Status of MCDB Project

L. Dudko, SINP MSU on behalf of LCG MCDB group

http://mcdb.cern.ch

OUTLINE:

- MCDB outlook
- MCDB usage
- Development plans

LCG MCDB group:

- S. Belov, JINR
- L. Dudko, SINP MSU
- A. Ribon, CERN
- A. Sherstnev, Univ. of Oxford

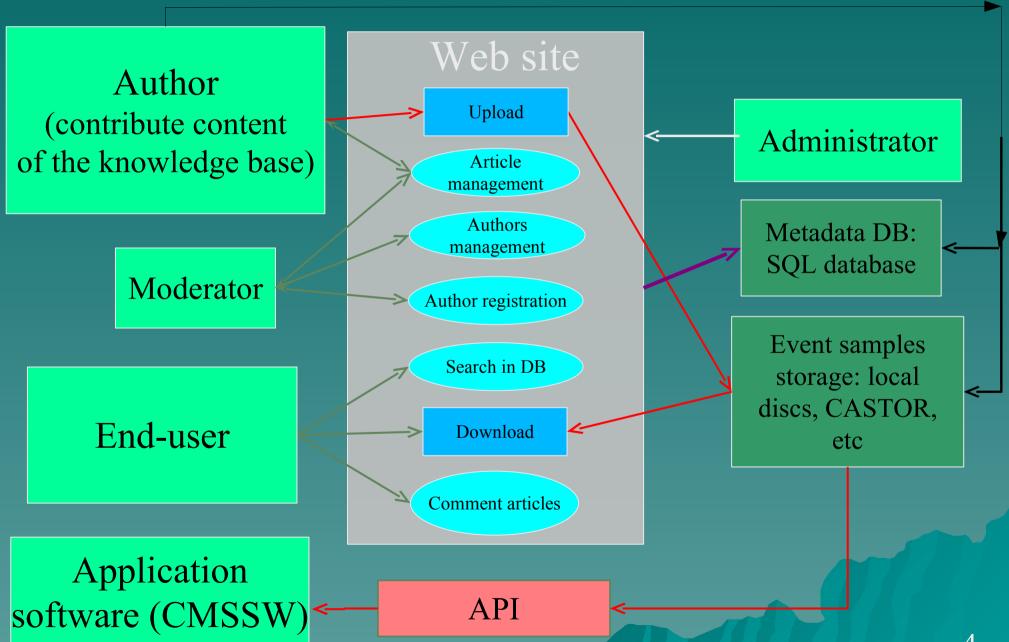
What is MCDB?

- The LCG MCDB knowledge base is developed to store and document sophisticated event samples simulated for the LHC collaborations by experts
- MCDB Provides infrastructure to keep MC samples and sample documentation
- Facilitates communication between MC experts and users in LHC collaborations

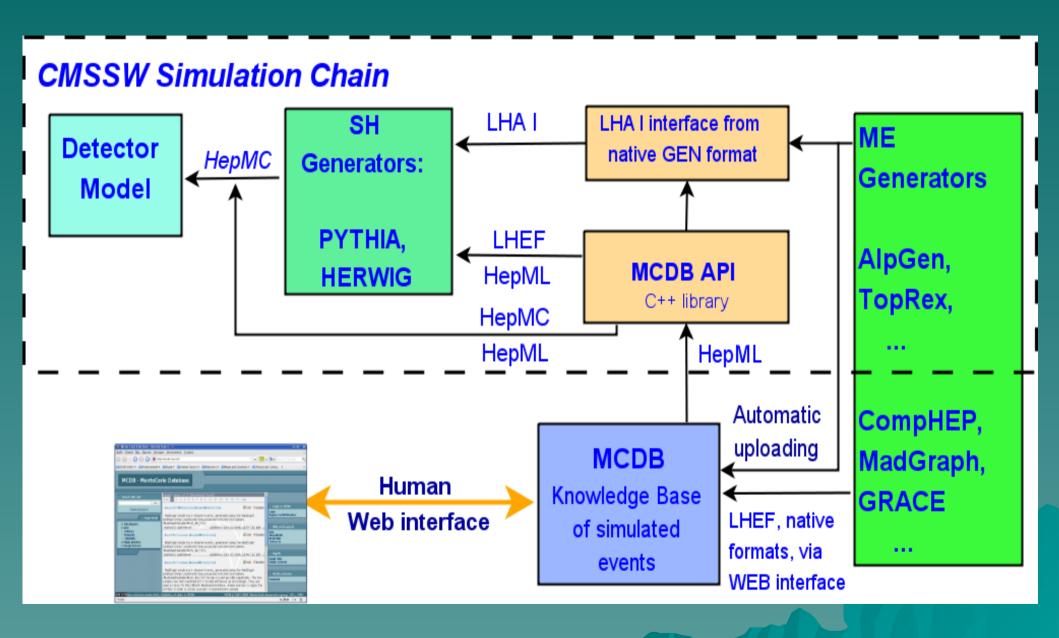
General Structure of MCDB?

- Core MCDB:
 - Web Content Management System
 - Data Base of the information (MySQL)
 - Storage of the samples (CASTOR)
- MCDB Interfaces
 - MCDB API: an interface to get information from MCDB automatically (by external software clients)
 - Uploading Interface: automated uploading of new samples and new information to MCDB
- All Interfaces are based on HepML specifications

LCG MCDB Scheme



MCDB in the Global Simulation Chain



Current Usage of MCDB (I)

- MCDB samples are available for all experiments, but for a while the main consumer is CMS.
 - All externally generated samples for the official CMS production should be accessible in MCDB in LHEF format
 - During the production, CMSSW process these events automatically by means of LHEInterface (MCDB API is the part of the interface – MCDBSource)

https://twiki.cern.ch/twiki/bin/view/CMS/MCDBCMSSWInterface

Current Usage of MCDB (II)

- On the side of Authors/Users statistics:
 - -36 registered authors (not only CMS people, but ATLAS and theory people as well)
 - -234 articles
 - More than 4000 unique visits during the last 3 years

