

DØ data processing within EDG/LCG

Torsten Harenberg¹, Kors Bos², Rob Byrom³, Steve Fisher³, David Groep², Willem van Leeuwen², Peter Mättig¹, Jeff Templon²

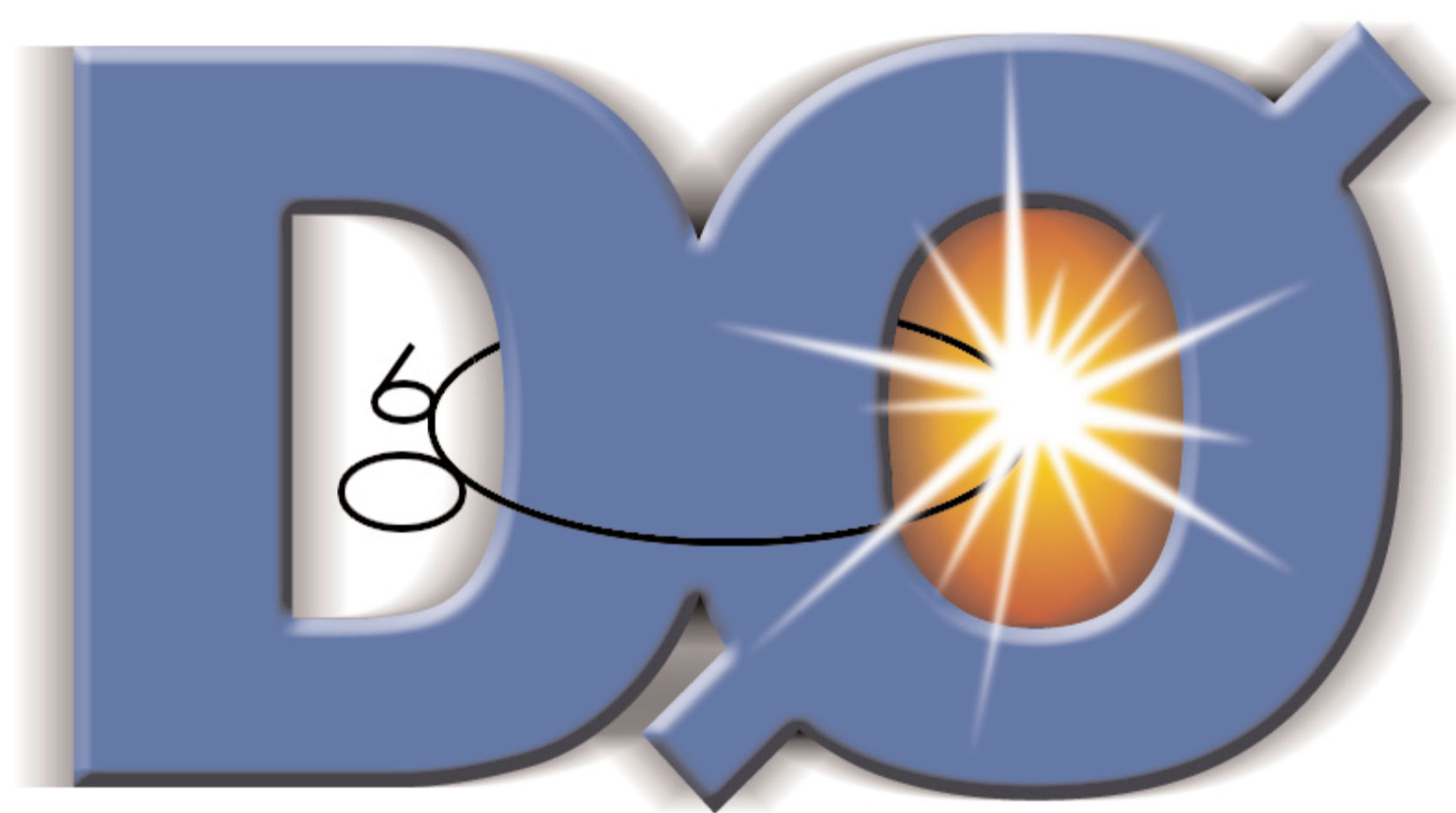
¹ Bergische Universität Wuppertal, Germany

² NIKHEF Amsterdam, The Netherlands

³ Rutherford Appleton Laboratory, U.K.

