

# Generator code support

- Atlas structure
- Code in CVS
- Structure of code
- Releases
- Tag collector
- Validation
- Issues for release 14.
-

# Historical remarks

- This may help explain why we have the current situation and how things fit
- Athena was adopted as the Atlas framework in ~2000.
- Almost all atlas simulations for Physics TDR used Pythia.
  - Pythia was the only generator fully integrated into the old fortran framework. ATLAS never used StdHep. Simulation ran off pythia common blocks
- All ATLAS code in new framework is C++ (except for some muon reconstruction code in F90)
- Generators was the user package inside Athena
  - I was MC group convener and involved with Athena development
    - I decided to use Generators as a use case
    - Fast simulation (atlfast) was first client.
  - We might now make some different decisions, but...

# Historical remarks II

- All generators were in fortran: needed a common interface to simulation
  - ATLAS never used HEPEVT (used the pythia common directly)
  - No candidate C++ event record
  - HepMC was an atlas product (written by Matt Dobbs and Hans-Jorgen Beck).
    - Two parts, The record itself and the I/O packages that are used to fill it from the generators
- We have MCEventCollection that contains a set of HepMC events
  - Thought that this might be needed for pileup or event mixing
  - In practice there is only one event in the collection.
- CLHEP agreed to take over HepMC and manage it
  - Serious problems with timely bug fixing: I/O packages not adopted and replacement packages not adequate
  - HepMC now removed from CLHEP, maintained separately in LCGApps

# Historical remarks III

- GeneratorEvent appears in Storegate with as McEventCollection with Key GenEvent: Particles with  $\text{ctau} > 1\text{mm}$  are declared stable, these are decayed by simulation: mostly  $K_S$  and Lambda. Note B is decayed in the generator
- This McCollection is read by simulation
- Four simulation packages
- Full Geant based
  - Copies GenEvent to TruthEvent and feeds this to Geant. Adds Geant secondaries to HepMC and uses GenParticle BarCode (int) to distinguish them
- Parameterized G4 (same as G4): new in 13.0.20
- Atlfast: simple parameterized response with simplified geometry using stable particles from generator
- Atlfast II. Full tracking, parameterized calorimetry, full geometry; new in 13.0.20
- There are a set of Truth Tools to enable users to extract Truth information: Used to calculate Jet and Emiss truth for example.

# ATLAS organization I

- Monte Carlo software is a separate project within Simulation:  
Coordinator is Adele Rimoldi
  - Atlas SPMB is primary technical coordination body for offline software, chaired by David Quarrie
  - Meets alternate Mondays at 17:00 (phone only), Meets on Friday PM of software weeks in person (when combined with CMB(computing management board))
  - Discusses issues across domains
  - There are very few generator issues that need to go to this level
    - HepMC interface changes (which affect everyone)
    - You will be invited as needed to report when you have an issue
    - Contact Adele or David in this case
- Adele will need to know who is the contact person. (ACTION FOR THIS MEETING)

# ATLAS organization II

- There is an Atlas Monte Carlo group: part of physics coordination
- Conveners: Jon Butterworth (ends 30 Sept 2008), Borut Kersevan (ends 30 Sept 2007), Osamu Jinnouchi (1 Oct 2007 to 30 Sept 2009)
- Forum for Physics groups to bring issues, requests for new versions and tools
- Responsible for Physics validation
- Responsible for deciding defaults of parameters: “tuning”
- Responsible for requesting code changes
- Meets monthly (approx) and always during atlas and T/P weeks
- You will need to report software issues at these meetings and discuss with them plans for releases, code changes, new tools etc.
-

# Code organization

- .External packages
  - The external generator libraries (e.g. Pythia) exist outside of atlas CVS.
  - Most are built by the LCG Generators Project
    - [/afs/cern.ch/sw/lcg/external/MCGenerators/](#)
    - Managed by Witek Pokorski (who is also an atlas member)
    - Used by other LHC experiments
    - Older releases used GENSER, which is now obsolete
    - [/afs/cern.ch/sw/lcg/app/releases/GENSER/](#)
    - There are some other external code in [/afs/cern.ch/atlas/external](#)
      - Used mainly for code that is not yet in MCGenerators
      - Testing and development, Bug fixes from MC authors.
      - The need for this area is now far less as the MCGenerators are more responsive to issues