Open Issues for Farther Development

- Content Management System:
 - Improvements and bugfixing for the Import Files Tab (tracking is in Savannah)
 - Reorganization of the templates system in order to be more convenient for the user
- Logging subsystem improvements
- Server hardware
 - HDD partitions reconfiguration

Issues for Development

Interfaces

- Move API to HepML 0.2 specifications
- Add HepML 0.2 parsing for the uploading
- Implementation of the Uploading from CMS Grid sites
- Storage subsystem
 - Currently MCDB uses CASTOR, but users discover some instabilities with CASTOR
 - Improvement of caching and, probably, a Grid storage is needed

General Plans of the Project

- Reorganization of the code tree and migration to SVN
- Prepare the software in the form of RPM packages with necessary installation scripts
- Publish the MCDB software in HepForge as an Open Source project
- Completely distinguish the HepML part as a separate code and publish LibHepML as a separate Open Source project

Conclusion

- MCDB project provides powerful tool to handle events from external ME generators
- MCDB is the official part of the MC simulation chain in CMS
- LCG MCDB team provides support for this project and necessary developments
- All of the developed software will be published as an Open Source and available for independent usage and developments

BACKUP SLIDES

The Major Features of LCG MCDB (I)

- Powerful WEB interface with Content Management System for the authors of event samples and endusers
- Tree graph of physics categories with articles published in MCDB to browse the database content
- Power search engine based on SQL/XML to search the content of the knowledge base
- Flexible and reliable authorization system based on CERN AFS/Kerberos logins or LCG GRID certificates
- SQL structure of event sample documentation
- BackUp of samples and SQL information

The Major Features of LCG MCDB (II)

- CASTOR is the native storage for the event samples
- Direct uploading of multiple files from AFS/CASTOR/ GRID (wild-card characters are possible) to LCG MCDB
- Direct downloading of files from LCG MCDB (CASTOR) with HTTP/RFIO/GridFTP/... (URI)
- Application Programming Interface (API) for the LHC collaborations environment software
- LHEF/HepML unification of event file format and sample description

MCDB Interfaces

- WEB Interface is convenient for interactive work
 - USERs can browse, download, comment
 - AUTHORs can upload and document new samples or modify the old one
- MCDB API is the automatic way to access
 the samples and their documentation from
 external software (e.g. CMSSW)
- Automatic uploading interface is the automatic way to upload and document new samples

Login to the MCDB - MonteCarlo Database Search this site FEEDBACK COMMENTS 📈 Edit 💌 Delete Go Please, provide your feedback comments on the LCG MCDB project, here Login to MCDB Advanced search published: 16th Ney 2005, 13:40 | author(s): Lev Doudko .: Login III Main MENU ☑ Edit 区Delete PROCESS PP->H->ZZ->4MU Top physics Registration The event sample simulates the liggs production with decay to four Exotic production muons (viz Z-bosons). It is created by the CompHEP Monte-Carlo generator. The Register as MCDB author Single top Higgs mass value is 500 G.V. All used physics parameters and applied cuts can Moderators list QCD tt a QCD be found in a prt file are article. B physics published: 19th Sep 2005, 09:42 | author(s): Alexander Sherstnev multijets Help and support Software 📝 Edit 🔼 Delete Help Requests About MCDB Higgs physics Contact us These events were preapred by CompHEP in a special hash-model, where 2 first Gauge bosons quark generations are unified to one of hash-quarks. See details in the aricle itself. Gamma and jets published: 29th Sep 2005, 14:51 | author(s): Alexander Sherstnev O 2gamma and jets W and jets New author ☑ Edit ☑ Delete QCD Z(2TAU)+3J EVENTS WITH ALPGEN2 WW and jets Z and jets registration ZZ and jets Events for the Z+3jets production. Z-boson decays to tau lepton pair. The events were prepared with ALPGEN Monte-Carlo generator. They can be used for the MLM ME-PS matching procedure, since generated with ickkw=1. All generation arameters and cuts applied can be viewed in the qcd_2tau3j_unw.par parameter Articles published: 18th Oct 2005, 12:20 | author(s): Alexander Niliienko . Categories abstracts

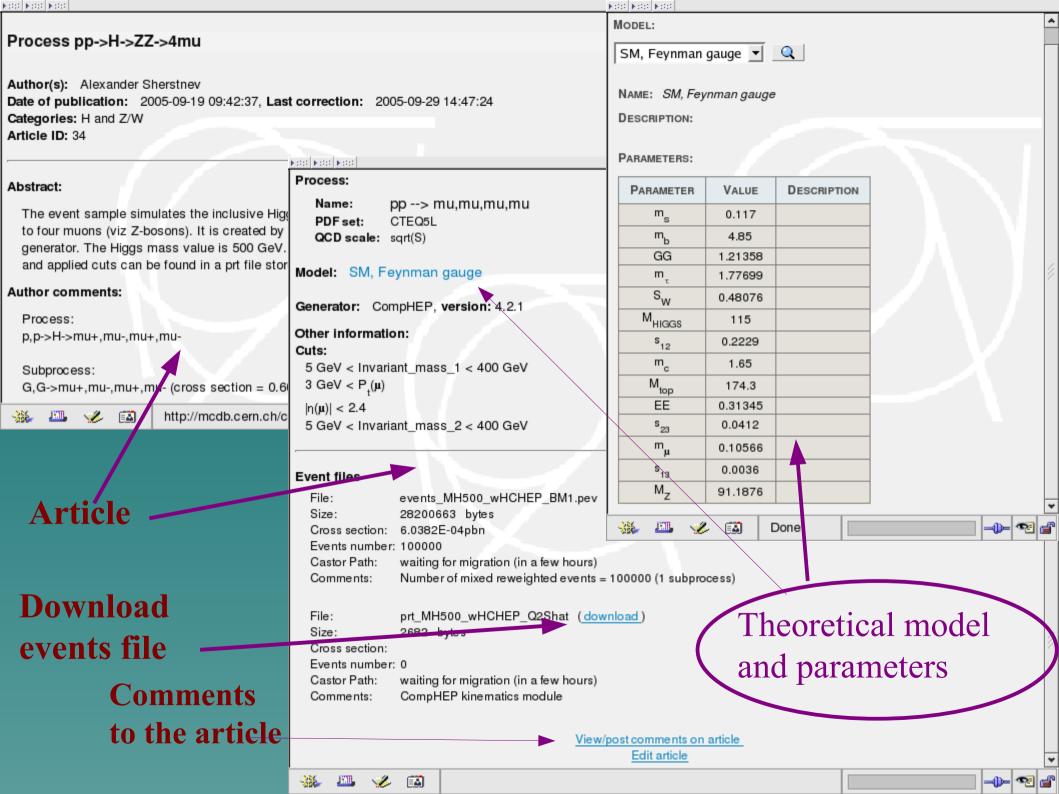
0000228 times visited since October 2005

