

LCG Applications Area

LCG Launch Week

12 March 2002

Torre Wenaus, BNL/ATLAS

LCG Applications Project Manager

torre.wenaus@cern.ch

<http://cern.ch/lcg/peb/applications>



Launch Week - Applications WG Agenda

- **WG1 (Applications) Tuesday morning Chair: John Harvey**
 - 9:00 - 9:15 Introduction, overview - John Harvey (15')
 - 9:15 - 10:00 Applications subproject - Torre Wenaus (45')
 - 10:00 - 10:30 Discussion (30')
 - Issues related to common project organization & execution
 - 10:30 - 11:00 Coffee
 - 11:00 - 11:20 Process RTAG - Fons Rademakers (15')
 - 11:20 - 11:40 Math library review RTAG - Fred James (15')
 - 11:40 - 12:00 Persistency Framework RTAG - David Malon (15')
 - 12:00 - 12:30 Discussion (30')
 - Work program, RTAGs, next steps

RTAG = Requirements and Technical Advisory Group



Personal Introduction

- ATLAS since 2000; before that STAR/RHIC (computing leader), BaBar (simulation leader), Geant4, GEM/SSC, L3/LEP
 - At CERN (L3/MIT) 1986-1992
- Physics Applications Software (PAS) Group Leader in the BNL Physics Dept
- US ATLAS Software Manager
- ATLAS Planning Officer
- Particle Physics Data Grid ATLAS Team Leader
- To sustain US responsibilities I will be on a cycle of 3 weeks at CERN, 1 week at BNL
 - Have a deputy, David Adams (BNL), for PAS leader and US ATLAS software manager
 - Will not continue as PPDG ATLAS team lead



Applications in Launch Workshop

- **Goals**
 - Agree on project scope, priorities
 - Develop high-level work plan (goals and milestones)
 - Define concrete next steps
 - Foster spirit of productive collaboration
- **Approach to this presentation**
 - Try to build a basis for productive, non-trivial discussion by avoiding bland generalities
 - while also avoiding surprises borne of inadequate consultation
 - Be as specific as possible in presenting possible common project areas; high level work plans; resource estimates, needed and available; prioritization; and what available resources + prioritization imply for what can get done
 - Proposals seen by SC2 members, experiments & project participants and adjusted in light of feedback
 - Make more specific work plan comments on areas covered by existing RTAGs, commensurate with level of presentations from RTAGs



Talk Outline

- Status overview
- Applications Area role, scope, Phase 1 activities
- Possible organization
- Applications work planning: RTAGs
- Project execution approach
- Personnel
- Next steps
- Supplemental slides on candidate RTAGs
- Questions/discussion during the talk is welcome



Status Overview

- In pre-PEB meetings, received Applications area input from the experiments
- Some Applications-related RTAGs established
- Applications Area Manager appointment in place
 - Supported by BNL (DOE) and CERN
 - Started (officially) March 1
 - Based at CERN from March 10
- Direct and email discussions with experiment computing coordinators, IT/API, IT/DB, EP, software developers on applications area, organization, RTAGs
 - I am observer in existing applications-related RTAGs
- Getting going on 'generic' planning
- Avoiding 'jumping the gun' with respect to SC2
 - 'High level workplan' in the form of 'candidate RTAG' suggestions in this talk
- Survey of existing and anticipated resources
- Hiring of people into LCG positions is in progress



Project Execution Board Role

The PEB

- acts on SC2 recommendations for common projects
- develops and gets agreement on strategy and workplan
 - goals, milestones, deliverables, schedule, resource allocation, prioritization
 - assembly of and buy-in from implementation teams
 - SC2 approval of strategy and workplan
- manages and tracks the progress and direction of the project
- ensures conformance with SC2 recommendations
- identifies areas for study or resolution by SC2

