ATLAS Analysis

Roger Jones
130 petabytes
600k datasets in various formats
355 million files
Various processing task categories
800 active users
Several analysis frameworks – no single “analysis chain” to preserve
For analysis, the most efficient level is to focus on the AOD/D3PD level
Slimming and skimming off common input datasets (e.g. at AOD or “D3PD” level)
All code and datasets used in a publication should be preserved under existing policy

ATLAS has no approved level-3 formats for external use, and such release will require such approval

We are concerned that anything released be useful as information, & not consume large amounts of collaboration effort (both in production and response)

As such, tools like Recast are more attractive
  The information incorporates the efficiency, acceptances and corrections – so is robust
  It also helps meet the internal requirement of full documentation of analyses
Plan to merge the AOD (end of recon chain) with the D3PD
Still have multiple derived DPDs representing skim/slim/thin and augmentation
Smart framework needed for this step
Documentation etc

- Paper & software documentation in twikis, indico, CDS
- Software in SVN
- Metadata on datasets in AMI
- Frontier conditions database access if needed
Prototype of one part of the analysis chain

Leverage existing tools and services where possible
- taking advantage of environment preservation lessons (e.g. DESY automated services)

Start with:
- A prototype service in ATLAS
- An existing analysis environment at a Tier 2 center
- Existing datasets on disk that can be stored locally or via FAX (federated sites)
- Validate, preserve and analyse the skim/thin/skim step
Sustained Usability Picture

Processing environment (capture):

Preserve

Time evolution

\[ \Delta t \]

\[ \Delta t \]

\[ \Delta t \]

\[ \Delta t \]
Slim/Skim/Thin & 
Preservation

Web interface at CERN
gets requests, shows their status

OracleDB at CERN
Stores requests, splits them in tasks, 
serves as a backend for the web site

Curation server
receives web queries, collects info on 
datasets, files, trees, branches

Executor at Tier2
gets tasks from the DB, 
creates, submits HTCondor SkimSlim jobs 
makes and registers resulting DS
Point of entry to collect skim & slim operation

Build & define a validation process from this

Execute using existing virtualized execution infrastructure at the Tier 2

Is data “still alive” monitoring infrastructure
Re-use, implement exe environment
Prototype Status & Outlook

- Focus on one aspect of analysis preservation
- Meet the experiment where it currently is:
  - Slim & Skim service
  - Production Tier 2 center (Midwest Tier 2) virtualization services
- Leverage existing preservation environments from other experiments & labs
- Use to identify the hard issues (technical and policy)
- Already spoken of as part of the new analysis model
Analysis Practical steps - RECAST

- Framework developed to extend impact of existing analyses
- Candidate for within-experiment and long-term analysis archival, encapsulating the full trigger & event selection, data, backgrounds, systematics
- Allow an existing analysis to be reinterpreted under an alternate model hypothesis
  - Complete information from original analysis, including the tacit information, contained in the data
  - Not optimized for the new model, but more reliable than a naïve reanalysis?

Recast seen as a very promising solution for preserving analyses and useful, cost effective preservation of information – addresses levels ~1~3
Test case analyses of different types being input to Recast