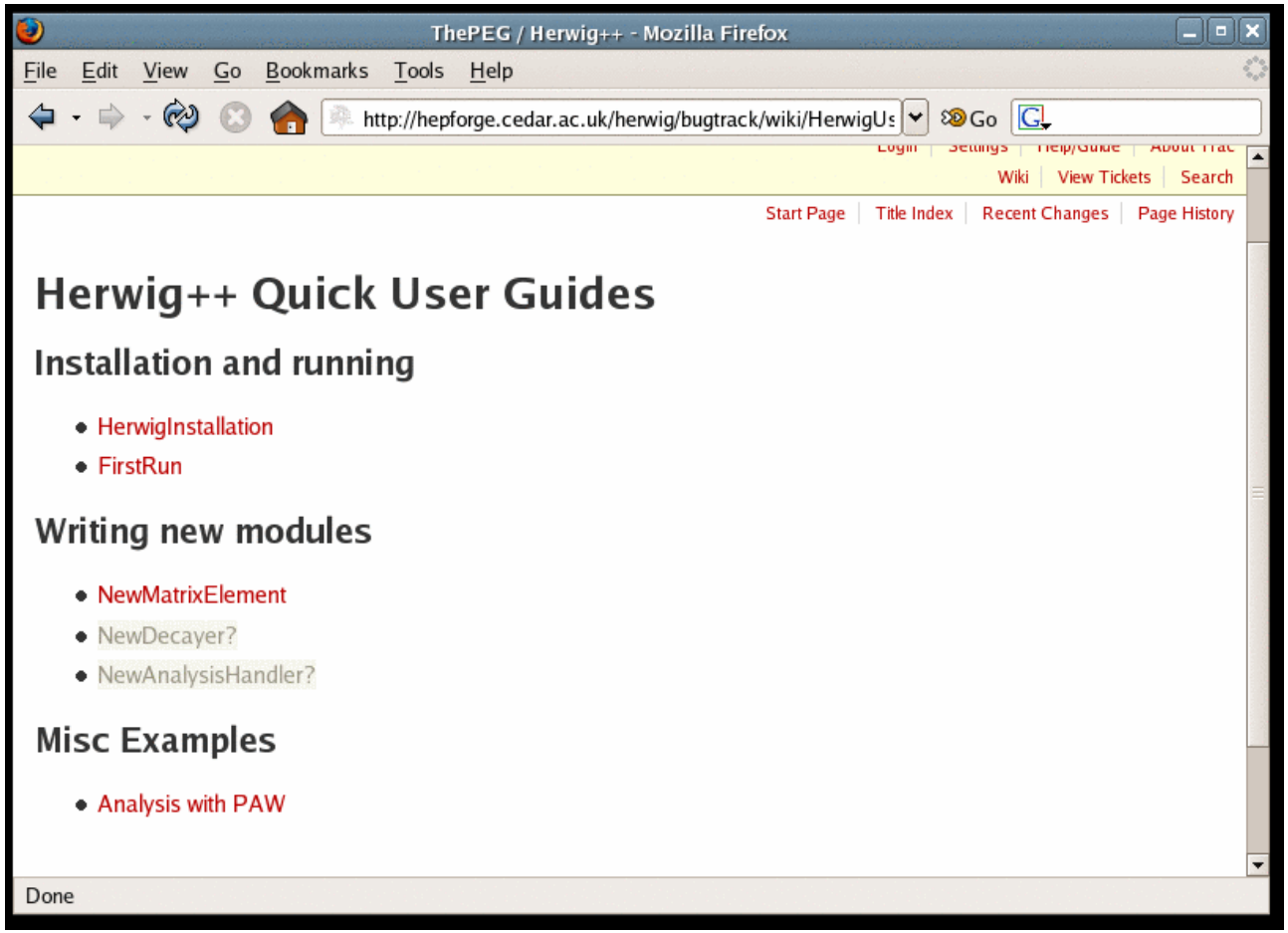


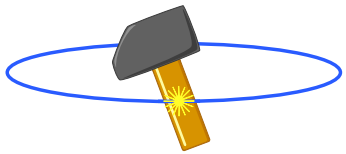


Durham University

HepForge tour

Project wiki

