Data Access and Management Software

David Malon
Argonne National Laboratory
malon@anl.gov

ATLAS Data Management

- Scope
- Organization
  - U.S. roles
- Status
  - U.S. contributions to date
- Plans
  - Proposed U.S. tasks
- Summary

DOE/NSF Quarterly review
January 1999

Scope of Database Domain

- Physics content databases
- Supporting infrastructure
- Joint responsibilities with other domains
- What is NOT included

Concrete Databases

- Event store (raw, reconstructed, simulated data supporting online, offline, and testbeam computing)
- Detector description (geometry)
- Conditions databases (calibration, alignment, run conditions)
- Statistics and analysis stores

Database Infrastructure

- Control/database interfaces, and generic database components needed to support those interfaces
- persistent side of transient/persistent interfaces
- physical data clustering and storage optimization
- data organization and indexing
- tertiary storage access & management

Database Infrastructure

- Infrastructure for distributed database development
- wide-area distributed data access
- database coding rules and endorsed practices
- schema evolution
- database administration
- data-store-specific infrastructure (e.g., Objectivity-specific components)
- *...more...*

Joint Responsibilities

- Event model
- Control/database interface
- Detector description model
- Interfaces to fabrication (production) databases

Outside the Scope

- Fabrication (production) databases
- Detector-on-the-floor database
- Service lines database

***Production database group** chaired by D. Ferrere & A. Petrilli has responsibility for these.*

***Interfaces** between these and detector description are joint responsibilities.*

Organization

- Prior to 1999 review, RD Schaffer headed the database domain and was very nearly its sole developer
- Like **Reconstruction** and **Simulation**, the **Database** domain was structured by ATLAS management after the 1999 software review to have a coordinator, and task leaders from each subsystem

Organization

- Global database coordination is now the JOINT responsibility of
  - David Malon (Argonne) 
  - RD Schaffer
- Subsystem database task leaders
  - Inner Detector: Stan Bentvelsen
  - Liquid Argon: Stefan Simion
  - Muon: Steven Goldfarb (Michigan) 
  - Tile: Tom LeCompte (Argonne) 
  - Trigger/DAQ: H.P. Beck

Subsystem database roles

- Content providers
  - e.g., subsystem geometry data
- Advice and consent with respect to core database strategies
- Testbed for core technologies
  - e.g., multiple approaches to transient/persistent mapping in tile testbeam

Subsystem database roles

- Core component prototyping
  - alignment databases will likely be prototyped in muon testbeam
- Sources of expertise within subsystems on data access
  - e.g., how to access TDR data from PASO

DOE/NSF Quarterly review
January 1999

Database efforts to date

- Prior to late summer 1999, database efforts were hampered by
  - lack of available personpower (approximately 1 person working)
  - need for precursory work, like defining an ATLAS event model or a generic detector description prior to building an event database or a detector description store
    - these tasks fell to the same person

Database efforts to date

- Primary products until summer 1999
  - visitor-pattern-based event data access
  - digitization data for SOME subsystems via this model from the combined performance TDR
  - hierarchical identifier design and initial implementation, and design ideas for a generic detector description model

Database efforts to date

- Above software was used for a 1-terabyte test of Objectivity, replicating data from these partial events 10 times
- Appointment of subsystem database coordinators made it possible to delegate work
  - people have been identified to complete data model for all subsystems from TDR data
  - detector geometry definition using XML is now well underway

U.S. contributions to date

- In summer 1999, the Argonne-led tile calorimeter testbeam pilot project provided first production use of Objectivity in ATLAS
  - innovative detector-centric architecture
  - support for multiple transient/persistent mapping strategies simultaneously
  - hot-swappable calibration strategies using strategy patterns

U.S. contributions to date

- Simona Rolli (Tufts) has been exploring and benchmarking Objectivity-specific approaches to scalability (e.g., use of segmented Varrays) in the context of an ATLAS event model
- BNL has been involved as an early user of event data access via PASO
- U.S. subsystem database coordinators (Steven Goldfarb, Tom LeCompte) are providing geometry data and event decoding effort

Database policy

- Official position is
  - use LHC-wide solutions wherever possible (e.g., RD45's HEPODBMS)
  - Objectivity is the ATLAS baseline solution
  - physics codes should be independent of database supplier insofar as this is reasonably achievable

Near-term Priorities

- Overall database domain design, coordination, oversight, and planning
  - two(!) CERN reviews of ATLAS computing this year
- Infrastructure for (distributed) development
- Database components to support database/control framework interfaces
- Event model and data source for 2000 prototype framework

Near-term Priorities

- Access to TDR data for reconstruction code development
- Detector description
- Initiation of effort to provide database support for simulation
- Production experience with Objectivity

DOE/NSF Quarterly review
January 1999

Infrastructure for (distributed) development

- A *sine qua non*
- Work is underway (joint U.S./CERN)
- Requires changes to SRT (ATLAS release management tool)
- Plan is to learn from BaBar

Database/Control

- The U.S. will likely have significant overall architectural responsibilities, and primary responsibility for the control framework. We must provide the data access components needed to allow this effort to succeed.