MCDB © 2005 Monte-Carlo Generators group, LCG, CERN

-0⊨| •••| 💣|



Maria Maria Maria Maria



New Authors HOW-TO

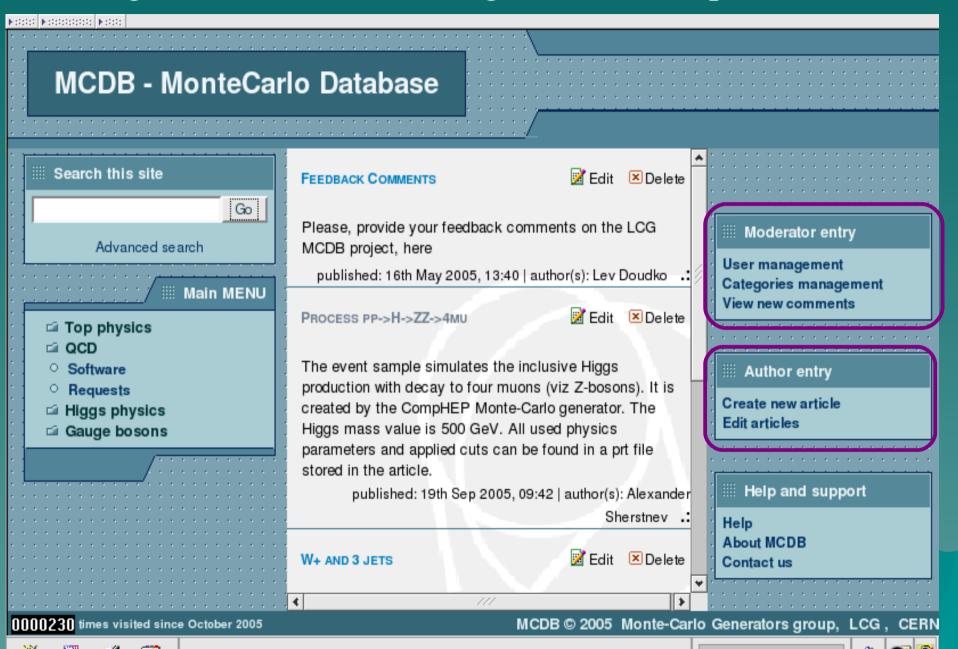
(only need if you plan to upload new samples)

- (1)Register as a new author, wait for the confirmation email
- (2)Login to the LCG MCDB authors area
- (3) Choose "Create New Article" in the authors menu
- (4) Fill the fields in the documentation templates, which will appear (title, generator, theoretical model, cuts, ...)
- (5) Upload your event files in the "Event Files" slice
- (6) Click "Preview/Save" slice and check the box "Publish"

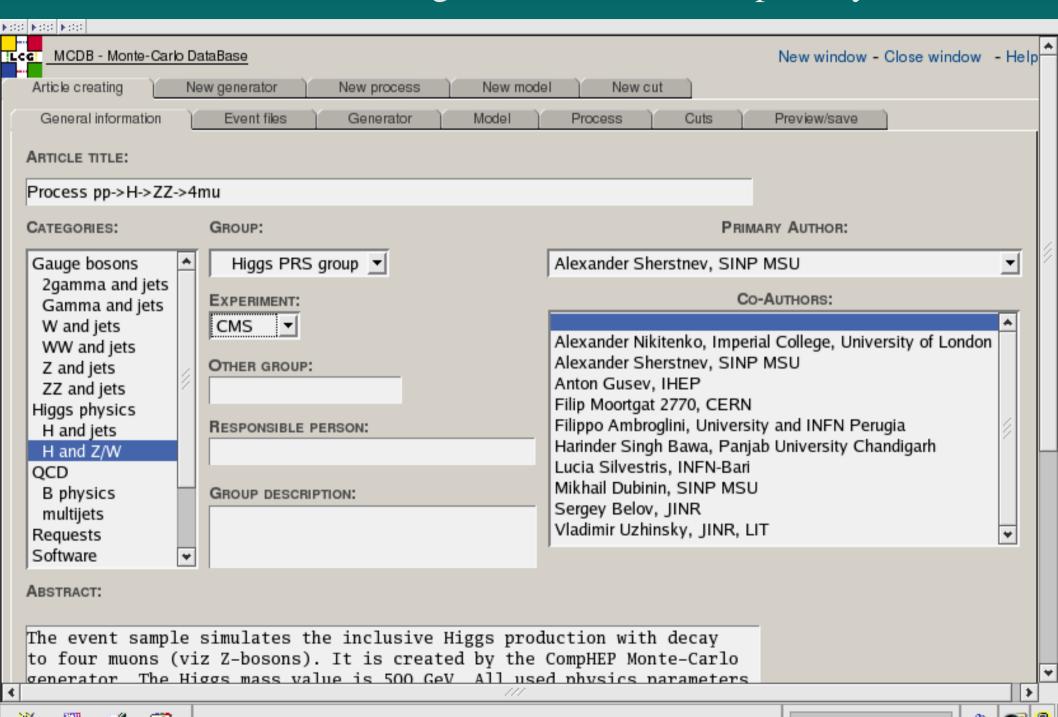
Notes:

- 1. Author needs valid CERN AFS login or LCG digital certificate to be authorized;
- 2. Author can store unfinished articles and resume to correct them in any moment;
- 3. Author can edit articles already published on the Web or do the documents publicly inaccessible for a while.