# Code organization

- .External packages (continued)
  - There are “glue” packages in the atlas CVS under External which point to the external areas. Example  
External/Pythia/cmt/requirements contains pointers to external code
-



# Code structure/building

- The interface packages are all contained in CVS under Generators
- Each package has /src /Package (header files) /cmt (configuration files) /share (example scripts) /doc and perhaps /test
- Example
- CMT is used to build code.
-

# Mainline Packages

- Used in large scale production by many users
- You should maintain all the main packages with the following exceptions
  - Tauola/Photos. This is now maintained by Witek Pokoroski. We use this in all productions so it is critical.
- The remainder are: Pythia, Herwig, Tauola/Photos, Mc@NLO, Alpgen. MadGraph Sherpa (see later for discussion)
- All these are in extensive use.
-

# Note on “Les Houches” types

- Alpgen, Mc@NLO and some others use external files
- These are made by running Alpgen stand alone making ascii files in standard “LesHouches” format
- Files are then processed by Herwig (or Pythia).
- In these cases, the full generator is not integrated into Athena, only the code needed to process the ascii files.
- 
-

# Note on Sherpa

- Although Sherpa is not a “Les Houches” generator: it makes full events, we run it externally
- Users make files containing Sherpa events and then process them with ReadSherpa\_i
- Sherpa should be fully integrated and this external processing abandoned, but there are issues:
  - A new set of events in sherpa requires a compilation step making .so that are then used by generation
  - Libraries are huge
  - Needed release patching: this problem is now solved, it is no harder than adding a new script
- Joe Virizi has been working with the Sherpa authors and he should be consulted
- The ReadSherpa\_i does work so replacing it is not high priority.

# Other Packages

- Hijing: Used for Heavy Ions
  - Managed by Misha Lelchouk/
    - They will require your help occasionally.
  - Not in production in release 12.
  - Special requirements due to large event sizes
- Isajet
  - You have to maintain this but
    - Not used much
    - No interface changes for many years
    - Should be updated for release 14 (low priority)
    - Author is member of ATLAS: F.Paige
-

# Other Packages II

- PythiaB
  - Modified version of Pythia used by B-physics group
  - More efficient generation of certain exclusive final states
  - Maintained by Maria Smizanska and James Catmore
  - You need to inform them of any changes in Pythia as they depend on it.
  - You do not need to maintain this, but will have to monitor bug fixing: It was not validated in release 11 and not used for production
  - Works in release 12

# Other Packages III

- ATLAS policy requires that all code used for publications must be available to all collaborators
- Simulation event samples must be made with validated code in a release. Therefore, there are some packages that are used by few people. **They are required to implement code in these cases**
  - The policy is that you have to aid them to get the package working, and put it into releases
  - They then maintain them.
  - If the code fails at some point, you have to decide whether to drop it from a release if the persons fail to maintain it.
  - If the code becomes of “general use”, you may have to take it over. The Monte Carlo group makes the final decision and sets priorities

# Other Packages IV

- GeneratorFilters
  - Set of code used to filter events at generator level
  - Example: W to enu and require e to be in tracking volume
  - Most of this code comes from physics groups as it is specific to process
  - You should receive this code, check it and tag it in releases
  - The code is stable, there have been no new algs for some months
- TruthHelpers
  - A set of predicates to be used with the HepMC iterators to aid users
  - IsGenStable: stable in the generator, excludes geant secondaries
  - IsGenInteracting: used in etmiss truth



# Other Packages V

- ParticlePropertiesService (Charles Leggett)
  - Part of Athena Core not Generators
  - You need to be aware of it
  - Used mainly by Geant and digitization, to identify properties of stable particles
  - Based on HepPDT
  - Recent changes to deal with Heavy Ion G4 secondaries.
  - Cannot be used as replacement for internal particle properties of Pythia
  - Could be used to enforce values between generators: Force same W mass for example
    - Appealing but never tried, not easy and always were higher priorities

# Other Packages VI

- Random Numbers
  - We use Athena Random Service to manage all random numbers
    - Used to be based on ranlux: problems with correlations;
      - Only relevant for single particle production
      - Solved by making 1M events per job!
    - Ranecu recently added as an option.
  - Service is part of AthenaCore
    - Stavropolous maintained it
    - Probably not an issue unless it breaks or bugs appear.