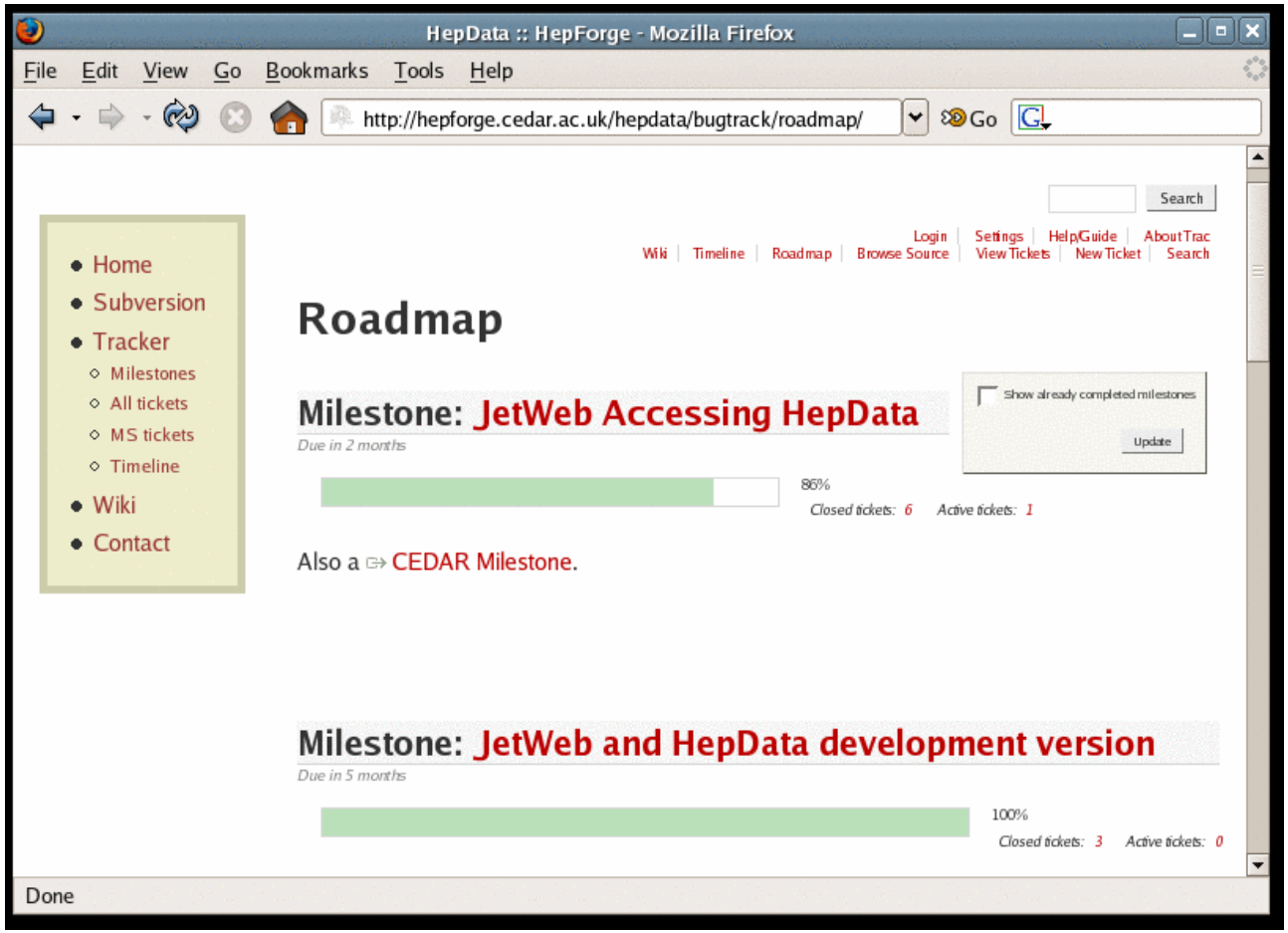


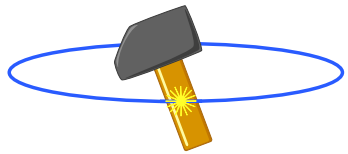


Durham University

HepForge tour

Project bug tracker: milestones

