

# Bookkeeping

Alessandra Forti

26<sup>th</sup> June 2003

# Outline

- Data set tables
- Data set code
- Goals for the beginning of August
- Task management
- PR/SP/luminosity/RQM interaction
- Conclusions

# Data Set tables

- Data set tables will contain the information relevant to the users (dataset names, collections, luminosity....). They are therefore the most important part of the bookkeeping.
- There is now a first schema. It is not yet in cvs but it will be in the next two days after having applied some modification required.
- There is already a script to fill it with SP4 collections. 1M collections have been inserted as first test.

# Data Set tables

- We need also one to load PR collections and July test collections.
- This will allow us both to test the schema and to understand more in detail what PR and SP bookkeeping modules need to publish in the tables and how.
- The insertion of the new test collections is important as well, if not more, because it also tests a reduced (adequate for the July test) support for multiple files.

# Data set code

- Data set code, I remind you, has different levels: higher level interfaces in the shape of command line scripts and lower level implementation with classes, methods and attributes:
  - there are already 2 classes in cvs corresponding to two of the tables in the table schema but they are more of a place holder and an example on how to use the framework than real objects.
  - We didn't have time to discuss about this at the workshop but I will send out a first proposal for a more developed model of the classes on Monday.

# Goals for the beginning of August:

- Insertion point of view:
  - replication of the GoodRuns package functionality in the database.
    - Insertion of collections
    - Creation of datasets with appropriate names
  - Adding luminosity information
- User Interface:
  - Data set retrieval
  - Data set luminosity calculation
  - Data set splitting into subsets according to the number of events for the tcl files.

# Task Management

- Data set is the first priority but we are working also on task management.
- There is a proposal still under private discussion for a table schema connected to the dataset table.
  - The use case on which this proposal was formed is the skimming and needs to be revised in the light of users production.
- There are not yet defined goals for this part (e.g. job submission but not job monitoring). It should be defined next week though.

# PR/SP/luminosity/RQM/Data Dist

- The interaction between the Dataset and these modules has been better defined.
  - PR and SP are not anymore so untouchable even if they will remain totally distinct from the dataset.
  - They will use their own data distribution methods.
  - RQM and luminosity modules from being independent will be integrated in PR as sub-modules.
    - If needed temporary methods to insert their info will be developed
  - We defined when PR and SP should “publish” and what we mean with it. See Tim talk.

# Conclusions

- There has been a lot of progress in the last 3 weeks.
- There are two new people: 1 has already started and will work on task management, the other will start in July.
- More important the level of communication between the members of the group has increased (I hope it will continue on this level)
- The workshop also helped to establish interaction between the dataset module and the production modules that were still vague.
- Down side we are late on the schedule of at least one month.

This document was created with Win2PDF available at <http://www.daneprairie.com>.  
The unregistered version of Win2PDF is for evaluation or non-commercial use only.