Data Management Overview

David M. Malon

malon@anl.gov

U.S. ATLAS Computing Meeting

Brookhaven, New York

28 August 2003

Outline

- Technology and technology transition
- POOL and POOL integration
- Detector description and primary numbers
- Interval-of-validity databases and conditions
- Data Challenge databases: Magda, AMI, metadata, and virtual data
- Near-term plans
- Interactions with online, and with Technical Coordination
- Staffing
- Conclusions

Technology transition

- ATLAS database strategy has been, consistently. to employ “LHC common solutions wherever possible”
  - Until last year this meant Objectivity/DB as the baseline technology
  - Objectivity/DB conversion services retained as a reference implementation until developer releases leading to 6.0.0; retired at the end of 2002
- Today’s baseline is LCG POOL (hybrid relational and ROOT-based streaming layer)
  - ATLAS is contributing staff to POOL development teams
  - All ATLAS event store development is POOL-based
- Transition period technology: AthenaROOT conversion service
  - Deployed pre-POOL; provided input to persistence RTAG that led to POOL
  - AthenaROOT service will, like Objectivity, be retired once POOL infrastructure is sufficiently mature

What is POOL?

- POOL is the LCG Persistence Framework
  - Pool of persistent objects for LHC
- Started by LCG-SC2 in April '02
  - Common effort in which the experiments take over a major share of the responsibility
    - for defining the system architecture
    - for development of POOL components
  - ramping up over the last year from 1.5 to ~10FTE
- Dirk Duellmann is project leader
  - Information on this and the next several slides borrowed from him
- See http://pool.cern.ch for details