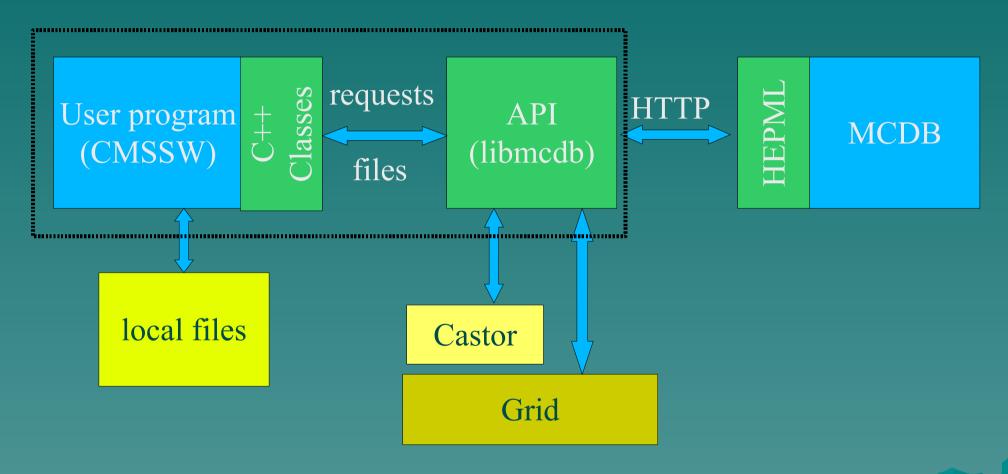
After authorized login to MCDB the additional entries will appear at the right side menu, according to the author permissions



Add/Edit Article link is the gate to the article template system



MCDB API - automatic access to MCDB samples from CMSSW production chain



Interface to download the samples and description from MCDB automatically. Realized in CMSSW LHEInterface and MadGraphInterface, CompHEPInterface (partially)

API STEP I: LibMCDB <=> MCDB

 HTTP request with ArticleID and file name from API to MCDB server

(e.g. http://mcdb.cern.ch/cgi-bin/xmlquery.cgi?article=116)

- HepML description and paths to samples, as an answer from MCDB to API
- API parses the HepML (XML) block and fill
 C++ classes with the description
- API makes local copy of the event sample, download the remote sample via RFIO, GridFTP, SRM, ...

API STEP II: LibMCDB => CMSSW

- API provides to CMSSW:
 - local path to file with input events
 - Sample Description in the form of C++ objects described in mcdb.hpp
- CMSSW process the local event file to the next level of simulation
- CMSSW passes the C++ objects with event description to the output files

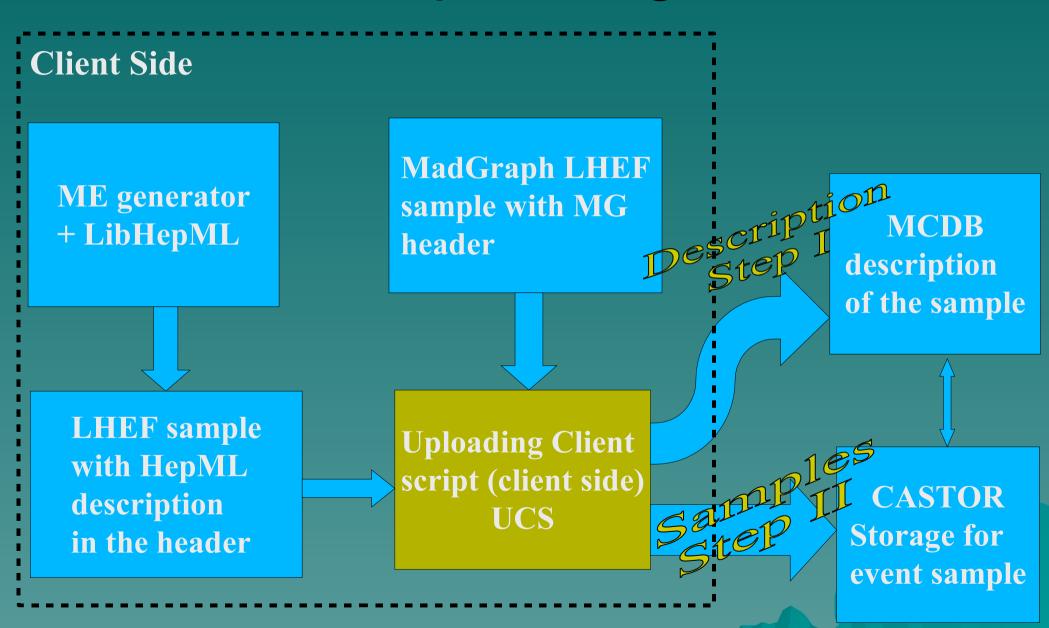
LibMCDB Implementation

(LibMCDB / MCDB API / LHEInterface / MCDBInterface)

- Standalone C++ package for everybody (available on MCDB web site)
- External LCG AA package in /afs/cern.ch/sw/lcg/external/mcdb/
- Integration in CMSSW: LHEInterface (see C. Saout talk), MadGraphInterface, CompHEPInterface

We support all of these implementations for a while.

Automatic Uploading to MCDB



Main tasks for Uploading Client

- Upload LHEF MadGraph/HepML samples and describe it in the new MCDB article, sample description creates automatically from the header of the first sample. The samples upload directly to CASTOR via SRM access.
- Upload additional samples to existent MCDB article and do not change the description. It is possible to upload all kind of files with this option, not only LHEF samples.

Main tasks for Uploading Client. Cont'd.