# Releases

- The code is divided into Projects which can be built separately
  - Projects do not map simply onto CVS
  - Almost all Generators are in Simulation project
  - The glue packages are in External.
  - Sometimes a dependency forces a package to be in an illogical place

# Releases II

- Nomenclature is A.B.C.D
  - A is major release
  - B=0 for a production release
  - C=0 for a development release (usually)
  - D is Cache number for patching after a release is built.
  - Cache is required for the production system (could be empty)
- Current activity
  - 12.0.7 Last validated production release. 12.0.7.2 is about to be released
  - 13.0.20. Next candidate production release. In bug fix mode. 13.0.20.1 has run, 13.0.20.2 to be built this week.
  - 13.10.0 development will become 14.0.10 as production candidate

# Releases III

- There are incremental nightly build for use in testing.
- You can put anything in development branch
- The bug fix branch (A.0.C) accepts tags only after they are approved
  - Need to be justified by Savannah bug report.
  - There are two nightlies (Val and XX). If succeeds in one, then can go to the other
  - Procedure is used so that the full build usually works first time
    - A full check out and build takes 24hrs
- There is an automatic mailing system that reports errors from the builds
  - The mails go to the package managers.
  - Can go to more than one person. You should have one person who gets all of them, should be the same person dealing with technical validation. I suggest not to remove anyone for the time being (Borut and I get most of them)

# Tag collector

- Management of tag collection is done via tag collector.
- All tags go via this web based system
  - Can be frustrating to use (my opinion)
  - I suggest that only one person does this
- Procedure is to make CVS tag then collect it
- <https://atlastagcollector.in2p3.fr:8443/AMI/servlet/net.hep.atlas.Database.Bookkeeping.AMI.Servlet.Command>

# Platforms

- We only support Linux
- Several SLC/Gcc versions
  - Occasional panics about “fortran going away”
  - 32 bit and 32bit compatibilty on 64bits
  - Still some issues with 64bit native
    - Mostly solved now for compilation
    - Not used for production: memory use is too large
  - I recommend RTT testing to deal with platform issues
- Some GRID sites have odd linux versions (Mandrake in Nordic)
  - Validation problems: example Herwig results used to be site dependent: Ok at the moment but
  - Single event validation is an issue

# Validation

- Nightly tests
  - Can only run very small numbers of events
  - Pointless if the code has not changed
  - Used mainly by packages with large dependency
  - Not used by Generators; Few dependencies, code does not change rapidly
- RTT (run time tester)
  - Runs against nightlies and recent builds
  - Robustness and reliability have been an issue: improved recently.
  - Has been used by Generators. Not extensively used.
- Full chain.
  - Runs same configuration as production jobs.
  - Used for final checks on integration before large scale production starts



# Validation II

- Large scale grid production (order 1M events)
- Validation reporting
  - Meeting every Tuesday by phone at 16:10
  - Alternates between technical and physics validation
  - Physics Validation is responsibility of the Monte Carlo group
  - Technical validation is your responsibility
-

# Technical validation

- Coordinator is Manuel Gallas
- You need to have a contact person
  - Will receive bug reports and issues from Full chain: these will be sent by Manuel
  - Should report on open issues at each meeting (Contact Manuel for details)
-

# Physics validation

- Coordinator is Wouter Verkerke
  - Main focus is performance
  - Expect to find only rare
  - Monte Carlo group reports to this.
  - You do not have to report here, usually there will not be issues relevant to you as they will have been fixed earlier
  - This validation is based off a large scale production sample
    - You should monitor this sample to ensure that all tools that are to be used in production are being used in the validation samples
  -
-

