LCG Applications Area – Overview, Planning and Resources

Torre Wenaus, BNL/CERN
LCG Applications Area Manager

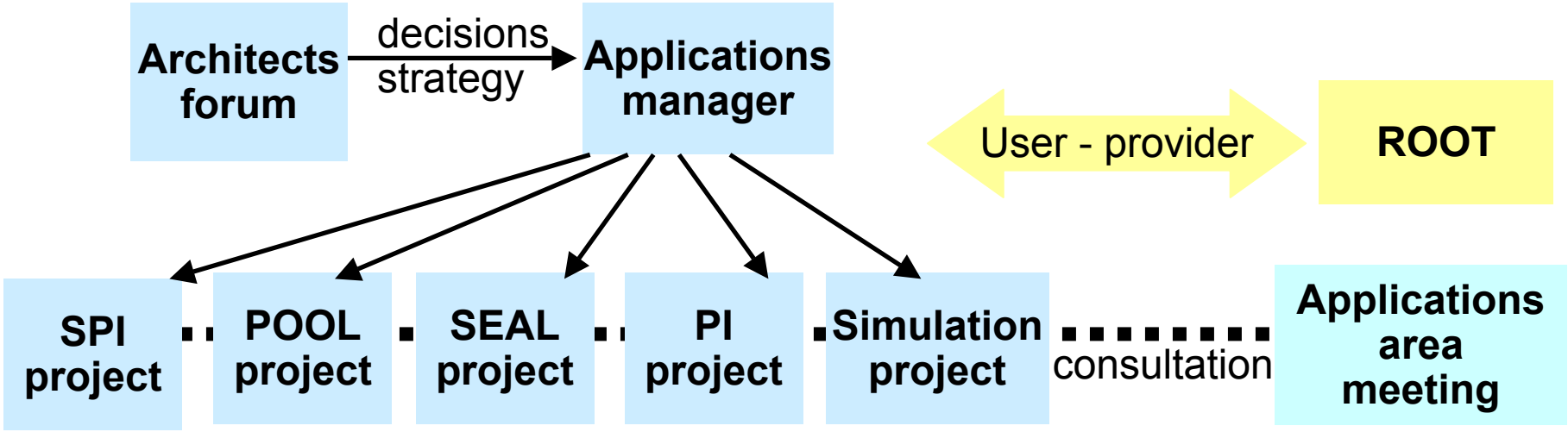
http://lcgapp.cern.ch

Applications Area Internal Review
October 20, 2003

Outline

- Applications area organization and overview
- Implementing the Architecture Blueprint
- Project planning
- Applications area projects
- Personnel resources, participation
- Concluding remarks

Applications Area Organisation



Management and Communication

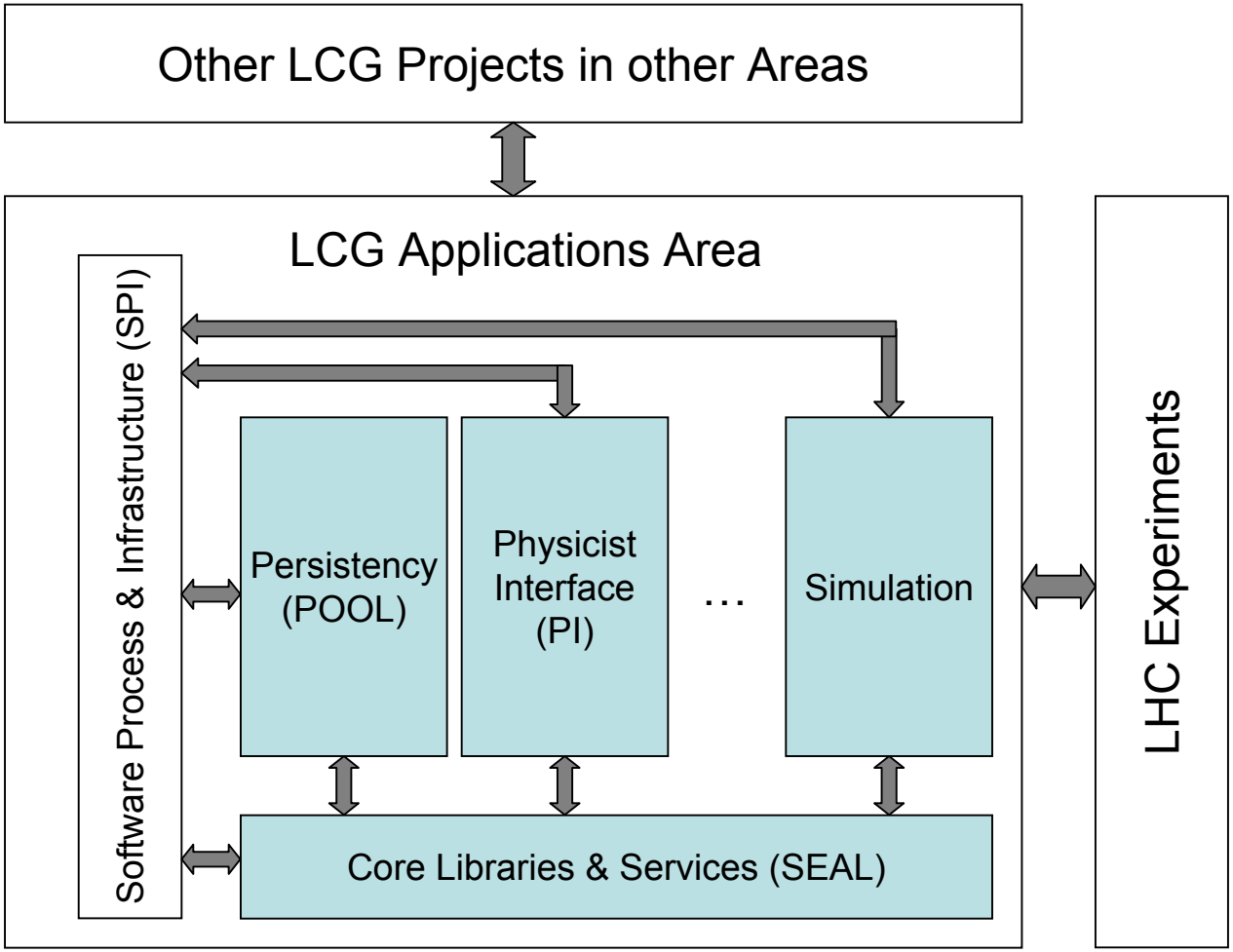
- Architects Forum
  - Attendees: architects, project leaders, EP/SFT leader
    - Computing coordinators invited; attend occasionally
  - Good atmosphere, effective, agreement generally comes easily. No problems so far.
  - Minutes public after internal circulation
  - Meetings 1-2/month
- Applications area meeting
  - 25-50 attendees local and remote
  - Project status, release news, activities of interest (internal and external), software usage and feedback
  - Meetings ~3/month
- Many meetings at project and work package level
  - We promote having them in the afternoon with a phone connection