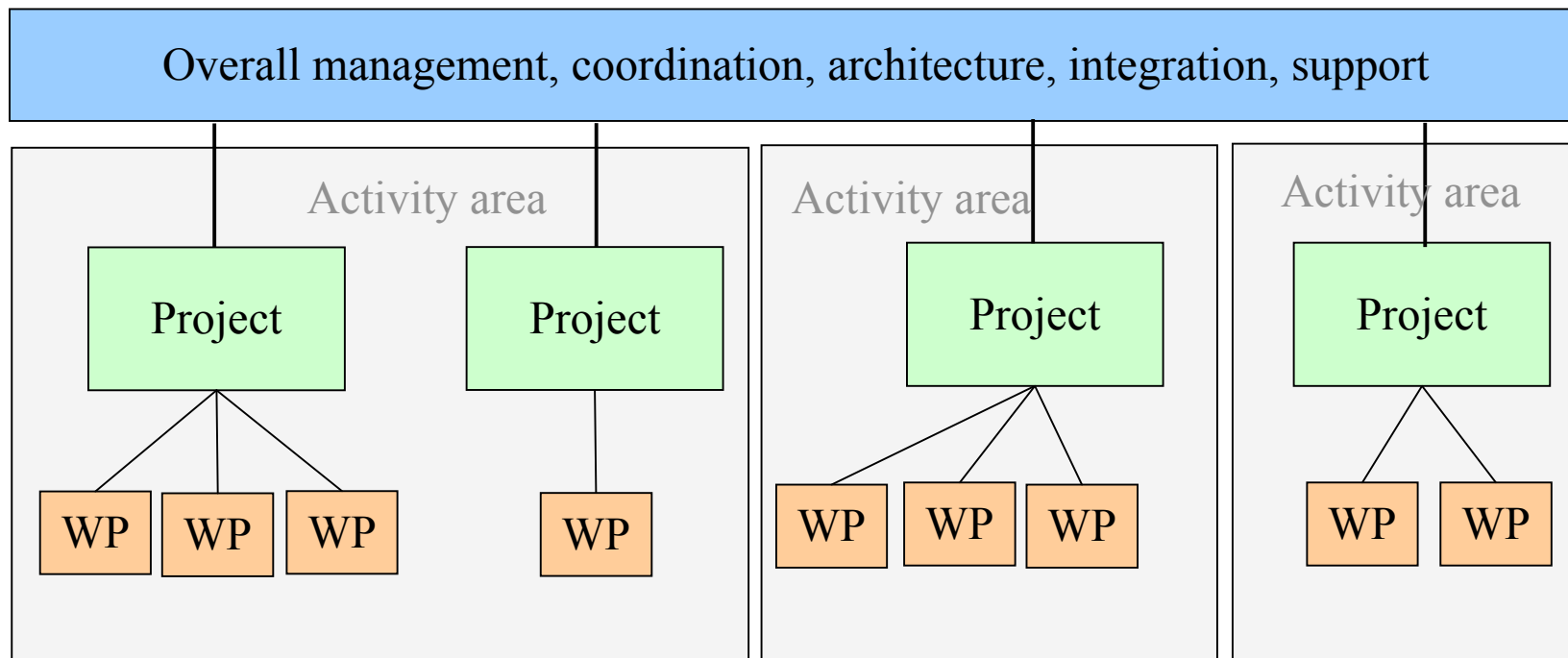


Applications Scope: Activity Areas

1. **Application software infrastructure**
 - physics software development environment, standard libraries, development tools, compiler expertise
2. **Common frameworks for simulation and analysis**
 - Geant4, other simulation codes and related needs
 - Development, integration of analysis toolkits & components
3. **Support for physics applications**
 - Dev, support of common software tools & frameworks
4. **Grid interface and integration**
 - Portal environment
 - Direct assistance to experiments at core/grid interface
 - **Adaptation of physics applications to grid environment**
5. **Physics data management**
 - Event data, metadata, conditions data, analysis objects



Possible Organization of Applications Activities



Example: **Activity area:** Physics data management
Possible projects: Hybrid event store, Conditions DB, ...
Work Packages: Component breakdown and work plan lead to Work Package definitions. ~1-3 FTEs per WP