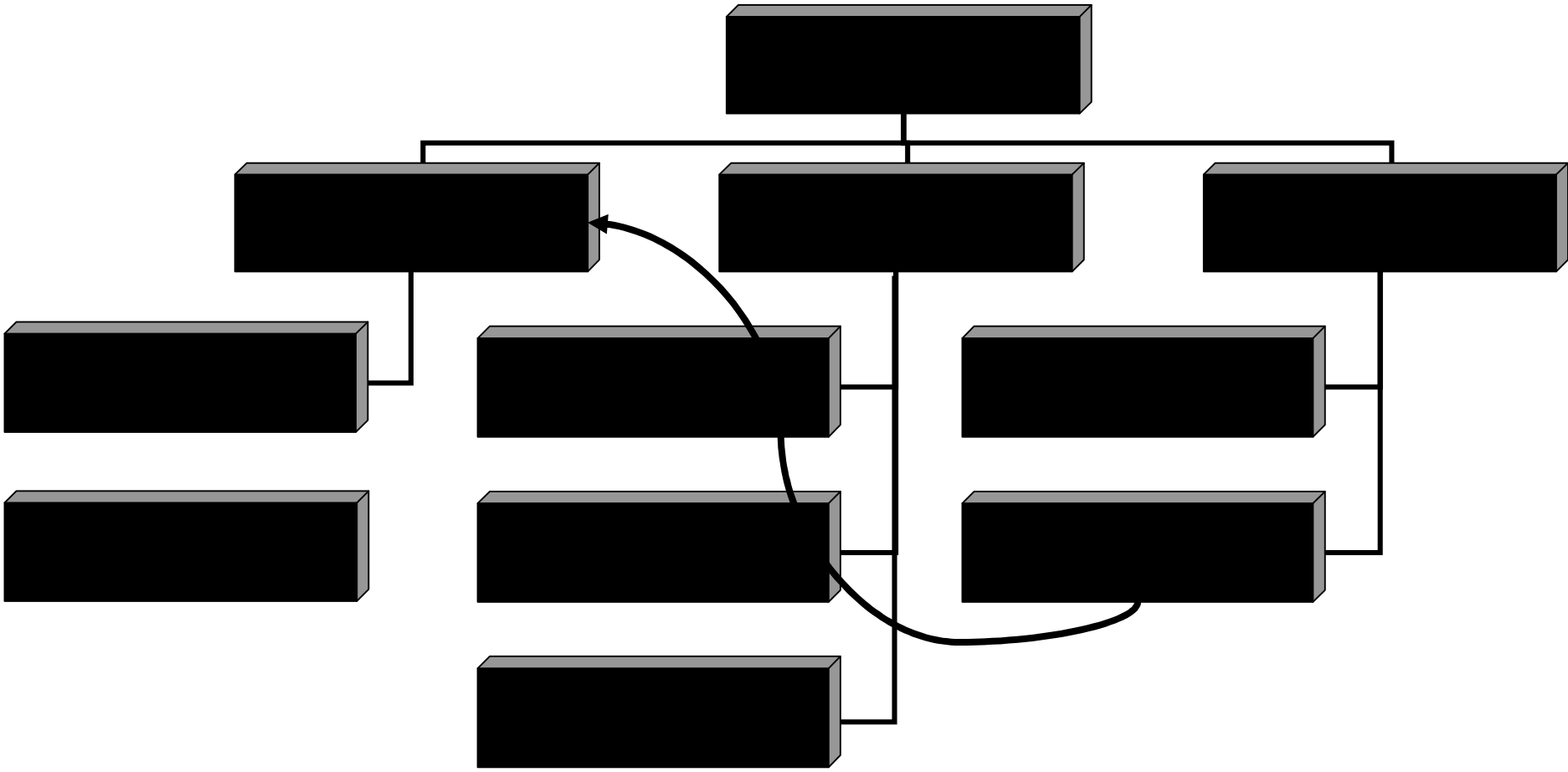
POOL and the LCG Architecture Blueprint

- POOL is a component-based system
  - A technology-neutral API
    - Abstract C++ interfaces
  - Implemented reusing existing technology
    - ROOT I/O for object streaming
      - complex data, simple consistency model
    - RDBMS for consistent metadata handling
      - simple data, transactional consistency
- POOL does not replace any of its components technologies
  - It integrates them to provides higher level services
  - Insulates physics applications from implementation details of components and technologies used today

POOL Work Package breakdown

- Storage Service
  - Stream transient C++ objects into/from storage
  - Resolve a logical object reference into a physical object
- File Catalog
  - Track files (and physical/logical names) and their descriptions
  - Resolve a logical file reference (FileID) into a physical file
- Collections and Metadata
  - Track (large, possibly distributed) collections of objects and their descriptions (“tag” data); support object-level selection
- Object Cache (DataService)
  - Track and manage objects already in transient memory to speed access

POOL Internal Organization



POOL integration into ATLAS

- AthenaPOOL conversion service prototype is available in current releases
- Scheduled to be ready for early adopters in Release 7.0.0
  - Based upon this month's "user release" of POOL
- POOL releases have been pretty much on schedule
  - Current release, 1.2.1, incorporates most recent LCG SEAL release
- Several integration issues are resolved in a stopgap fashion; much work remains
- Further LCG dictionary work (SEAL) will be required to represent ATLAS event model

ATLAS POOL/SEAL integration

- Many nuisance technical obstacles to POOL integration into ATLAS