Focus on Experiment Need

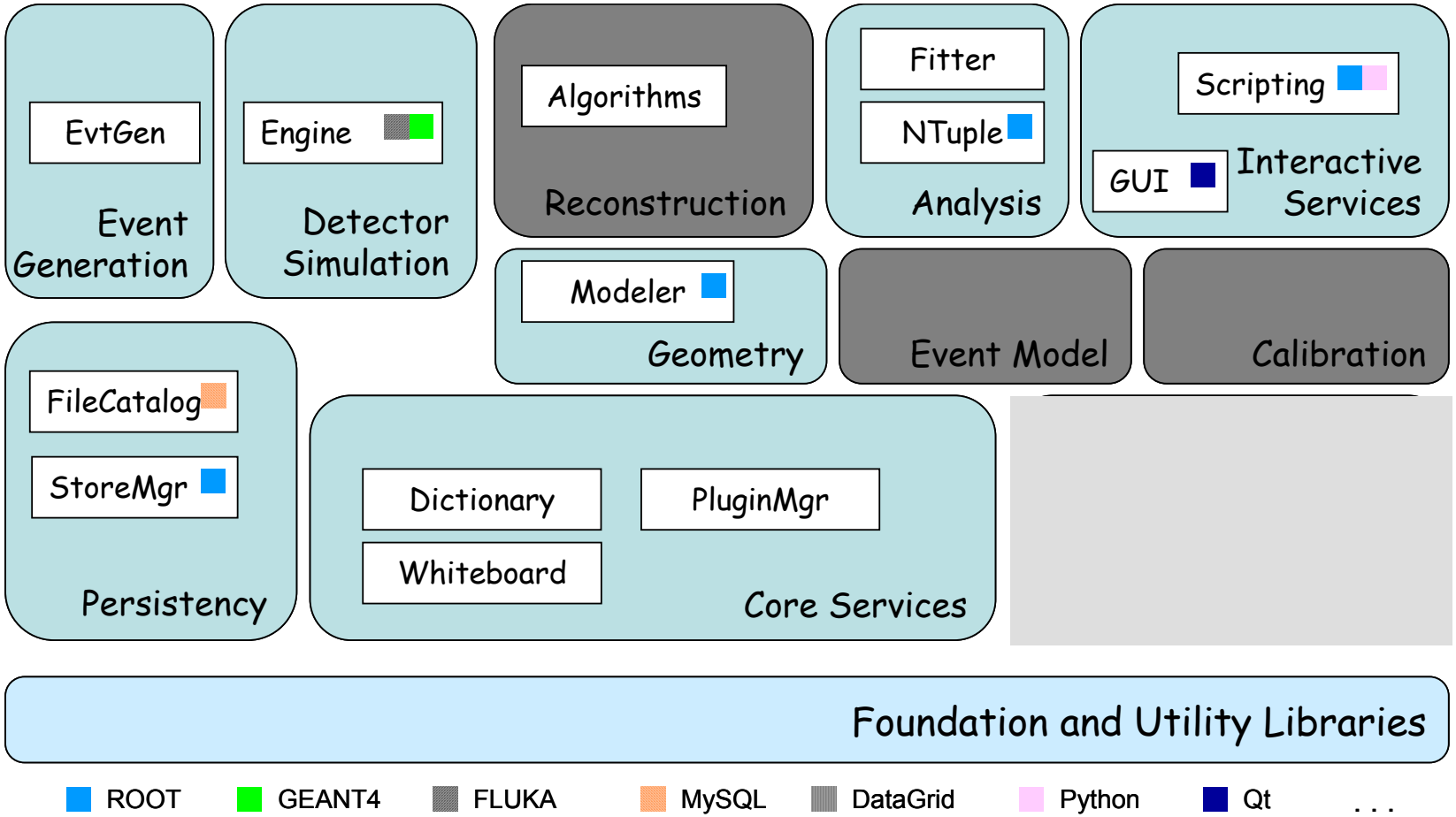
- Project structured and managed to ensure a focus on real experiment needs
  - ✓ **SC2/RTAG** process to identify, define (need-driven requirements), initiate and monitor common project activities in a way guided by the experiments themselves
  - ✓ **Architects Forum** to involve experiment architects in day to day project management and execution
  - ✓ **Open** information flow and decision making
  - ✓ **Direct participation** of experiment developers in the projects
  - ✓ Tight **iterative feedback** loop to gather user feedback from frequent releases
  - **Early deployment and evaluation** of LCG software in experiment contexts
  - **Success defined by experiment adoption** and production deployment