Possible Organization of Applications Activities – Management & Architecture

- **Applications area manager**
 - Overall management and coordination
 - Planning and overseeing a coherent architecture for Applications software and infrastructure
 - Will enlist help in this from experiment architects (in particular)
- **CERN line managers (Applications-involved IT, EP groups)**
 - Personnel line management for CERN-based participants
 - Technical management of project activities comes from the project, in close consultation with line management
- **Project leaders**
 - Responsibility for planning and executing the project
 - Project is 'hosted' by the experiment/group of its leader
- **Work package team leaders (where team size warrants)**
 - Responsible for designing, developing work package deliverables



Applications area activity in LCG Phase 1

- **Prepare the LHC computing environment**
 - Provide the common tools and infrastructure for the physics application software
 - Grid interfaces to application software
 - Integrate applications software into progressively more complex grid prototypes to validate the computing model
 - Seek opportunities for re-use of project results
- **Participate in and support experiment data challenges**
 - Deploy and evaluate tools, infrastructure and grid interfaces in data challenge production and analysis environments
- **Participate in writing the LCG Technical Design Report**
 - Will describe the full LHC Computing Grid to be built in Phase 2



Applications work planning: RTAGs

- The PEB acts on SC2 recommendations for common projects
- First PEB Applications work planning activity is recommending common project areas to SC2
 - i.e. suggesting RTAGs
 - and associated rough schedule and prioritization
- Here follow
 - Summary of experiment input received in the PEB
 - Summary of existing RTAGs
 - Suggestions for new candidate RTAGs
 - With proposed timeline (prioritization)
 - Details in supplementary slides



Experiment Input on Applications

- Had input from four experiments in a pre-PEB meeting
- Good, specific suggestions for Applications area scope and execution received. The following are taken directly from the experiment talks presented to the PEB:
- Acknowledge existing efforts from experiments to define common activities
- Support for external packages: XML parser, Python,...
- Support for physics packages: Event generator tools, CLHEP, particle properties
- Support for ROOT, FLUKA
- Event data management: ROOTIO and metadata
- Conditions data management: review existing project



Experiment Input (2)

- More input taken directly from experiment talks to the PEB:
- Analysis tools, assessment of Anaphe, AIDA
- Software development tools
- Common components: data dictionary, geometry
- Time scales important: requirements should be received by PEB quickly and work started
- Clear schedules, milestones and deliverables
- Early deployment and frequent release iterations



Existing RTAGs in the Applications Area

- RTAG activity launched in some areas seen as high priority by the experiments (SC2)
- Persistency framework
 - Interim report
- Process for managing LCG software
 - Essentially complete
- Math library review
 - Interim report
- Grid RTAG being formed, relating partially to Applications
- No details here... see later talks
- Just a few 'execution' comments on persistency framework...



Persistency Framework Comments

- All experiments express interest in participating in the persistency framework project following on from the RTAG
 - Participation will have to be quantified soon
- Timescale for delivery of a basic (but production-capable) hybrid event store is short, but should be attainable
 - ATLAS: Late September 2002 (DC1 phase 2)
 - CMS: Next year (but interest in something usable by November this year)
 - ALICE: end of the year
 - LHCb: middle 2003
- I think completing the full, scalable persistency framework will consume the duration of LCG phase 1



Candidate RTAGs in the Applications Area

- The process by which the PEB becomes active in a given area begins in the SC2
 1. SC2: define and establish RTAG
 2. RTAG: develop requirements and submit report to SC2
 3. SC2: digest report and submit final requirements to PEB
 4. PEB: organize and execute the project
- So, the approach taken here to suggesting a 'high level workplan' without stepping on this process is to list candidate RTAGs and a corresponding prioritized timeline, as input to this workshop and the SC2
 - List is fine grained; timeline indicates some possible mergings
 - Some staggering where strong overlap of people is likely (but not enough if I've underestimated the overlap)
 - Balance between over-fragmentation and overly broad, complex RTAGs is difficult
- Comments on candidate RTAGs in supplementary slides