# Bug reporting

- Savannah is used
- You need someone to manage the generators area and assign/fix/close bugs
  - Probably should be same person reporting to Manuel.
-

# Documentation

- Supposed to use Doxygen, but
  - Your clients are users not developers
  - Examples in /share area are most useful
  - Pdf in /doc linked to MC group pages. Tagged with releases
  - Documentation standdown took place after Stavropoulos left: But no complaints so it is probably OK
  - Tutorials are also available: you may be asked to update or give these as needed.
    - Originally prepared by Stavropoulos
    - Catmore is in overall charge of tutorials
- Check occasionally that physics workbook has no obvious problems
  - You are not responsible for this but if it is wrong it will cause you trouble
- After release 14 goes into bug fix mode, I suggest you review all the documentation

# MC Production issues (evgen only)

- This is mainly the responsibility of the MC group production manager
- Production samples are required to use code in releases
  - Including scripts needed to run jobs
- Production system is based on “Job Transforms”
  - Python scripts that wrap athena jobs
  - Are now used by (new) users who want to run existing code outside the prod system
- Evgen wrapper script is `csc_evgen_trf.py`
  - Takes as input other (sample specific) scripts that specify the particular sample
  - Physics group production managers prepare the Sample specific fragments, validate them and then submit for inclusion in CVS
  - These go into a Cache (usually every two weeks) that is installed on top of a production release

# MC Production issues (evgen only) II

- The scripts are different in release 12 and release 13 as job configuration changed: migration to “configurables”
- Note that release 11 and older used bash scripts (these are not completely gone as some people are still using old releases)
- Release 12 in Generators/DC3\_joboptions/
- Release 13 in (release 14 will use same area)
- Provides a provenance for all MC samples and a set of examples for users.
- We are no longer accepting new samples with release 12
  - Force validation of 13
  - New evgen samples can be made with 13 and backported to 12 for simulation if needed (must use ascii format to do this)

# MC Production issues (evgen only) III

- The core scripts and the basic python system was set up by Martin Woudstra (still on atlas)
- In release 13, the responsibility was devolved to the package owners
- You have to maintain the core. Borut Kersevan has been doing this and might be willing to continue. Changes to the core should now only be needed for new generators
- The tagging of the fragments and submission to the Cache is done by the MC production manager who also defines the jobs



# MC Production issues (evgen only) IV

- Most evgen production (pythia, herwig) uses only scripts as input
- However “Les Houches” accord generators are different
  - Production of the 4 partonic 4 vectors takes place outside Athena
    - Done by “experts”
    - Files put on grid in DDM
    - Athena grid jobs then process them with Herwig/pythia and make pool files
    - Can use enormous amounts of CPU, may need specialized jobs and file concatenation
  - This procedure will be reviewed over the next 6 months or so:
    - There are serious arguments in favor of bringing everything inside Athena

# Release 12 status

- The current validated production release is 12.0.7
- The last cache is 12.0.7.2 whose release is imminent
- There will be no more 12.0.7 caches: any remaining bugs will not be fixed
-

# Release 13 status

- Differences from release 12
- Changes in python means
  - Release 12 Jobopts do not work with 13
  - Changes needed to production scripts
- Generators worked in 13.0.10 (with some exceptions)
- 13.0.20.1 has produced validation samples
  - Expect feedback at today's validation meeting
  - You may have to fix some bugs
  - **ACTION.** Please check that all the examples in the /share areas actually work,
    - Some of them may not have been migrated to “configurables”

# Release 14 issues/actions

- Build date unclear: depends on progress in 13 validation: End of October??
- Development uses 13.x.0 branch
  - Expect to use this for first data
- Must migrate the LCG version of HepMC
  - This is most critical item
  - Migration needed as new C++ generators use this version
  - Interfaces different (5 vectors rather than 4 vectors).
    - Schema evolution must be managed.
      - Existing data from releases 11, 12 and 13 must be readable
    - Since the C++ generators are not ready, this could be delayed if there are problems
- Migrate Mc@nlo to new version (released 3 weeks ago)