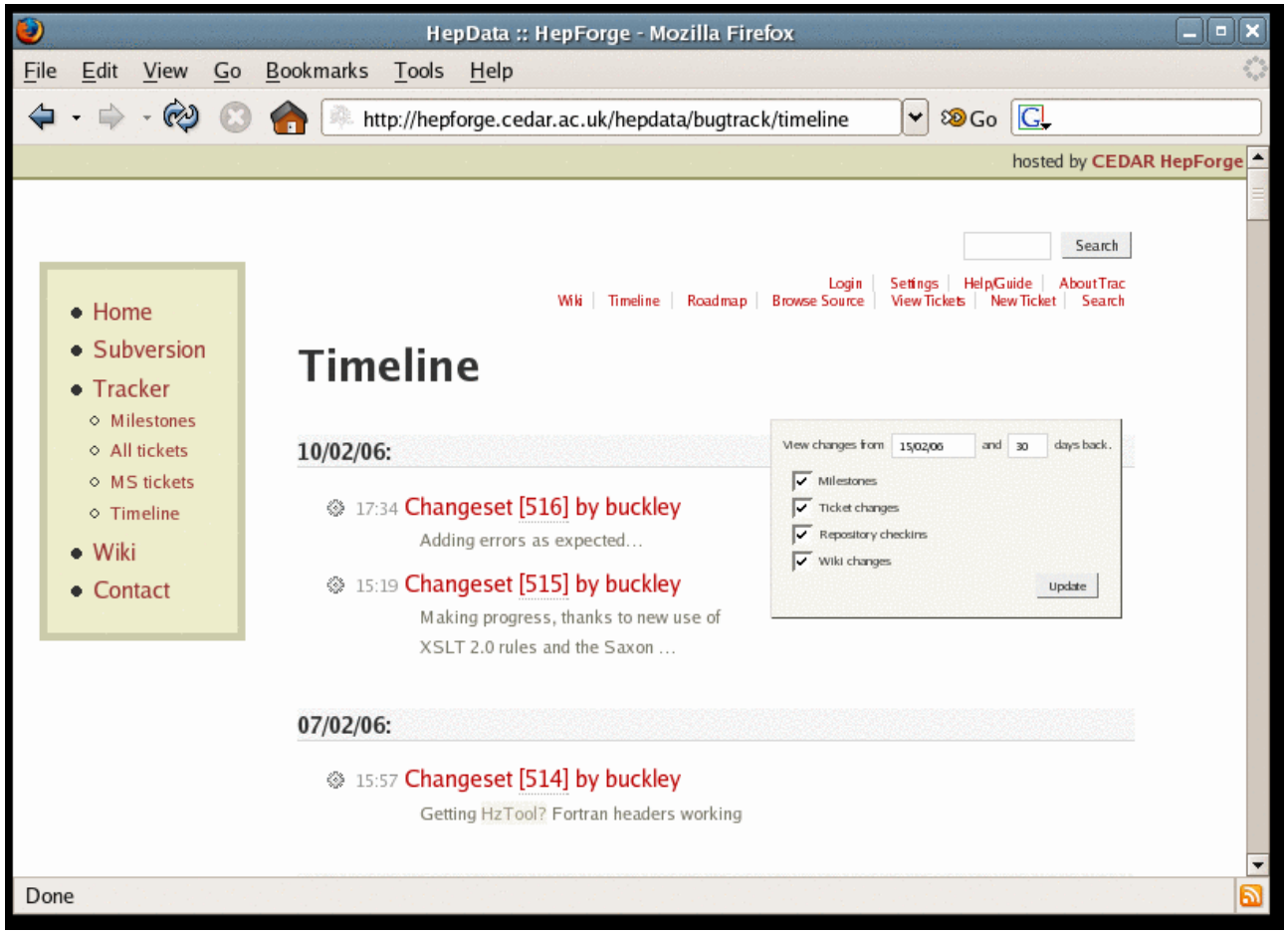


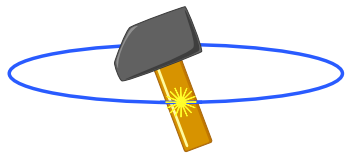


Durham University

HepForge tour

Project timeline (integrated with SVN)