Candidate RTAGs

Simulation tools	Non-physics activity; ask SC2
Detector description, model	Description tools, geometry model
Conditions database	If necessary after existing RTAG
Data dictionary	Key need for common service
Interactive frameworks	What do we want, have, need
Statistical analysis	Tools, interfaces, integration
Visualization	Tools, interfaces, integration
Physics packages	Important area but scope unclear
Framework services	If common framework is too optimistic...
C++ class libraries	Standard foundation libraries



Candidate RTAGs (2)

Event processing framework	Hard, long term
Distributed analysis	Application layer over grid
Distributed production	Application layer over grid
Small scale persistency	Simple persistency tools
Software testing	May be covered by process RTAG
Software distribution	From central 'Program Library' to convenient broad distribution
OO language usage	C++, Java (..?) roles in the future
Benchmarking suite	Comprehensive suite for LCG software
Online notebooks	Long term; low priority



Candidate (& existing) RTAGs by Activity Area

- **Application software infrastructure**
 - Software process; math libraries; C++ class libraries; software testing; software distribution; OO language usage; benchmarking suite
- **Common frameworks for simulation and analysis**
 - Simulation tools; detector description, model; interactive frameworks; statistical analysis; visualization
- **Support for physics applications**
 - Physics packages; data dictionary; framework services; event processing framework
- **Grid interface and integration**
 - Distributed analysis; distributed production; online notebooks
- **Physics data management**
 - Persistency framework; conditions database; small scale persistency
- **(Some assignments may be debatable)**



Comment on Post-RTAG Involvement of Experiment Experts and Architects

- RTAGs used for initial requirements and technical assessment
- Project then acts on what emerges from the SC2
- Experiments are then directly involved in project work, in 'hosting' projects, in management through the PEB, and in the feedback of a tight development cycle
- And there is the review/oversight process

- But a mechanism is needed to draw experiment experts & architects together with the project team when important issues & decisions arise during project execution
- Need not (should not?) be formal, and should not turn the RTAGs into permanent bodies
- e.g. meeting as necessary to expose an issue, gather input, and make a decision? Or use regular planning meetings?