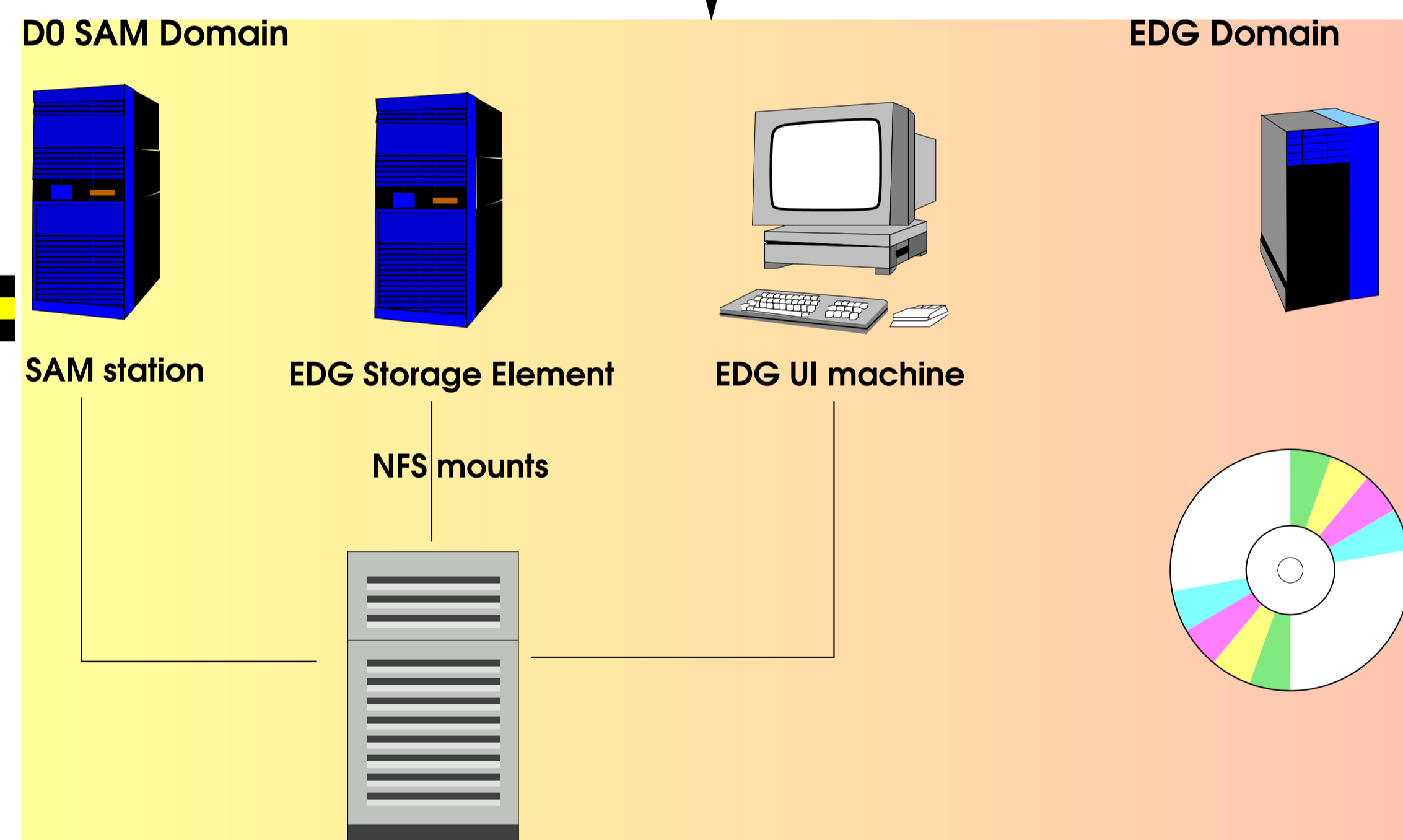
For the first time, data from a running experiment has been successfully reprocessed in a full-scale generic grid infrastructure.

In September 2003, the DØ experiment at Tevatron has launched a reprocessing effort. In total 519,212,822 of the experiment's events have been reprocessed to use the new perceptions of the detector's behaviour. Out of these events 97,619,114 have been reprocessed at remote sites. For the first time, the European DataGRID has been used to re-process a part of these events as an evaluation of the EDG application testbed



At Fermilab
Requests are centrally generated and passed to the EDG cluster team.