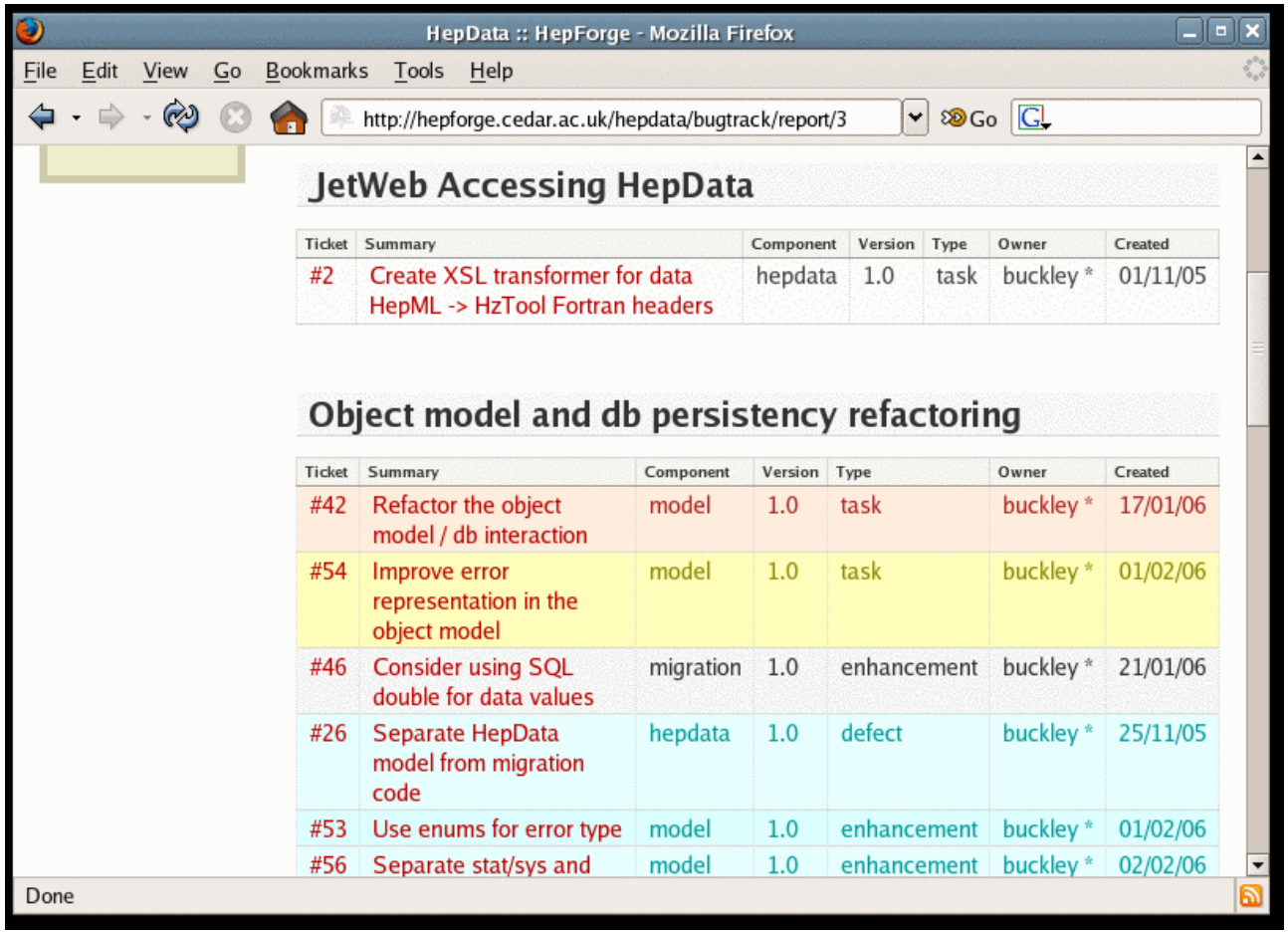


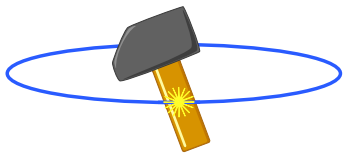


Durham University

HepForge tour

Project bug listing