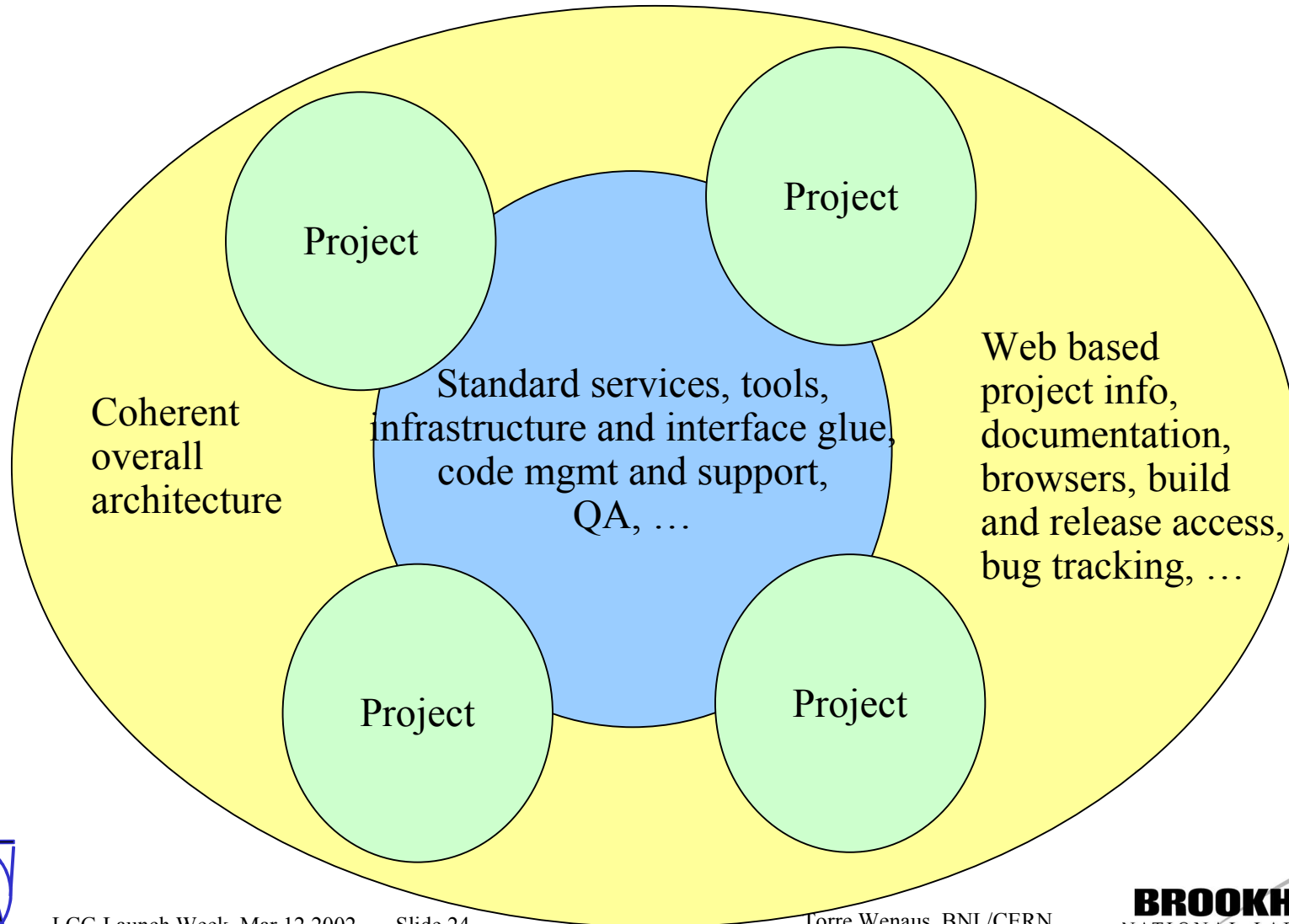


Project Execution Approach

- Incremental development with tight release cycle
 - Early exposure to users, regular feedback, iterative
- Full involvement of the experiments
 - Including 'hosting' projects in experiments, with host providing leadership, expertise, resources; initial prototyping and deployment environment; commitment to enfranchising other experiments at both development and deployment stages
 - All experiments closely involved in early specification phase
- Technical management of the project by the project
- 'Inclusive and responsive' culture
- Clear and open project planning and tracking
- (Non-commercial) source codes must be public
 - With an appropriate (LCG wide?) open source license
- Developers are designers, and vice versa
- Cohesion across projects; coherent overall architecture



Cohesion Across Projects



Centralized, Unified Support Services

- LCG common project software should be managed and maintained in a unified way at CERN
- Clearly establish CERN as a central repository for LHC software
- Clear, centralized guidance, documentation and distribution for third party components, both open and commercial
 - Guidance on remote usage: what needs to be installed (or bought), and how
 - Tools for very easy automated installation at remote sites. Any physicist should be able to do it
- Releases tested with validation/regression test suite covering all software
 - To be developed at the same time as the software



Software Environment Commonality

- 'Common culture' in software environment among the experiments should be an objective
 - Would considerably ease the sharing of common software
 - Software development centers participating in LCG common software development must use the LCG tools and environment
 - To the extent these are also the tools of the experiment, this is simplified
- A priority in implementing the LCG software infrastructure and process should be making its components available to and usable within the experiments themselves
 - For those experiments that wish to do so



Communication

- Web as communication medium will receive a lot of attention
 - *Current*, informative web content
 - Applications page is <http://cern.ch/lcg/peb/applications>
- Mailing lists will be open and archived
 - Applications mailing list, archive and HyperNews mirror is up; see the Applications web page
- Calendar of project related activities
- Work breakdown and schedule/tracking on web
- Extremely high threshold for making anything 'private'
 - I can't think of anything which should be private in the PEB Applications area



Oversight

- Project Manager gives brief PEB-wide report to SC2 monthly meeting
- Applications Area Manager gives ~30' report to SC2 meeting every 3 months
- Regular written reports
- LCG Project annual review
 - Stability on the annual review committee will aid in yielding helpful, insightful reviews