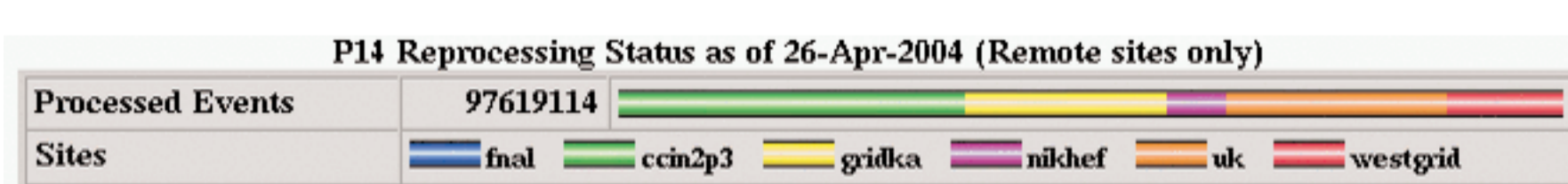
SAM/EDG gateway
m input Data into EDG DMS
m adapt DØ software
m monitor the job
m put results back into SAM



As DØ has its own data management system called "SAM", some sort of channel between SAM and the EDG data management system is required. The approach used is shown to the left, where a certain storage area, physically present on a back-end server machine, is visible both from a SAM-enabled machine ("SAM station") and from EDG machines at the same site. This has been achieved at NIKHEF.