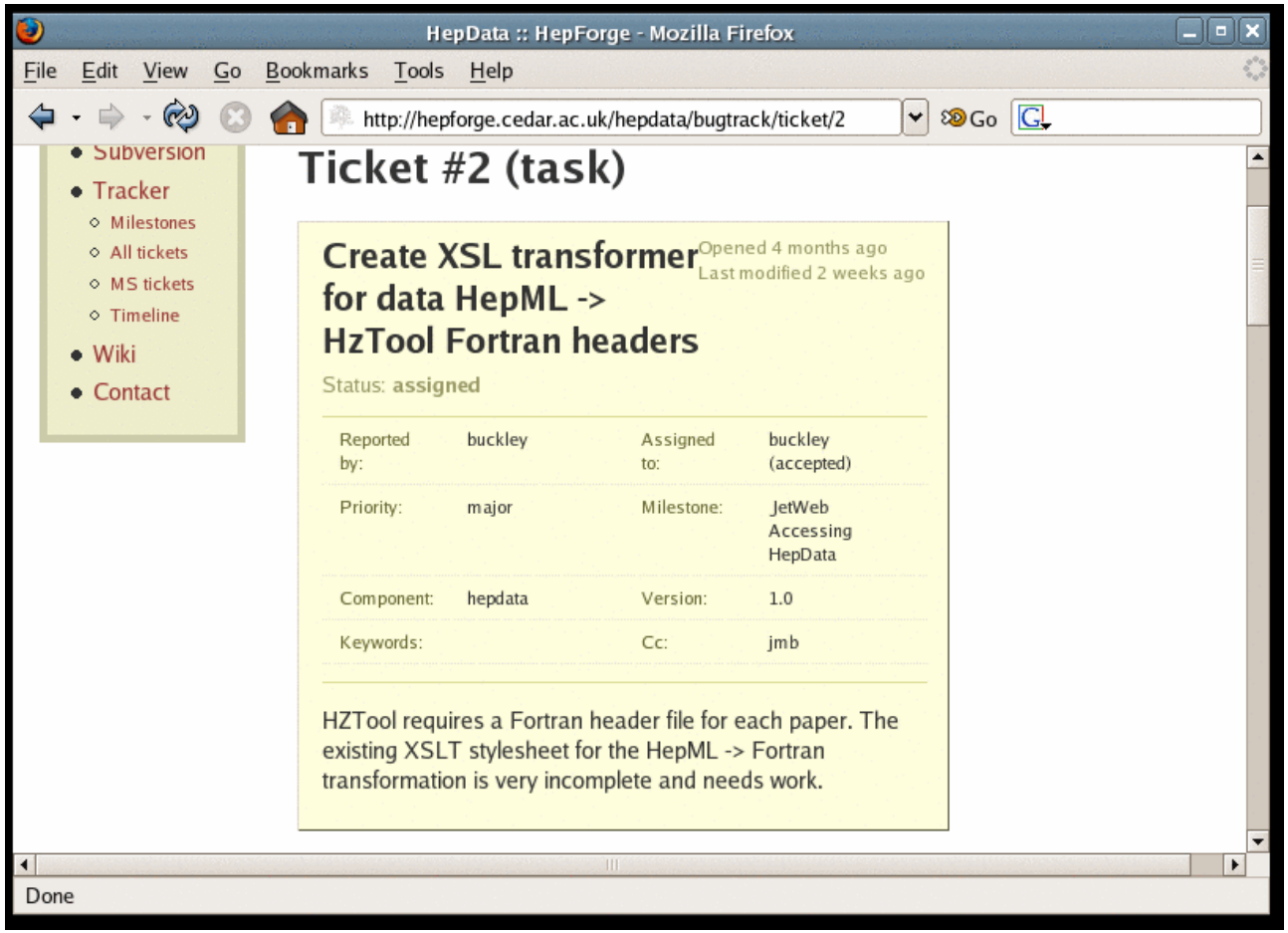


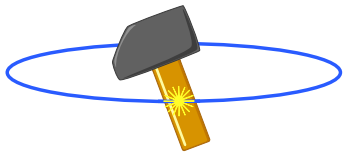


Durham University

HepForge tour

Project bug details (1)