Workshops and Meetings

- Anticipate weekly Apps-wide meeting reviewing overall status, and project meetings on similar schedule
- Regular workshops covering project activities and planning
- In the near term:
 - Hybrid event store workshop
 - Highest priority project which will emerge from the Persistency Framework RTAG/SC2 guidance
 - Process/infrastructure implementation planning meeting
 - Review existing infrastructure and tools and plan changes
 - Guided by Process requirements from RTAG/SC2



Planning and Tracking

- Will use XProject software project planning tool for
 - Work breakdown of projects and work packages
 - Schedule, milestones, progress tracking
 - Resource loading, tracking personnel assignments
 - Developed in US ATLAS (by me) and in use by US & Int'l ATLAS, US CMS, under evaluation by others
- Serves as a front-end which will feed whatever project planning tool(s) are used PEB-wide
 - Can feed MS Project, PPT etc.
- XProject has been extended to support PEB Applications Area planning
- Placeholder (empty) work breakdown and schedule are in



Original Estimates of Personnel Requirements

- Original manpower estimates from Sep 2001 proposal
 - To be completely overhauled, but gives some idea of numbers and foreseen distribution
 - Numbers are 2002 targets
 - Estimate is of total resources needed
- Application software infrastructure - 5 FTEs
- Common frameworks for simu & analysis - 13 FTEs
- Support for physics applications - 9 FTEs
- Physics data management - 9 FTEs
- **Total: 36**
- **36 = (8 existing IT/API + 19 new LCG) + (5 existing IT/DB/PDM + 4 new LCG)**



Currently Identified Applications Personnel

- The situation is fluid and organizational structures are still being determined, but this is how it looks now
 - Fellows, associates are included
- IT/DB Physics Data Management section: 4 people
- IT/API Applications for Physics and Infrastructure
 - API-Geant4: 9 people
 - API-Anaphe: 6.5
 - CERNLIB: 1.5
 - Total API applications area: 17 people
- New LCG hires in apps: 6 identified, 2 arrived so far
 - Remainder arriving between now and October
- Total identified LCG applications personnel to date: 27



LCG Applications Area Personnel Contributions

- Commitments on the part of many countries to contributing applications area personnel
- Two particularly notable contributions - because they are large and quite well established
 - United Kingdom
 - 5 or 6 of first 10 hires in applications
 - Another round opening with ~15 further hires
 - Italy
 - Recruitment started for ~15 people
 - Should be good source of applications people
- Overall, total number expected is consistent with applications area target



Use of Manpower in the Immediate Term

- Tentative project assignments where abilities dictate
- Training, learning
- Setting up tools and infrastructure
 - Repositories and code management tools
 - Code documentation and browsing (LXR, cvsweb, bonsai?, ...)
 - Problem tracking (with Remedy)
- Documentation and web content



Next Steps

- ... for discussion and resolution at this workshop
- We want to get off to a quick start on setting up software process infrastructure and initiating (particularly) the persistency framework project once requirements emerge from SC2 in these areas
- RTAG process has to move quickly if project work is to be enabled in other areas



Supplemental slides

- On candidate RTAGs
- To the extent I have any time left in my talk, I will go over the first couple of these, and leave the rest
 - Available for the discussion following the RTAG talks
 - For online perusal and feedback



RTAG?: Simulation tools

- Geant4 is establishing a HEP physics requirements body within the collaboration, accepted by SC2 as a mechanism for addressing G4 physics performance issues
- However, there are important simulation needs to which LCG resources could be applied in the near term.
- By the design of LCG, this requires SC2 delivering requirements to PEB
- John Apostolakis has recently assembled G4 requests and requirements from the LHC collaborations
- *Proposal:* Use these requirements as the groundwork for a quick 1-month RTAG to guide near term simulation activity in the project, leaving the addressing of physics performance requirements to the separate process within Geant4



RTAG?: Simulation tools (2)

- Some possible activity areas in simulation, from the Geant4 requests/requirements received from the experiments, which would be input to the RTAG:
 - Error propagation tool for reconstruction ('GEANE')
 - Assembly and documentation of standard physics lists
 - Python interface
 - Documentation, tutorials, communication
 - Geant4 CVS server access issues
- The RTAG could also address FLUKA support
 - Requested by ALICE as an immediate priority
 - Strong interest expressed by other experiments as well