Event data source for 2000 prototype framework

- Propose to use calorimeter data from the combined performance TDR
- Requires significant event model effort
  - important for U.S.-led architecture effort
  - good match to BNL's work on LAr OO reconstruction
  - generalizable to and prototype for overall ATLAS event definition, which will be a joint simulation/reconstruction/event filter/core software task

Access to TDR data

- Plan is to make all TDR hits and digis available through interim control framework (PASO) early this year
- Needed for reconstruction code development
  - Frank Merritt (Chicago) is tile reconstruction coordinator 
  - Srinivasan Rajagopalan (BNL) leads the *ab initio* LAr reconstruction effort 

Access to TDR data

- Project has been underway for a while
  - coordinated by RD Schaffer, who also did most of the inner detector work
  - used in 1-terabyte milestone
  - BNL LAr was an early client via PASO
- Subsystem database coordinators will oversee completion

Detector Description

- Precursor to geometry database efforts
- Current work is directed at providing an XML description of detector geometry
  - good enough to describe shapes
  - unclear yet how this will work with respect to logical organization of the detector and support for multiple detector views and detail filtering
  - needs a design review cognizant of emerging XML technologies (XML Schema)

Detector Description

- Geometry data provided via subsystem database coordinators
- XML data will be used to feed a GENERIC detector description (independent of application views) that needs to be defined and implemented
- later: detector description DATABASE

Database support for simulation

- ATLAS Geant4 simulation efforts are just beginning
  - time is right to begin the associated database effort
- U.S. has overall responsibility for
  - LAr simulation (Misha Leltchouk) 
  - Inner detector simulation (Fred Luehring) 
  - Monte Carlo generators (Ian Hinchliffe) 

Database support for simulation

- Propose to begin with Monte Carlo input events
  - good match to U.S. event generator responsibilities, and to Isagen work done by Boston/Harvard
  - one piece of eventual multicomponent ATLAS event
  - gentle introduction to event collection management
- Next steps: hits, then digis, then geometry

Production use of Objectivity/DB

- ATLAS MUST gain more production experience with Objectivity
- Only current production use is for tile calorimeter testbeam data
  - U.S. led tile calorimeter pilot project 
- This will change after control framework can be used for production
- Propose to continue to use tile testbeam as testbed for core db technologies

Production use of Objectivity/DB

- Content development and functional extensions will come from subsystem efforts, not core
- Propose to extend work to joint calorimeter testbeam if available effort within subsystems can be identified

FY2000 Support Request

- Argonne
  - s David Malon (1.0 FTE)
  - s John Christiansen (0.5 FTE)
  - s Guy Pandola (0.5 FTE)
  - u Global ATLAS database domain design, coordination, and planning
  - u Database development infrastructure
  - u Database components in support of control framework
  - u Initial database tasks in support of Geant4 simulation

FY2000 Support Request

- **Brookhaven**
  - **New hire (0.5 FTE)**
  - **Event model and event data in support of control framework and calorimeter reconstruction**
- **University of Michigan**
  - **New hire (0.5 FTE)**
  - **Detector description and related database work; muon subsystem prototyping of core database technologies**

Summary

- **The U.S. is well integrated into global ATLAS database efforts**
- **The U.S. has been asked to provide leadership to the overall ATLAS database enterprise, and to coordinate database efforts for two detector subsystems**

Summary

- Have chosen tasks well matched to U.S. responsibilities in other areas
- Have endeavored to delegate content-specific work and prototyping to subsystems
- Priorities
  - support for overall ATLAS database design, coordination, and planning
  - database development infrastructure

Summary

- Priorities continued
  - database components to support control framework, including event model prototyping and event data access
  - participation in detector description infrastructure DESIGN (some tool implementation at Michigan)
  - initial database efforts in support of simulation
  - production use of Objectivity via testbeam data

Summary of Milestones

- **Control/database infrastructure milestones, event model milestones coupled to control framework release milestones**
  - **May 2000 prototype reconstruction framework**
  - **Sep 2000 alpha reconstruction framework**
  - **Oct 2001 V2 reconstruction framework**
- **Detector description milestones**
  - **Mar 2000 XML DTD design review**
  - **Jun 2000 Complete subsystem physical geometries in XML**
  - **Oct 2000 generic model supporting geometry and logical detector organization**
  - **Mar 2001 readout geometries specified and available via generic model**
  - **Jul 2001 detector description via Objectivity/DB**

Summary of Milestones

- **Databases and testbeams**
  - **use of testbeam data as testbed for core technologies is an ongoing activity; Chicago/Argonne/CERN/Protvino collaboration for tile testbeam data is well underway**
  - **Summer 2000 databases for joint calorimeter testbeam IF effort can be identified within subsystems**
  - **Summer 2001 prototyping of conditions databases (alignment) in muon testbeam**
- **Database support for simulation**
  - **Oct 2000 initial (limited) database support for ATLAS Geant4 simulation**
  - **Oct 2001 database support for Monte Carlo events, hits, digitizations, and geometry**