- Not long-term concerns, but in the first year, they consume a great deal of time
  - Integration of POOL into ATLAS/Athena has not been trivial
- Examples
  - Conflicts in how cmt/scram/ATLAS/SPI handle build environments, compiler/linker settings, external packages and versions, …
  - Conflicts between Gaudi/Athena dynamic loading infrastructure and SEAL plug-in management
  - Conflicts in lifetime management with multiple transient caches (Athena StoreGate and POOL DataSvc)
  - Issues in type identification handling between Gaudi/Athena and the emerging SEAL dictionary
  - Keeping up with moving targets (but rapid LCG development is good!)

ATLAS contributions to POOL

- ATLAS has principal responsibility for POOL collections and metadata work package
- Principal responsibility for POOL MySQL and related (e.g., MySQL++) package and server configuration
- Also contributing to foreign object persistence for ROOT
- Contributions to overall architecture, dataset ideas, requirements, …
- Related: participation in SEAL project's dictionary requirements/design effort
- Expect to contribute strongly to newly endorsed common project on conditions data management when it is launched

Notes on relational technology

- POOL relational layer work is intended to be readily portable—to make no deep assumptions about choice of relational technology
- Collections work is currently implemented in MySQL; file catalog has MySQL and Oracle9i implementations
- Heterogeneous implementations are possible, e.g., with Oracle9i at CERN and MySQL on small sites
  - CERN IT is an Oracle shop
  - Some planning afoot to put Oracle at Tier1s, and possibly beyond
- Note that non-POOL ATLAS database work has tended to be implemented in MySQL; like POOL, avoiding technology-specific design decisions