Conclusions / Outlook

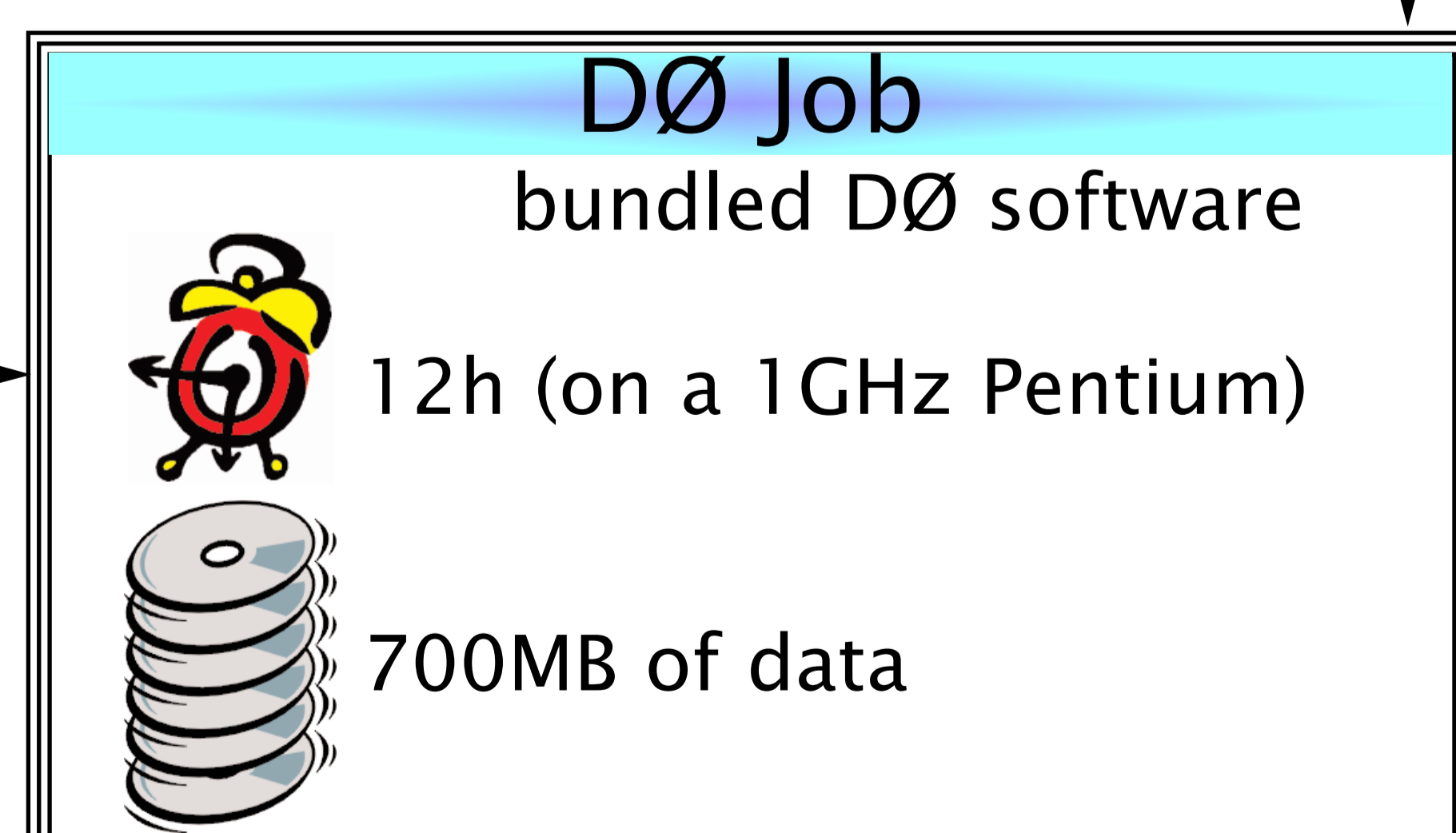
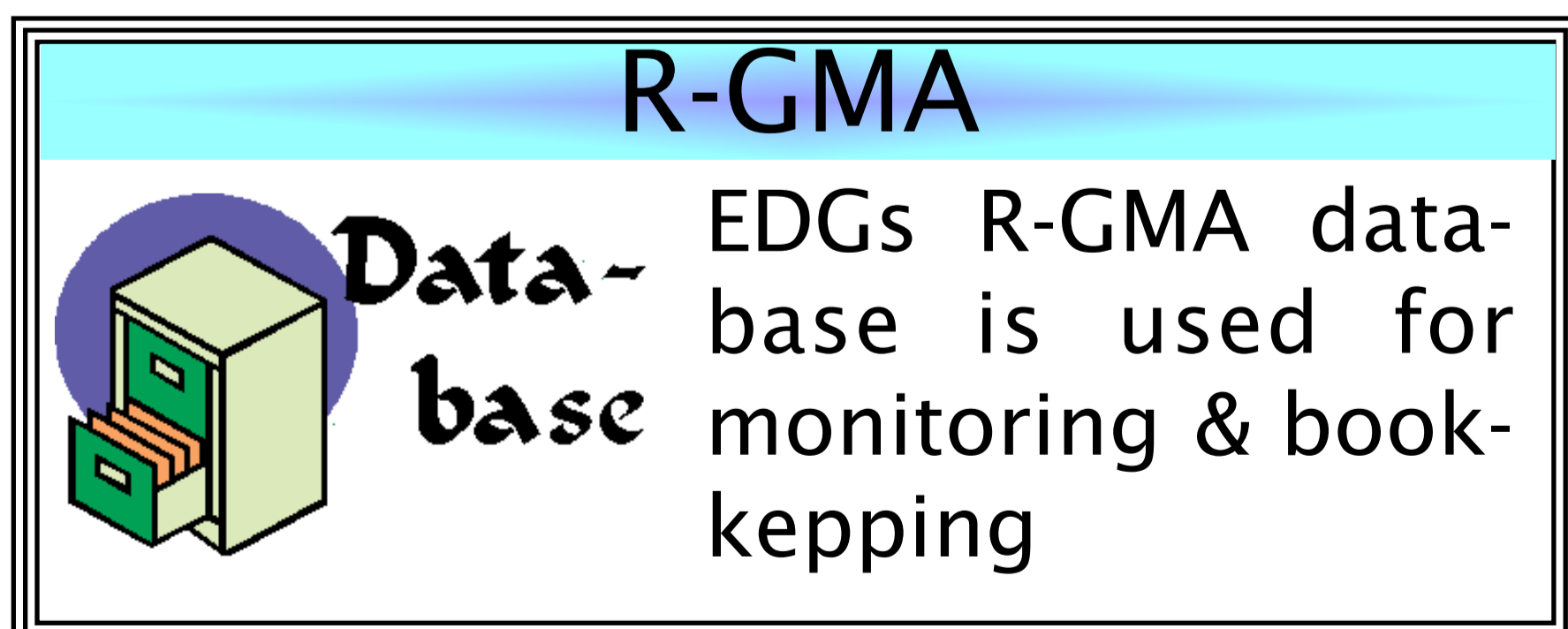
EDG has proved its usability for the first time on real data and we valuable experience could be collected. The preprocessed events are included in NIKHEF's contribution to the effort:



Several bugs and instabilities were found and reported, including R-GMA and WP5 SE. We continued to work on LCG resources with Monto-Carlo production.

We used EDG's own R-GMA data-base for Monitoring & Bookkeeping and constructed four tables:

- "submission table" - records the submission of jobs to the Resource Broker
- "job start table" - holds the time the job started on a Worker Node together with process ID and many more
- "job end table" - information is published immediately before the job stops
- "command table" - a command list table for debugging purposes



The DØ software has been adapted to run in the EDG framework. Only a few changes has to be made. Missing libraries were included and some extra packages were shipped (pyxml and Python 2.1).

However, arrounding wrapper scripts were written to handle the in- and output and put/get it from/to EDG's Data Management System (DMS).