# HepMC migration

- HepMC has two parts
  - HepMC and HepMCIO
  - HepMC should move to the lcg version
  - HepMCIO
    - Loads the event record from the fortran generators (HEPEVT) into HepMC
    - Lack of a true HEPVT standard means a mess
      - Problems with Tauola/Photos, particularly with Herwig
      - Problems with Pythia handling of documentaries: mstp(128)=? (atlas has used all possible values at some times in the past!)
      - HEPEVT used internally in HERWIG.
    - I suspect that the IO packages in lcg do not work properly and the ATLAS ones should be used.
      - It is too late to go through a full validation.
      - Ours are well tested
      - Let CMS debug the lcg ones!

# Release 14 issues/actions II

- Pythia and Herwig versions are stable (no changes needed):
- Tauola/Photos stable.
- May need not to migrate LhaPDF
- Sherpa
  - Investigate running entirely within Athena.
- EvtGen
  - B-decays inside Pythia/Herwig not used by Tevatron
  - Should this be default for Pythia and Herwig?
    - B-tagging efficiency depends on it at 20% level!
  - Supported by LCG MCGenerators
  - Interface to Herwig/Pythia exists (Juerg Beringer) in release 12: not used for extensive production
  - Has been used by B-physics group for exclusive modes (this used private version of EvtGen)

# Release 14 issues/actions III

- Recently Phojet is now being used by minbias group: needs to be updated/fixed. Was not used recently.

# New C++ generators

- These will need to be incorporated at some point
- .Sherpa
  - Already running in release 12
  - Expect Sherpa to support HepMC fully: ascii format could then be dropped
  - Run in Athena and make pool files directly
  - Release 14?
- Herwig ++
  - Jeremy Lys has interacted with authors and reported in MC meetings
  - Code is not ready for production use in 14
  - May be stable enough to put in release 14 for testing purposes
  - Makes HepMC and used LCG version, so HepMC migration must happen first
  - Make decision based on release 14 schedule and feedback from MC group
  - Interface should be easier than fortran version: parameter config comes from external file: we can deal with this in the same way as “python fragments”



# New C++ generators

- Pythia 8
  - Needs feedback from MC group
  - Uses HepMC
  - Not for production in 14.
  - Not stable enough for testing?

# People who are involved

- Ian Hinchliffe: acted as Manager, interface to simulation and worked on code as needed. (Approx 30% effort prior to 1 October 2006, less since then). I can be consulted. After 1 October, will remain as physics production manager.
- Giorgos Stavropolous. Full time. Left Atlas June 2007. Joined Nestor, can be contacted in case of questions.
- Borut Kersevan: Outgoing MC convener General expert. Good relations with Pythia team. Recently has worked mainly on production script migration for release 13. Acts as Monte Carlo production manager. Production script maintenance should be taken over by you. MC group will need decide about production manager. Borut has worked extensively on the HepMC IO packages.
- Maria Smizanska/James Catmore. Maintains Pythia B. **Continuing**
- Juerg Beringer (EvtGen). You should discuss with him what he intends to do
- Sebastien Binet: wrote persistency and tools to remove particles from HepMC in consistent way: available for consulting
- Witek Pokorski. LCG generators. Tauola and Photos. Will continue.
- Misha Lelchouk/Andrzej Olszewski: Heavy Ions: continuing.

# Migrating generator versions

- Follow recommendations from Monte Carlo Group
  - Only bug fixes are allowed in bug fix release: major version changes must wait for next major release.
  - Fortran versions of Herwig/Pythia are frozen, only bug fixes.
  - Others

# Persistency

- Atlas uses Pool
- Since all generators are wrapped in the same McEventCollection using HepMC, one persistency solution works for all generators
- We convert from double (transient) to float (persistent) to save space: important for AOD where MC truth is 30% of total.
- There is a transient/persistent split to allow for schema evolution
  - This will have to be used when you move to the CLHEP version of HepMC
  - Existing data from release 11, 12 and 13 must be readable with future releases