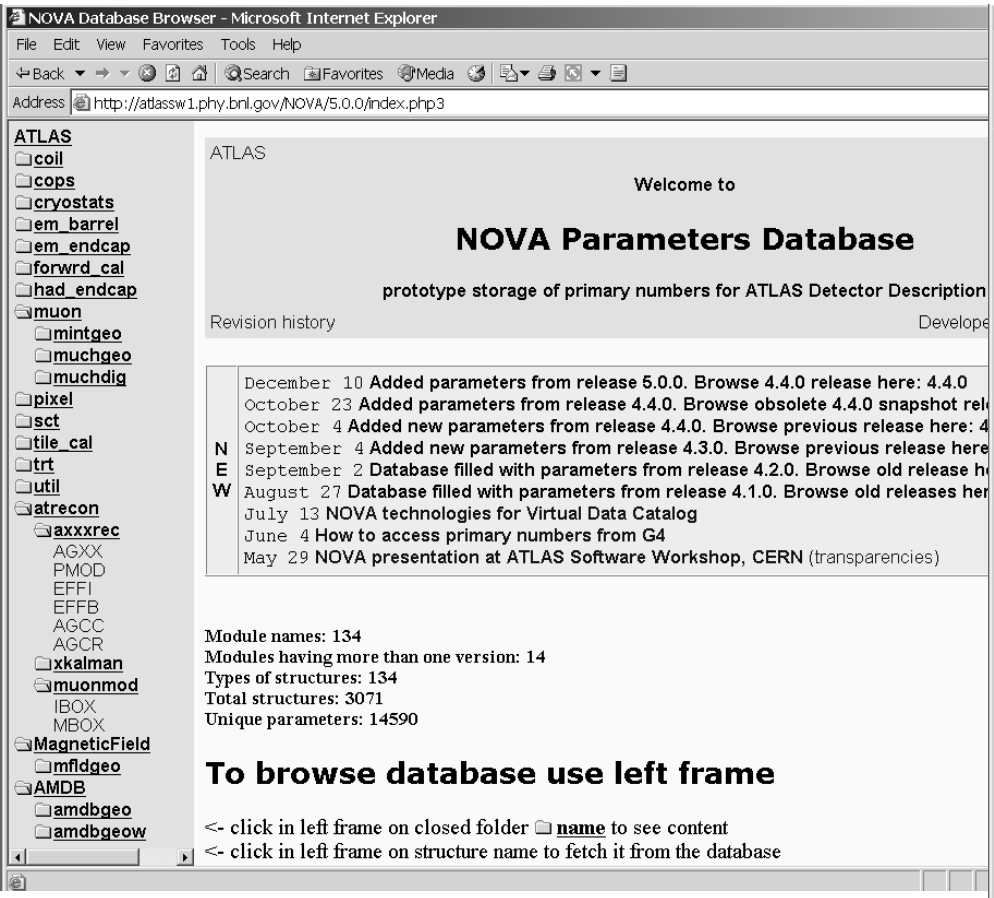
Detector description databases

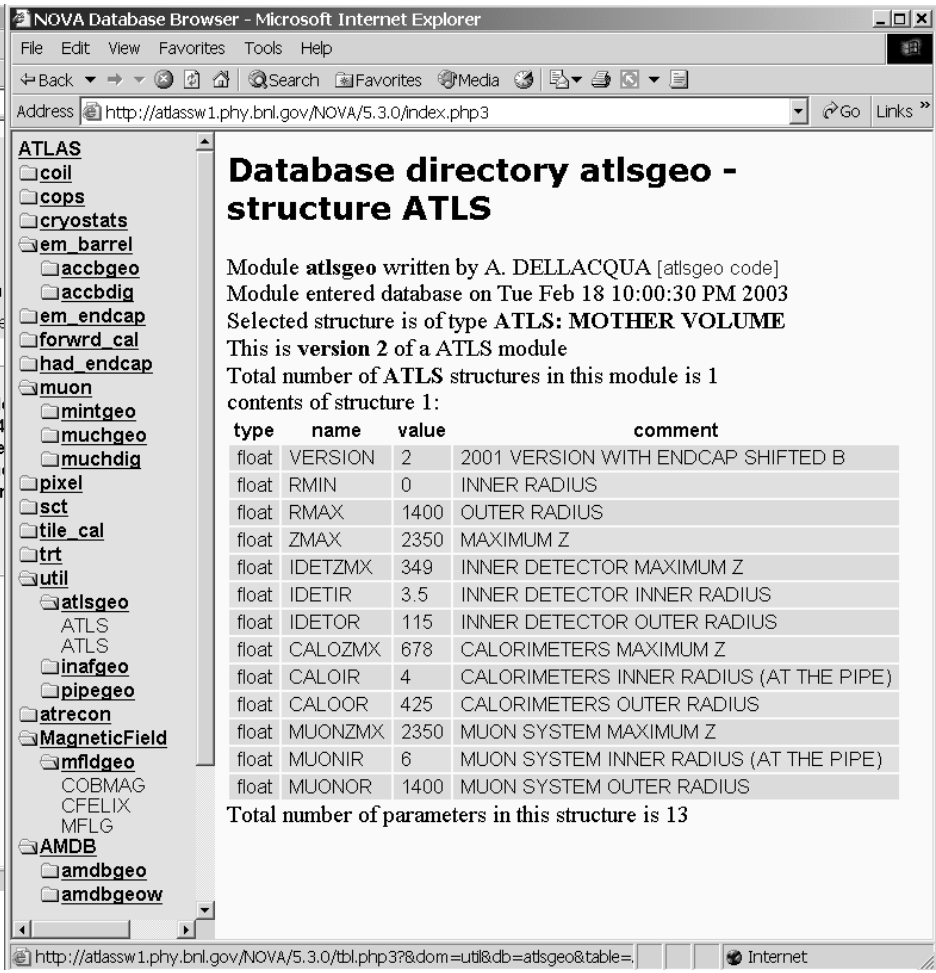
- "Primary numbers" (numbers that parameterize detector description) database deployed based upon NOVA
  - Leverages externally developed software
    - Begun as a BNL LDRD project (Vaniachine, Nevski, Wenaus)
    - Current implementation based upon MySQL
  - NOVA also used for other ATLAS purposes
- Used increasingly in Athena directly (NovaConversionSvc), via GeoModel, and by standalone Geant3 and Geant4 applications
- New data continually being added
  - Most recently toroids/feet/rails, tiletest data, materials

NOVA work

- Automatic generation of converters and object headers from database content integrated into nightly build infrastructure
- Work needed on input interfaces to NOVA, and on consistent approaches to event/nonevent data object definition
- Sasha Vaniachine is the principal contact person