- Replace the description of the samples in the existent MCDB article and upload new files. Data Set Name or Article ID are identifiers for the MCDB Article
- Describe LHEF sample in new article (automatically create new article), but do not upload any file.

Step I: Uploading Script (UCS) <=> MCDB

- UCS parses LHEF to find the header block
- UCS authorizes in MCDB and pass LHEF header to MCDB via HTTP (POST)
- MCDB parses the LHEF header (HepML or MadGrpah) to find the specific elements of sample description (According to SQL structure)
- MCDB creates new article and specific directory on CASTOR
- MCDB returns to UCS ArticleID and GRID directory to upload the sample

Step II: Uploading Script (UCS) => CASTOR

- UCS Authorizes on CASTOR with Grid certificates: grid-proxy-init
- UCS uploads the sample(s) to the specific incoming CASTOR directory via lcg-cp
- UCS sends request to MCDB to register the uploaded file in the corresponding MCDB article

Possible Uploading Issue:

CASTOR team provides only Storage Resource Manager (SRM) access, by default UCS uses srm://srm-cms.cern.ch members of other VO have to use some other gates (srm-public, srm-atlas, ...), special UCS option is provided: --srmserver

Authorization Mechanisms

- Possible UCS => MCDB authorization
 (the corresponding login or DN should be registered in MCDB first as an MCDB author)
 - CERN AFS login/password for UCS
 - LCG PKCS12 certificate + password
 - LCG usercert.pem/userkey.pem from ~/.globus/ require password
 - grid-proxy-init (default; no password)
- UCS => GRID authorization on CASTOR
 - grid-proxy-init (first check with grid-proxy-info)

Notes and Limitations

- One LHEF header sample description corresponds to one article in MCDB
- UCS takes samples description from the first file and creates MCDB article
- All other files are just copied to CASTOR and attached to the same article, but number of events and sample cross section parsed for each sample.
- Different physics processes should be uploaded by different runs of UCS and described in different MCDB articles
- In the first version of UCS only MadGraph header is possible, HepML is in progress
- MCDB Article identifies by Article ID or Data Set Name (the first one has a priority)

HOW TO Use UCS, Task I

The main task is to describe the set of LHEF (MadGraph or HepML header) files in MCDB as the new article and upload the files to specific CASTOR directory.

```
./upload2mcdb.pl file1 file2 ...
```

-debug

Description for the new MCDB article will be taken from the header of file1 but all other files will be uploaded and attached to the same article.

```
-dsname DataSetName #specify Data Set Name
-header [MG, hepml] #specify type of LHEF header (MG - MadGraph, hepml - HepML header)
-authors AFSlogin1,AFSlogin2,... #set additional authors for the article
-category Category1,Category2,... #set MCDB Category where to attache article (default is CMS08MG)
-not2web #do not post Article to WEB (keep in MCDB), default is post right after it is described
-verbose #be verbose during the run
```

#print additional information during the session

32

HOW TO Use UCS, Task II

<u>Upload more samples to the existing MCDB article</u>

- <u>Requires ArticleID or DataSetName to identify where to attach the samples</u>
- Do not change the description
- Not only LHEF format is possible

upload2mcdb.pl [-artid N] [-dsname DataSetName] --uploadonly file1 file2 ...

Possible options:

- -verbose #be verbose during the run
- -debug #print additional information during the session

HOW TO Use UCS, Task III

Replace the description in the existing article and upload new files

upload2mcdb.pl [-artid N] [-dsname DataSetName] --replace file1 file2 ...

```
Possible options:
-dsname DataSetName
                            #specify Data Set Name (analog ArticleID)
                            #specify type of LHEF header (MG - MadGraph,
-header [MG or hepml]
                               hepml - HepML header)
                                    #set additional authors for the article
-authors AFSlogin1, AFSlogin2,...
                                   #set MCDB Category where to attache
-category Category 1, Category 2,...
article (default is CMS08MG)
                            #do not post Article to WEB (keep in MCDB),
-not2web
                              default is post right after it is described
                     #be verbose during the run
-verbose
                     #print additional information during the session
-debug
                                                                         34
```

HOW TO Use UCS, Task IV

<u>Describe sample in new article but do not upload any file to CASTOR.</u>

<u>The description is taken from the header of the file</u>

upload2mcdb.pl --descriptiononly file1

Possible additional options:

-dsname DataSetName #specify Data Set Name -a [login, pkcs12, cert, globus] #type of authorization in MCDB, default is

globus

-header [MG, hepml] #specify type of LHEF header (MG -

MadGraph, hepml - HepML header)

-authors AFSlogin1,AFSlogin2,... #set additional authors for the article

#set MCDB Category where to attache

article (default is CMS08MG)

-not2web #do not post Article to WEB (keep in MCDB), default is

post right after it is described

-verbose #be verbose during the run

-category Category 1, Category 2,...

-debug #print additional information during the session

35

HOW TO Get Help on Uploading Client Script

Get the Uploading Client Script (download section of MCDB) wget http://mcdb.cern.ch/distribution/upload2mcdb.pl