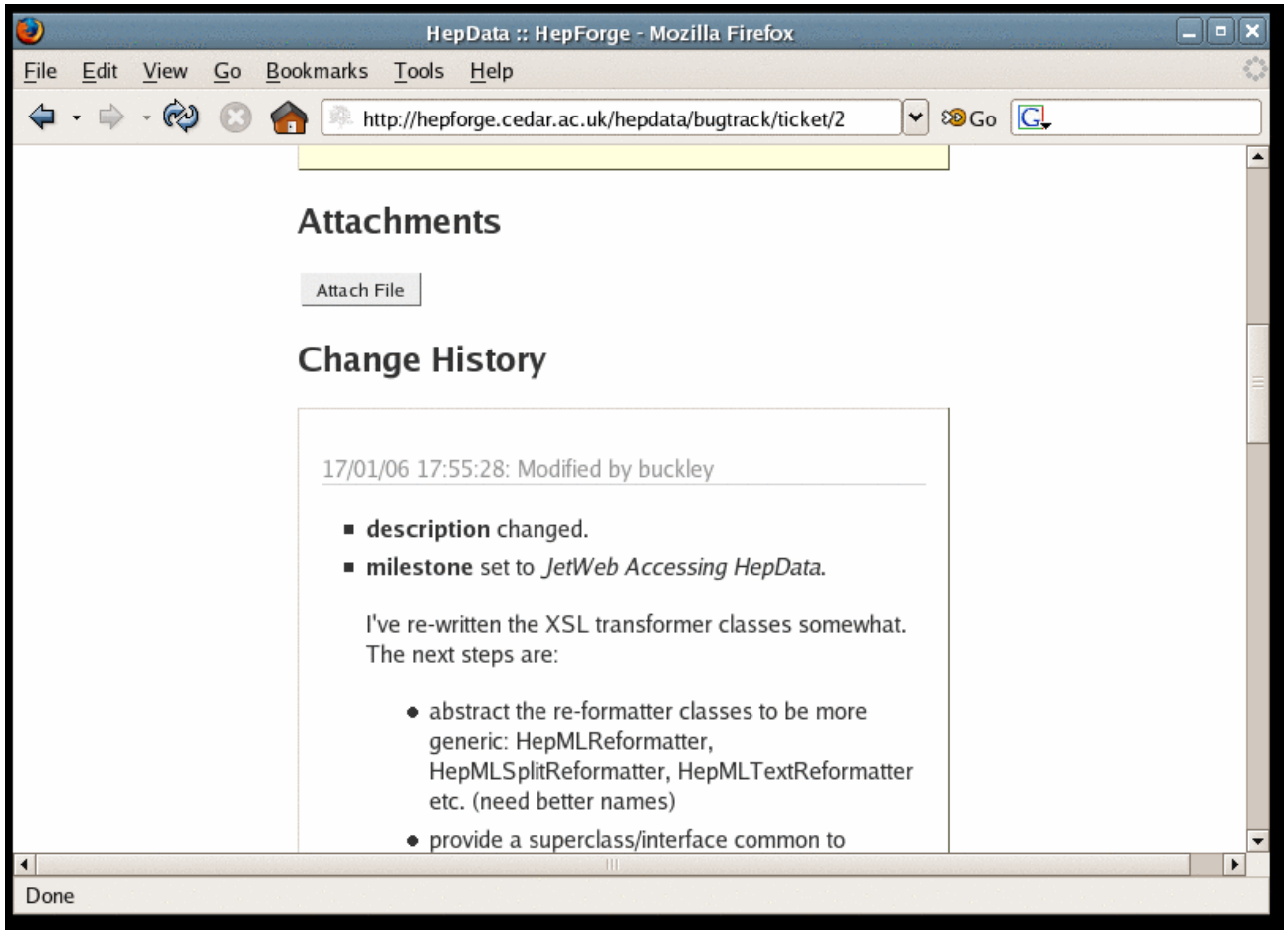


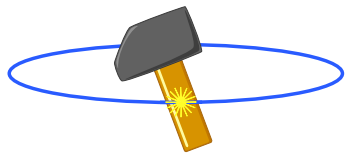


Durham University

HepForge tour

Project bug details (2)