NOVA browser screenshots





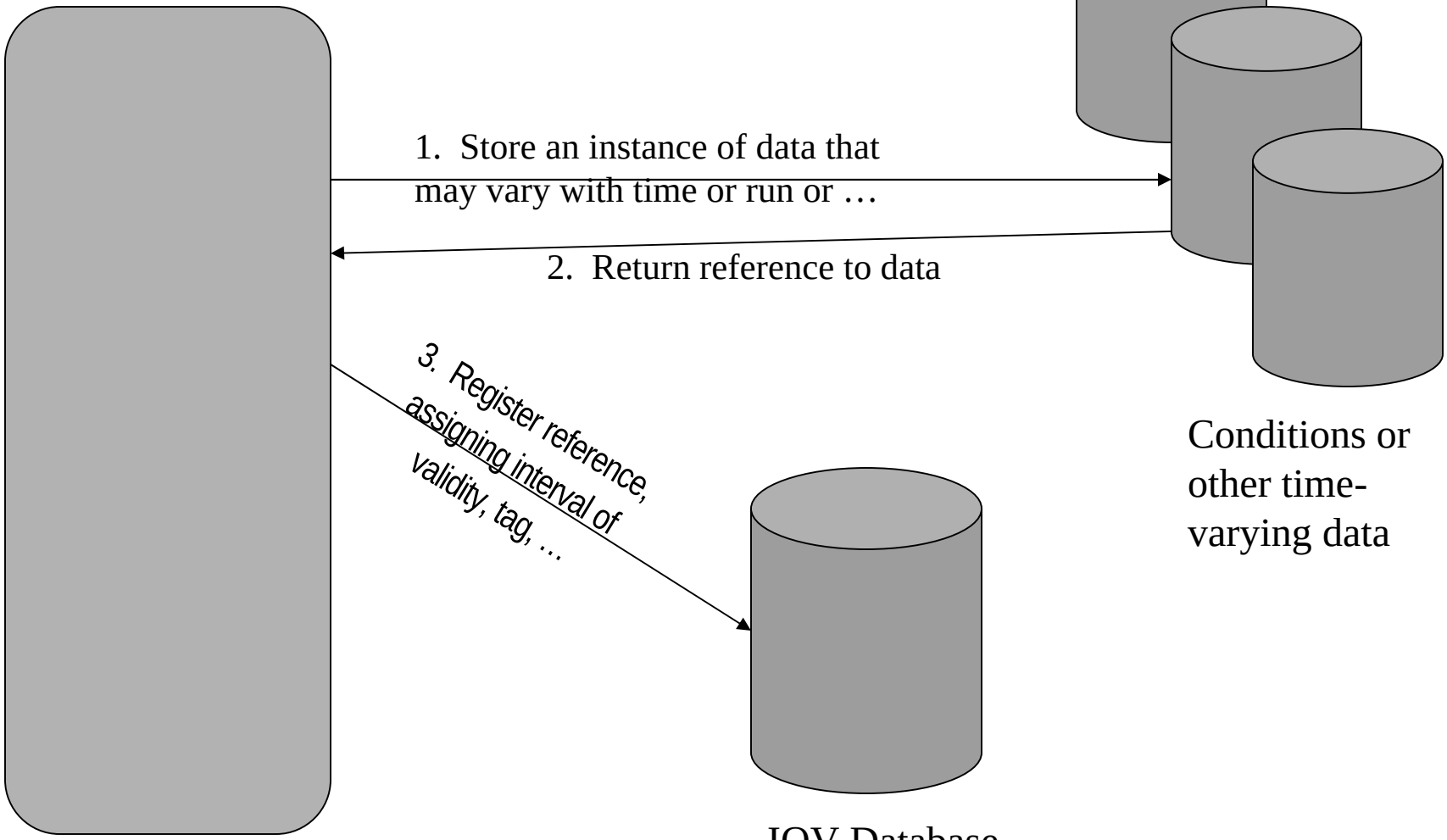
Interval-of-validity (IOV) databases

- ATLAS was a principal contributor of requirements to an LHC conditions database project organized under the aegis of RD45
- LCG SC2 in June endorsed a common project on conditions data, to begin soon, with this work as its starting point
- Lisbon TDAQ group has provided a MySQL implementation of common project interface
- ATLAS online/offline collaboration
  - Offline uses Lisbon-developed implementation in its releases as an interval-of-validity database
  - Offline provides an Athena service based upon this implementation
- Prototype (with interval-of-validity access to data in NOVA) in releases now
- Due for early adopter use in Release 7.0.0