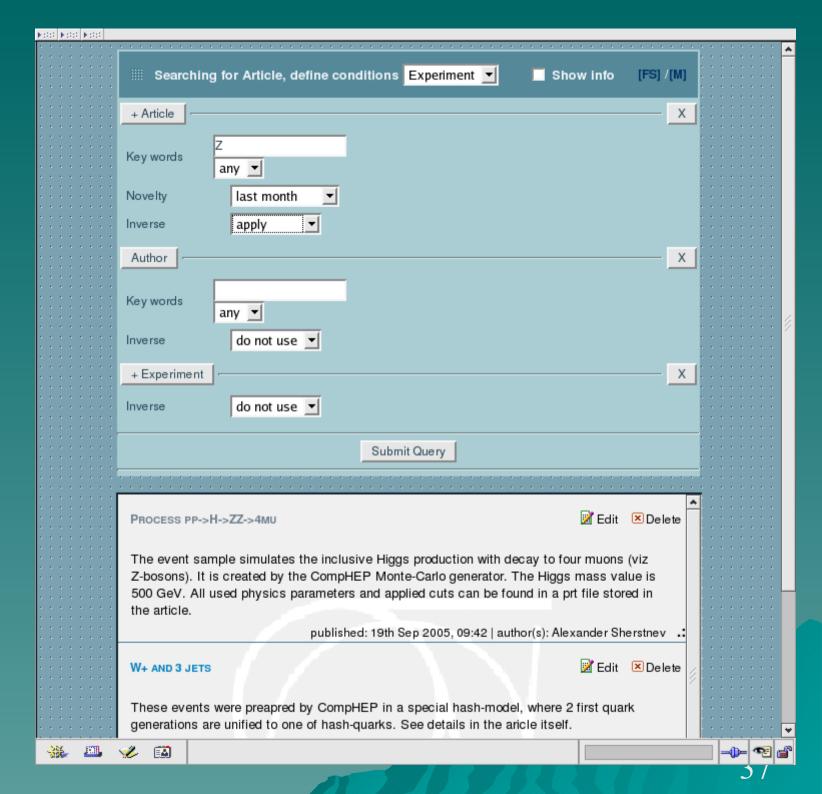
Short help. Get the list of available options:

./upload2mcdb.pl -h

Long help. Get the detailed description with examples:

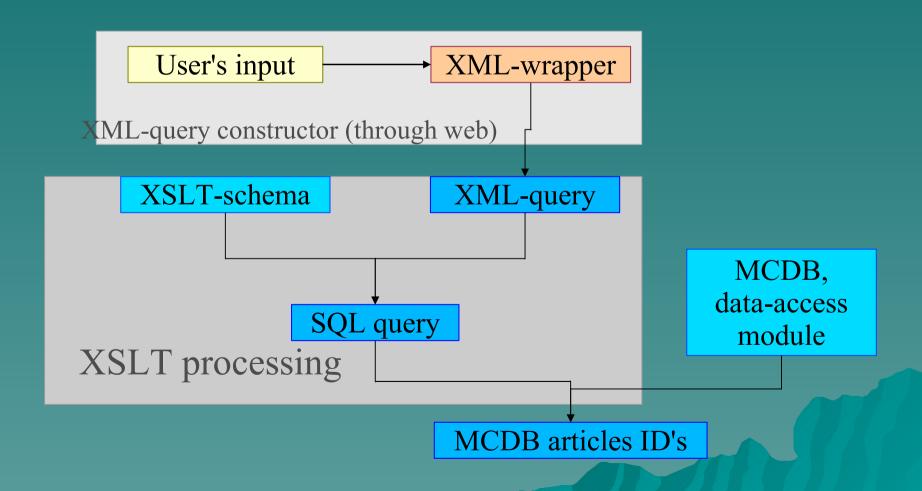
./upload2mcdb.pl -help

Advanced Search Query



MCDB Search Engine

- dynamic query construction wizard (JavaScript/XML/SQL)
- Search by many possible criteria with complicated relations between DB-objects



Users Comments Interface

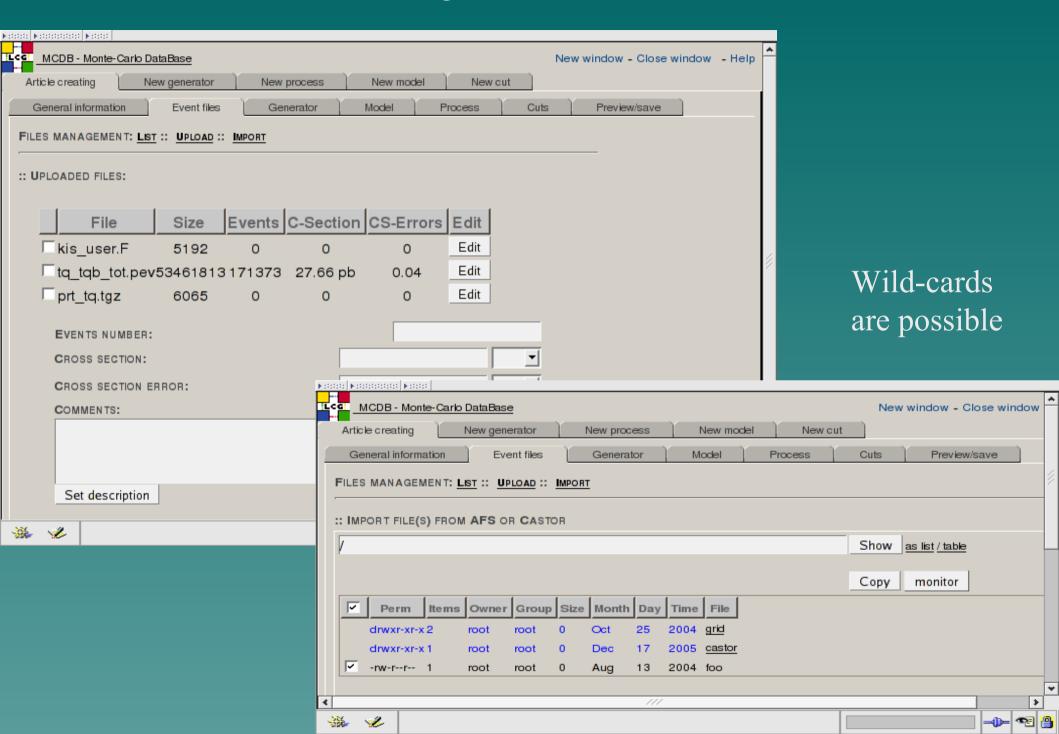
Fidd Fidd Fidd				
PROCESS PP->H->ZZ->4MU	☑ Edit 🗵 Delete			
The event sample simulates the inclusive Higgs production with decay to four muons (viz Z-bosons). It is created by the CompHEP Monte-Carlo generator. The Higgs mass value is 500 GeV. All used physics parameters and applied cuts can be found in a prt file stored in the article.				
published: 2005-09-19 09:42	:37 author(s): Alexander Sherstnev 😀			
Comments: [Moderate]				
YOUR NAME:				
E-MAIL:				
COMMENTS:	_			
Send				
🔆 🕮 🌿 🔝 Done	-0- 12- 12- 12- 12- 12- 12- 12- 12- 12- 12			