RTAG?: Detector geometry & materials description and modeling services

- Write the product specification for detector geometry and materials description and modeling services
 - Specify scope: eg. Services to define, provide transient access to, and store the geometry and materials descriptions required by simulation, reconstruction, analysis, online and event display applications, with the various descriptions using the same information source
 - Identify requirements including end-user needs such as ease and naturalness of use of the description tools, readability and robustness against errors e.g. provision for named constants and derived quantities
 - Explore commonality of persistence requirements with conditions data management
 - Identify where experiments have differing requirements and examine how to address them within common tools
 - Address migration from current tools



RTAG?: Conditions database

- Will depend on persistency RTAG outcome
- Refine the requirements and product specification of a conditions database serving the needs of the LHC experiments, using the existing requirements and products as a reference point. Give due consideration to effective distributed/remote usage.
- Identify the extent to which the persistency framework (hybrid store) can be directly used at the lower levels of a conditions database implementation.
- Identify the component(s) and interfaces atop a common persistency foundation that complete the conditions database



RTAG?: Data dictionary service

- Can the experiments converge on common data definition and dictionary tools in the near term?
- Even if the answer is no, it should be possible to establish a standard dictionary service (generic API) by which common tools can interact, while leaving free to the experiments how their class models are defined and implemented
- Develop a product specification for a generic high-level data dictionary service able to accommodate distinct data definition and dictionary tools and present a common, generic interface to the dictionary
- Review the current data definition and dictionary approaches and seek to expand commonality among the experiments. Write the product specifications for common (even if $N < 4$) components.



RTAG?: Interactive frameworks

- Frameworks providing interactivity for various environments including physics analysis and event processing control (simulation and reconstruction) are critical. They serve end users directly and must match end user requirements extremely well. They can be a powerful and flexible 'glue' in a modular environment, providing interconnectivity between widely distinct components and making the 'whole' offered by such an environment much greater than the sum of its parts.
- Develop the requirements for an interactive framework common across the various application environments
- Relate the requirements to existing tools and approaches (e.g. ROOT/CINT, Python-based tools)
- Write a product specification, with specific recommendations on tools and technologies to employ
- Address both command line and GUI interactivity



RTAG?: Statistical analysis interfaces & tools

- Address requirements on analysis tools
 - What data analysis services and tools are required
 - What is and is not provided by existing tools
- Address what existing tools should be supported and what further development is needed
 - Including long term maintenance issues
- Address role of abstract interfaces to statistical analysis services
 - Are they to be used?
 - If so, what tools should be interfaced to a common abstract interface to meet LHC needs (and how, when, etc.)
- Address requirements and approaches to persistency and data interchange



RTAG?: Detector and event visualization

- Examine the range of tools available and identify those which should be developed as common components within the LCG Applications architecture
- Address requirements, recommendations and needed/desired implementations in such areas as
 - existing and planned standard interfaces and their applicability
 - GUI integration
 - Interactivity requirements (picking)
 - Interface to visualizing objects (eg. Draw() method)
 - Use of standard 3D graphics libraries
- Very dependent on other RTAG outcomes



RTAG?: Physics packages

- Needs and requirements in event generators and their interfaces & persistency, particle property services, ...
- Scope of the LCG in this area needs to be made clearer before a well defined candidate RTAG can be developed



RTAG?: Framework services

- While converging on a common event processing framework among the LHC experiments may be impractical at least on the near term, this does not preclude adopting common approaches and tools for Framework services
 - Examples: message handling and error reporting; execution monitoring and state management; exception handling and recovery; job state persistence and recording of history information; dynamic component loading; interface definition, versioning, etc.
- Seek to identify framework services and tools which can be developed in common, possibly starting from existing products.
- Develop requirements on their functionality and interfaces.