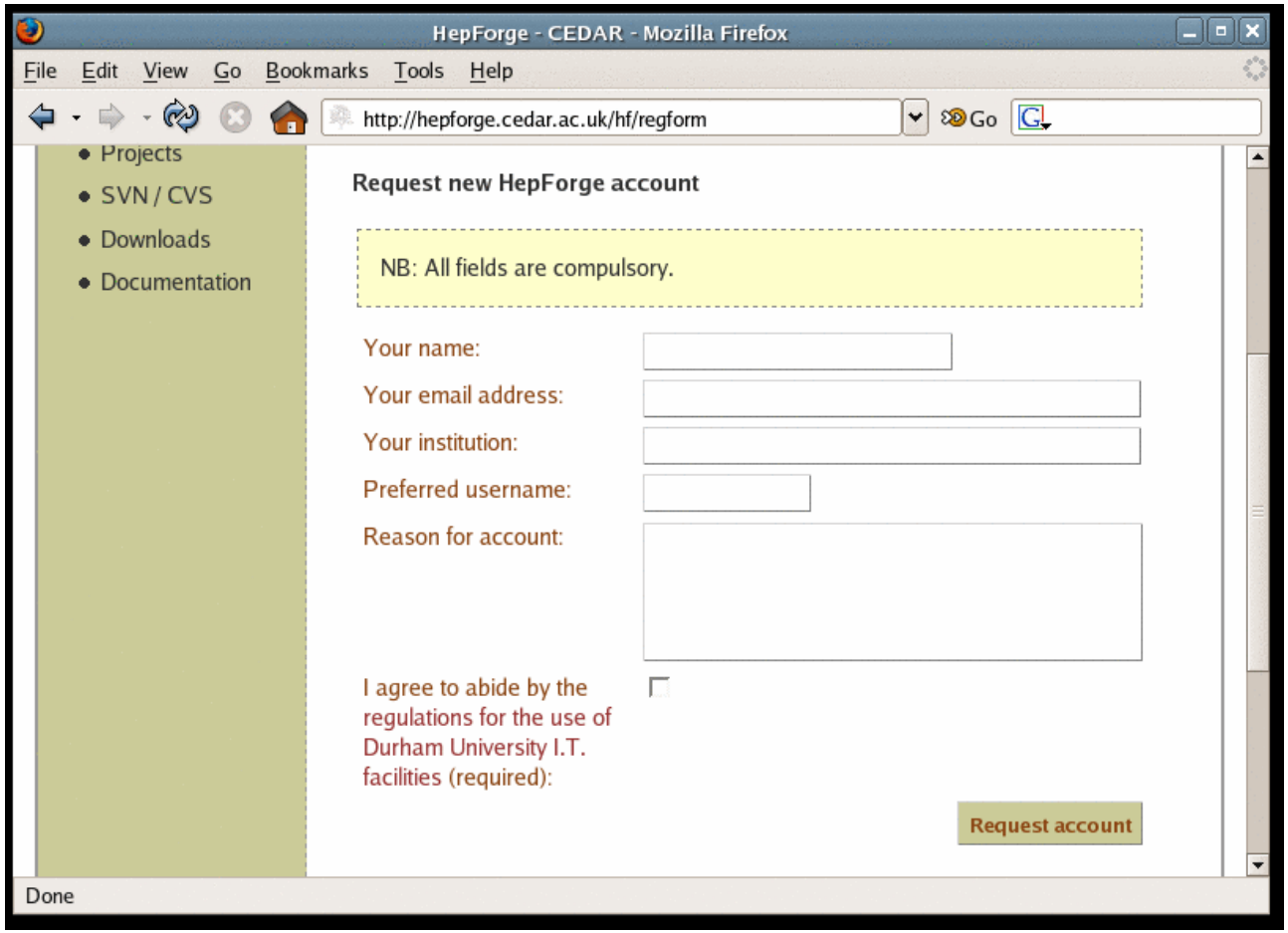


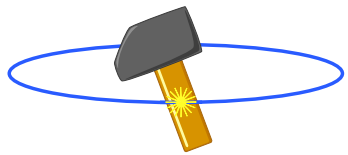


Durham University

HepForge tour

The HepForge registration form! Think about it. . .





Durham University

Features still to come

- Keyword and category project metadata
  (for HepCode & general user convenience)
- Web interface to project metadata
  (keywords, description, “pretty” project name. . .)
- FAQ, news, LaTeX . . . filters
- SSL-encryption of protected project areas
- And many others (see HepForge’s own HF project!)

Durham University

Registration

Please think about using HepForge!

- 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

HepForge summary

- Designed to be easy to use, pick those features useful to you and ignore the rest
- Designed to encourage modular, reuseable, well-documented HEP software
- If you have a project which fits those descriptions, think about using HepForge, it could save you a lot of hassle
- Herwig, Pythia, ThePEG, SoftSusy, LHAPDF, Hoppet, FastNLO, ExHuME. . . already do!
- (Plus all the CEDAR sub-projects, of course)

Durham University

Summary

Summary

- CEDAR is primarily a generator tuning system, combining JetWeb and HepData
- HepData in its new incarnation will use HepML and other standard interfaces
- We need user (expt) input about what you want us to be do for LHC data!

- HepForge is available for HEP software development now!
- Eventually will be used to implement the HepCode repository
- Feedback has 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

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
  - code highlighting, Markdown...
- Download manager (personalisable)

Durham University