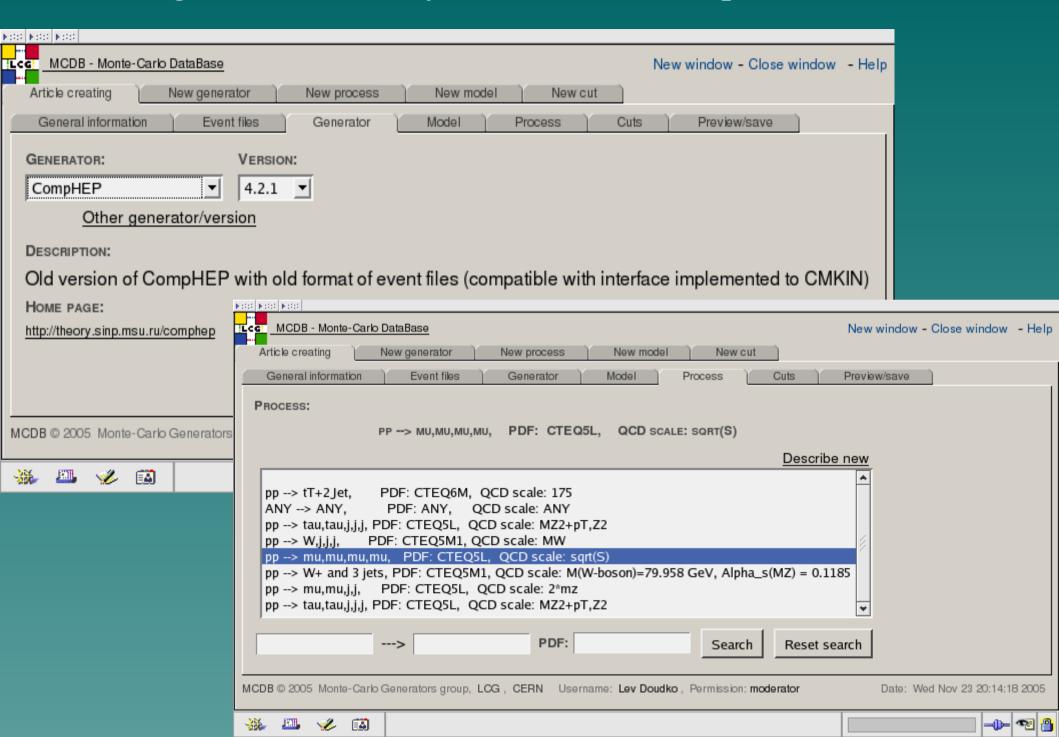
- Form to send a request for the authorization as new LCG MCDB author.
- Necessary only you want to uplo new MC samples

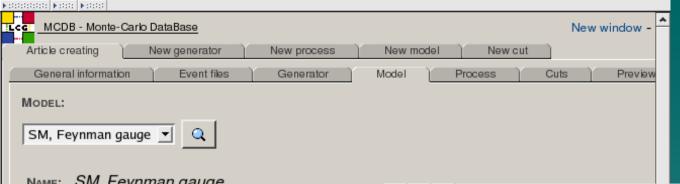
FREE FREE FREE		
[™] Please provid	egistration le following information to register e required (and at least one of *)	*
First name:		
Last name:		
CERN AFS login: *		
DN from LCG certification	ate: *	1/2
Experiment:		
Group:		
Organization:		
E-mail:		
	examine you personal LCG certificate or load it to your browser (instructions here) Register Cancel	
₩ ₽ ₩	Done — • • •	<u> </u>

Event files slice to manage event files attached to the article



MC generator and Physics Process description slices





Physics model parameters and applied cuts slices

Name: SM, Feynman gauge

DESCRIPTION:

PARAMETERS:

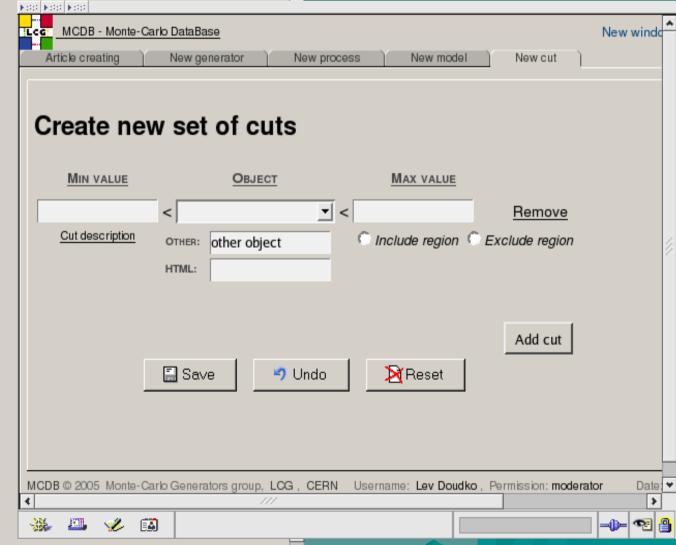
Управление гр

PARAMETER	VALUE	DESCRIPTION
m _s	0.117	
m _b	4.85	
GG	1.21358	
m _τ	1.77699	
S _W	0.48076	
M _{HIGGS}	115	
S ₁₂	0.2229	
m _c	1.65	
M _{top}	174.3	
EE	0.31345	
S	0.0412	
m _µ	0.10566	
S ₁₃	0.0036	
M _Z	91.1876	

MCDB @ 2005 Monte-Carlo Generators group, LCG, CERN Username: Lev Doudko, Permission: moderator