RTAG?: C++ class libraries

- Address needs and requirements in standard C++ class libraries, with recommendations on specific tools
- Provide recommendations on the application and evolution of community libraries such as ROOT, CLHEP, HepUtilities, ...
- Survey third party libraries and provide recommendations on which should be adopted and what should be used from them
- Merge with Framework Services candidate RTAG?



RTAG?: Event processing framework

- There is no consensus to pursue a common event processing framework in the near term. There is perhaps more agreement that this should be pursued in the long term (but there's no consensus on a likely candidate for a common framework in the long term)
- This looks at best to be a long term RTAG
- Two experiments do use a common event processing framework kernel (*Gaudi*)
- Many difficult issues in growing N past 2, whether with *Gaudi*, *AliRoot*, *COBRA* or something else!



RTAG?: Interfaces to distributed analysis

- Develop requirements on end-user interfaces to distributed analysis, layered over grid middleware services, and write a product specification
 - Grid portals, but not only; e.g. PROOF and Jas fall into this category
 - A grid portal for analysis is presumably an evolution of tools like these
- Focus on analysis interface; address the distinct requirements of production separately
 - Production interface should probably be addressed first, as it is simpler and will probably have components usable as parts of the analysis interface



RTAG?: Distributed production systems

- Distributed production systems will have much common ground at the grid middleware level. How much can be done in common at the higher level of end-to-end distributed production applications layered over the grid middleware?
 - Recognizing that the grid projects are active at this level too, and coordination is needed
- Survey existing and planned production components and end-to-end systems at the application level (AliEn, MOP, etc.) and identify tools and approaches to develop in common
- Write product specifications for common components, and/or explicitly identify specific tools to be adapted and developed as common components
- Include end user (production operations) interface
 - Grid portal for production



RTAG?: Small-scale persistency & databases

- If not covered by the existing persistency RTAG, and if there is agreement this is needed...
- Write the product specification for a simple, self-contained, low-overhead object persistency service for small-scale persistency in C++ applications
 - Marshal objects to a byte stream which may be stored on a file, in an RDBMS record, etc.
 - In implementation, very likely a simplified derivative of the object streamer of the hybrid store
 - For small scale persistence applications, e.g. saving state, saving configuration information
- Examine the utility of and requirements on a simple, standard, easily installed and managed database service complementing the persistency service for small scale applications
 - MySQL, PostgreSQL etc are casually adopted for simple applications with increasing frequency. Is it possible and worthwhile to converge on a common database service



RTAG?: Software testing tools & services

- How much commonality can be achieved in the infrastructure and tools used
 - Memory checking, unit tests, regression tests, validation tests, performance tests
- A large part of this has been covered by the process RTAG



RTAG?: Software distribution

- May or may not be adequately addressed in the process RTAG
- Requirements for a central distribution point at CERN
 - A 'CERN LHC Program Library Office'
- Requirements on software distribution taking into account all tiers
- Survey and recommend on the various approaches, their utility, complementarity
 - Tarballs (DAR)
 - RPMs and other standard open software tools
 - Role of AFS, asis
 - Higher level automated distribution tools (pacman)



RTAG?: Evolution of OO language usage

- Long-term evolution of C++
- Role for other language(s), e.g. Java?
 - Near, medium and (to the extent possible) long term application of other languages among LHC experiments
 - Implications for tools and support requirements
- Identify any requirements arising
 - Applications, services to be developed in common
 - Third party tools to be integrated and supported
 - Compilers and other infrastructure to be supported
 - Libraries required



RTAG?: LCG Benchmarking suite

- Below threshold for an RTAG?
 - Every LCG application should come with a benchmarking suite, and should be made available and readily usable as part of a comprehensive benchmarking suite
- Develop requirements for a comprehensive benchmarking suite of LCG applications for use in performance evaluation, testing, platform validation and performance measurement, etc.
 - Tools which should be represented
 - Tests which should be included
 - Packaging and distribution requirements



RTAG?: Online notebooks and other remote control / collaborative tools

- Identify near term and long term needs and requirements common across the experiments
- Survey existing, planned tools and approaches
- Develop recommendations for common development/adaptation and support of tools for LHC