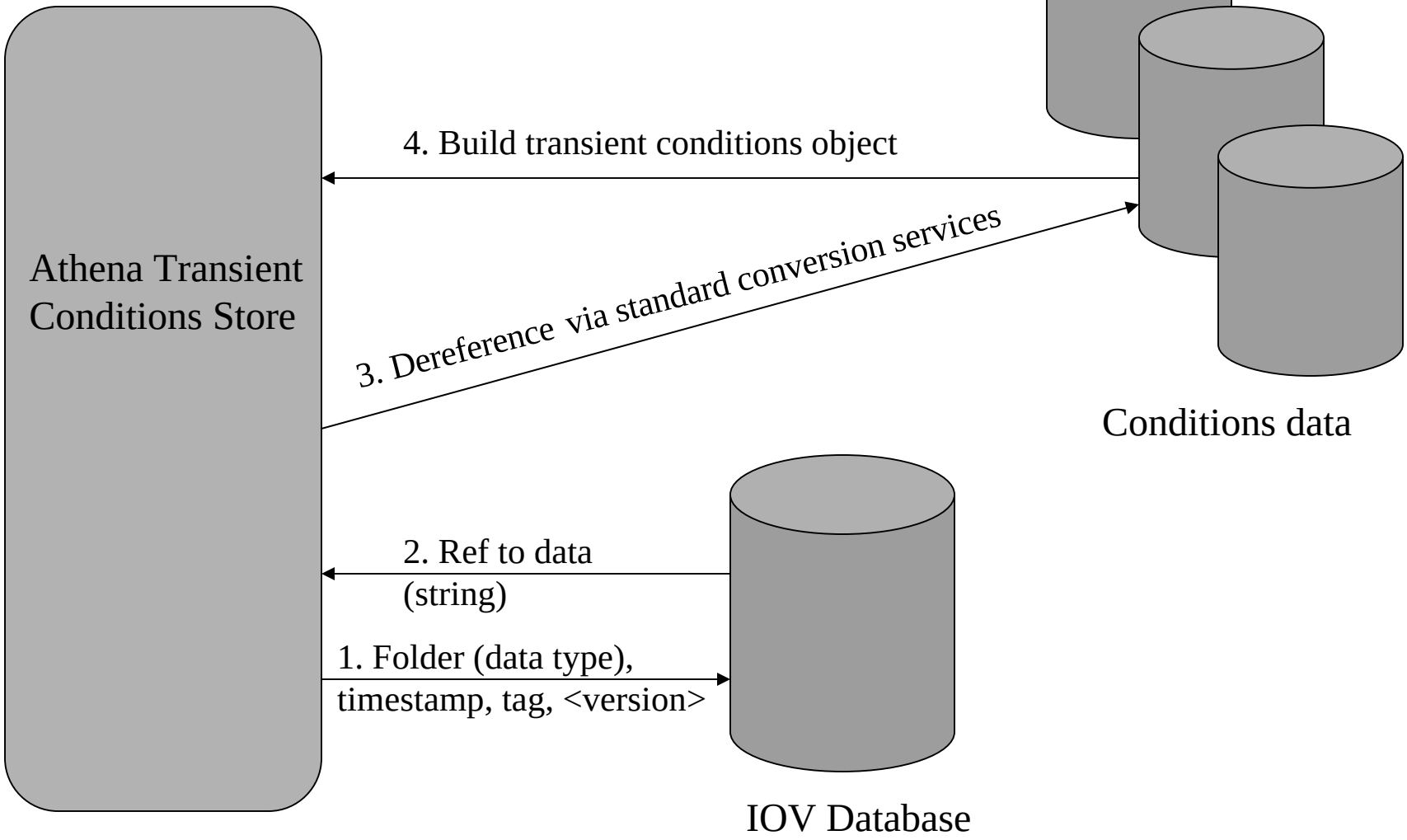
IOV databases

- ATLAS database architecture extends “usual” thinking about conditions service
- Interval-of-validity database acts as a registration service and mediator for many kinds of data
- Example:
  - Geometry data is written in an appropriate technology (e.g., POOL), and later “registered” in the IOV database with a range of runs as its interval of validity
  - Similarly for calibrations produced offline
  - No need to figure out how to represent complex objects in another storage technology (conditions database) when this is a problem already solved for event store technology (POOL)

Conditions Data Writer



Conditions data client



Conditions and IOV people

- LAL Orsay (Schaffer, Perus) is leading the IOV database integration effort
- LBNL (Leggett) provides the transient infrastructure to handle time validity of objects in the transient store (StoreGate) with respect to current event timestamp
- Hong Ma (liquid argon) is an early adopter of the Athena-integrated IOV database service
- Joe Rothberg (muons) is writing test beam conditions data to the database *outside* of Athena, for later Athena-based processing

Conditions Data Working Group

- A Conditions Data Working Group has been newly commissioned, headed by Richard Hawkings (calibration and alignment coordinator)
- Charged with articulating a model for conditions/calibration data flow between online/TDAQ and offline, including DCS, for understanding and recording rates and requirements, and more—not just conditions data persistence
- Contact Richard (or me, I guess) if you'd like to contribute

Production, bookkeeping, and metadata databases

- Data Challenges have provided much of the impetus for development of production, bookkeeping, and metadata databases
- Strong leveraging of work done under external auspices
- MAGDA (BNL) used for file/replica cataloging and transfer
  - Developed as an ATLAS activity funded by PPDG
  - Magda/RLS integration/transition planned prior to DC2
- AMI database (Grenoble) used for production metadata
  - Some grid integration of AMI (e.g., with EDG Spitfire)
- Small focused production workshop held at CERN earlier this month to plan production infrastructure for Data Challenge 2
  - Rich Baker, Kaushik De, Rob Gardner involved on the U.S. side
  - Report due soon