Mozilla-bin [4] ▲

Evolution - Почта mcdb Ihcb.odp -



Date: Wed♥

k us

PARAMETERS OF EVENT SAMPLE DESCRIPTION MCDB XML Scheme inside HepML specifications

- General information
 - Title
 - Abstract
 - Authors
 - Experiment and/or Group
- Physics process
 - Initial state
 - Final state
 - QCD scale
 - Process PDF
- Event files
 - Physics process/subprocesses
 - File name
 - Events number
 - cross section and uncertainty

- Used generator
 - Name and version
 - Description
 - Home page address
- Theoretical model
 - Name
 - Description
 - Set of parameters and their values with author's descriptions
- Applied cuts

MCDB API C++ Classes

http://mcdb.cern.ch/doc/API/public/mcdb.hpp

```
class Generator{
                           public:
namespace mcdb
                             Generator();
                             ~Generator();
                             string& name();
class MCDB;
                             string& name(const string&);
                             string& version();
class Article;
                             string& version(const string&);
                             string& homepage();
class File;
                             string& homepage(const string&);
class Author;
                            class Process{
                            public:
class Cut;
                              Process();
class Generator;
                              ~Process();
                              string& initialState();
class Model;
                              string& initialState(const string&);
                              string& finalState();
class Process;
                              string& finalState(const string&);
                              string& factScale();
class Subprocess;
                              string& factScale(const string&);
                              string& renormScale();
                              string& renormScale(const string&); };
```

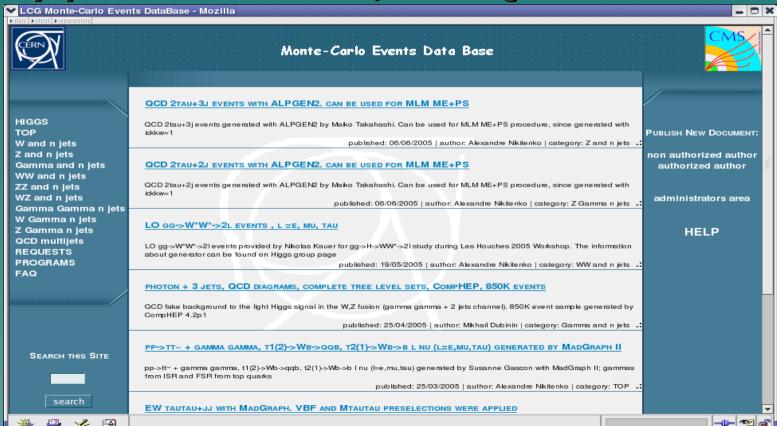
string& pdf();

string& pdf(const string&);

```
class Model{
 public:
    Model();
    ~Model();
    class ModelParameter;
    string& name();
    string& name(const string&);
    string& description();
    string& description(const string&);
    vector<ModelParameter>& parameters();
    vector<ModelParameter>&
parameters(const vector<ModelParameter>&)
    class ModelParameter
     public:
        ModelParameter();
        ~ModelParameter();
        string& name();
        string& name(const string&);
        string& value();
        string& value(const string&);
                                   45
```

History: CMS MCDB

- http://cmsdoc.cern.ch/cms/generators/mcdb/
- Operated in CMS during the OSCAR/ORCA era, widely used by the Higgs group
- Only parton level files; AFS storage;
 Only phonetic search; No SQL



Documentation

- Main Web Page http://mcdb.cern.ch
 - Description of the project
 - Users and Authors HOW-TOs
 - Developers documentation
- Wiki

https://twiki.cern.ch/twiki/bin/view/CMS/MCDBCMSSWInterface

https://twiki.cern.ch/twiki/bin/view/LCG/LCGMCDB

- [hep-ph/0404241] LCG MCDB proposal
- [hep-ph/0604120] LCG MCDB report (p.200-204)
- [hep-ph/0703287] LCG MCDB description
- Core software supported by LCG Software Project Infrastructure (MySQL; CASTOR; CGI; Perl; Apache)
- Mailing lists USERS: lcg-mcdb-users@cern.ch Developers: project-lcg-mcdb@cern.ch

Summary

- Three MCDB interfaces are ready:
 - Interactive WEB interface (everything is possible)
 - MCDB API is the automatic way to process the event samples from MCDB during the production (not only LHEF samples)
 - Uploading Interface is the automatic way to upload and describe LHEF sample(s) in MCDB (Not only LHEF samples can be uploaded but not described automatically. First version of UCS works with MG header only)
- LibHepML is the proposed standard way to describe the sample automatically in ME generator (development version is available in MCDB)

MCDB Software

- Stable versions of MCDB server and MCDB API are available in the download section: http://mcdb.cern.ch/distribution
- Development versions are in MCDB CVS: simu.cvs.cern.ch:/cvs/simu/GENSER/MCDB http://simu.cvs.cern.ch/cgi-bin/simu.cgi/simu/GENSER/MCDB/
- Integration of MCDB API in CMSSW CVS is in CMSSW/src/GeneratorInterface/MCDBInterface
- All of the MCDB software will be published as an OpenSource in HepFourge

Where to Find MCDB Uploading Interface Scripts

Client Part of Uploading Interface: upload2mcdb.pl

CVS: http://simu.cvs.cern.ch/cgi-bin/simu.cgi/simu/GENSER/MCDB/distribution/

Stable Version: http://mcdb.cern.ch/distribution/

Server Part of Uploading Interface: upload_server.cgi

CVS: http://simu.cvs.cern.ch/cgi-

bin/simu.cgi/simu/GENSER/MCDB/cgi-bin/authors/

HepML in LHEF

- J. Alwall et al., A standard format for Les Houches Event Files (2006) [hep-ph/0609017] and [hep-ph/0703287]
- Event Sample Structure:

```
<LesHouchesEvents version="1.0">
  <header>
    <hepml>
        <!-- HepML sample description here -->
    </hepml>
  </header>
  <init> ... </init>
  <event> ... </event>
  <event> ... </event>
  . . . . . . . .
</LesHouchesEvents>
```