Metadata handling

- ATLAS Metadata workshop held 23-25 July in Oxford
- Issues:
  - Metadata infrastructure to be deployed for Data Challenge 2
  - Integration of metadata at several levels from several sources
    - Collection-level and event-level physics metadata
    - Collection-level and event-level physical location metadata
    - Provenance metadata
    - …
  - Common recipe repository/transformation catalog to be shared among components
- Workshop report due soon (overdue, really)

Virtual data catalogs

- Data Challenges have been a testbed for virtual data catalog prototyping, in AtCom, with VDC, and using the Chimera software from the GriPhyN (Grid Physics Networks) project
- Shared “recipe repository” (transformation catalog) discussions are underway
- Recent successes with Chimera-based ATLAS job execution on “CMS” nodes on shared CMS/ATLAS grid testbeds
- More from the grid folks (Rob Gardner?) on this
- Work is needed on integration with ATLAS database infrastructure

Near-term plans

- Focus in coming months: deploy and test a reasonable prototype of ATLAS Computing Model in the time frame of Data Challenge 2
  - Model is still being defined—Computing Model working group preliminary report due in October(?)
  - Note that DC2 is intended to provide an exercise of the Computing Model sufficient to inform the writing of the Computing TDR
- Ambitious development agenda required
  - See database slides from July ATLAS Data Challenge workshop
- Tier 0 reconstruction prototype is a principal focus, as is some infrastructure to support analysis

Event persistence for Tier 0 reconstruction in DC2

- Persistence for ESD, AOD, Tag data in POOL (8.0.0)
  - ESD, AOD, Tag not yet defined by reconstruction group
- Athena interfaces to POOL collection building and filtering infrastructure (7.5.0)
- Physical placement control, and placement metadata (to support selective retrieval) (7.5.0)
- Support for writing to multiple streams (e.g., by physics channel) (7.5.0)
- Support for concurrent processors contributing to common streams (8.0.0)
- Cataloging of database-resident event collections in son-of-AMI database, and other AMI integration with POOL (7.5.0)
- Magda←→RLS←→POOL integration (7.5.0++?)

Conditions database development

- Extensions needed to Athena/StoreGate to support writing calibrations/conditions from Athena
  - AthenaPOOL service should be capable of handling the persistence aspects by Release 7.0.0/7.1.0
- Work underway in the database group on organizing persistent calibration/conditions data, infrastructure for version tagging, for specifying in job options which data are needed
  - Limited prototype capabilities in Release 6.3.0
  - Responsibility for model moving to new calibration/alignment coordinator
- Some exercise of access to conditions data varying at the sub-run level by Tier 0 reconstruction is planned for inclusion in DC2

Conditions database futures

- LCG conditions database common project will start soon
- ATLAS development agenda will be tied to this
  - Expect to contribute strongly to common project requirements and development
- Some DCS and muon test beam data are already going into the ATLAS/Lisbon implementation of the common project interface that will be the LCG conditions project's starting point; liquid argon testing of this infrastructure also underway

Coordination with online and TC

- Historically, demonstrably good at the component level (cf. the joint conditions database work with Lisbon), but largely *ad hoc*
- Now formalized with new ATLAS Database Coordination Group commissioned by Dario Barberis, with representation from online/TDAQ, offline, and Technical Coordination
- Conditions Data Working Group also launched, with substantial involvement from both offline and online
- Successful joint conditions data workshop organized in February in advance of this

Staffing

- Current census: small groups (2-3 FTEs) at Argonne, Brookhaven, LAL Orsay; about 1 FTE at Grenoble working on tag collector and AMI databases for production
  - Enlisting involvement from British GANGA team; cf. metadata workshop at Oxford last month
  - U.S. ATLAS computing management has worked hard to try to increase support for database development, but we all know how tough the funding situation is
  - Trying to leverage grid projects wherever possible
- Lack of database effort at CERN is conspicuous, and hurts us in several ways
- Data Challenge production and support is a valuable source of requirements and experience, but it reduces development effort
- Not clear that expected staffing levels will allow us to meet DC2 milestones
- More on this at the LHC manpower [sic] review next week

Conclusions

- Work is underway on many fronts: common infrastructure, event store, conditions and IOV databases, primary numbers and geometry, metadata, production databases, …
- Many things will be ready for early adopters soon (7.0.0/7.1.0)
- Development agenda for Data Challenge 2 is daunting
- We need help, but we cannot pay you
- If we survive DC2, look for persistence tutorials at the next U.S. ATLAS computing workshop