Substantive evaluation and feedback from experiment integration/validation efforts now in progress

Apps Area Projects and their Relationships



Applications Domain Decomposition



Products mentioned are examples; not a comprehensive list

Project activity in all expected areas except grid services

Implementing the Architecture Blueprint

- **Use what exists:** almost all work leverages existing software
  - ROOT, Gaudi/Athena components, Iguana components, CLHEP, Aida, HepUtilities, SCRAM, Oval, NICOS, Savannah, Boost, MySQL, GSL, Minuit, gcc-xml, RLS, …
- **Component-ware:** followed, and working well
  - the basis of SEAL, POOL, PI development as components of a coherent overall architecture
  - SEAL components successfully integrated into POOL, PI
- **Object dictionary:** In place
  - Meeting POOL needs for persistency
  - Application now expanding to interactivity
  - ROOT and LCG agree on convergence on common dictionary; should see activity in this over the next year
- **Object whiteboard:** Still to come
  - Serious design discussions underway

Implementing the Architecture Blueprint (2)

- **Component bus/scripting environment:** both Python environment and its integration with ROOT/CINT progressing well
- **ROOT**
  - The 'user/provider' relationship is working
  - Good ROOT/POOL cooperation – POOL gets needed modifications, ROOT gets debugging/development input
  - ROOT will be the principal analysis tool; full access to its capability is a blueprint requirement and will be provided
    - Directly and via interfaces such as PyROOT, Aida_ROOT
    - Work still needed to complete uniform access to functionality from ROOT (CINT) as well as LCG (Python)
  - ALICE directly using ROOT as their framework

Applications Area Project Planning

- Applications area overall plan (Version 2 – August 2003)
- Project plans for the various projects
- WBS, schedules
  - WBS follows the project/subproject/work package structure
  - Schedule defines milestones and deliverables, monitors performance
    - Three levels:
      - Level 1 (LHCC reporting – 4 total)
      - Level 2 (PEB/SC2 reporting – ~2/quarter/project)
      - Level 3 (applications area internal tracking)
    - Experiment integration/validation milestones monitor take-up
- Personnel resources spreadsheets
- Future resource needs estimations/planning (in development)
- Risk analysis

Level 1 (LHCC) Milestones

Date	Description	WBS	ID#
2003/6/15 Done v=15	**General release of POOL hybrid data store**	lcg:1.1.2	1.2

Keys: ms major lcg1 lcg2 lcg3 lcg app pool WBS: lcg:/lcg/app/data

Done! Time variance 15 days

The first public, production-capable release of the persistency framework. This will be a release offering basic hybrid persistency services, documented and packaged using the SPI-defined templates and tools, for general use by the experiments in production environments.

The release should support production usages with 1M file counts in the distributed catalog, 20k file counts in local XML catalogs, 50TB data volumes, 300 concurrent clients for DB served catalogs, access time for insert/lookup better than 30ms, and distributed operation at 10 sites.

Previous releases were internal releases targeted at developers and experts; this is the first public release with the robustness, documentation, packaging and support requirements inherent in a public release. Specific feature set for this release defined by a milestone four months earlier.

Distributed analysis Milestones will be (re)introduced Based on ARDA RTAG outcomes

2005/3/1	**Full function release of persistency framework**	lcg:1.1.2	1.7

Keys: ms major lcg1 lcg2 lcg3 lcg app pool WBS: lcg:/lcg/app/data

Completion of the fully functional POOL persistency framework.

Deliverables in terms of feature set, performance and scalability for this milestone will be finalized as an outcome of the Computing TDRs of the experiments in 2004.

Level 2 Milestones

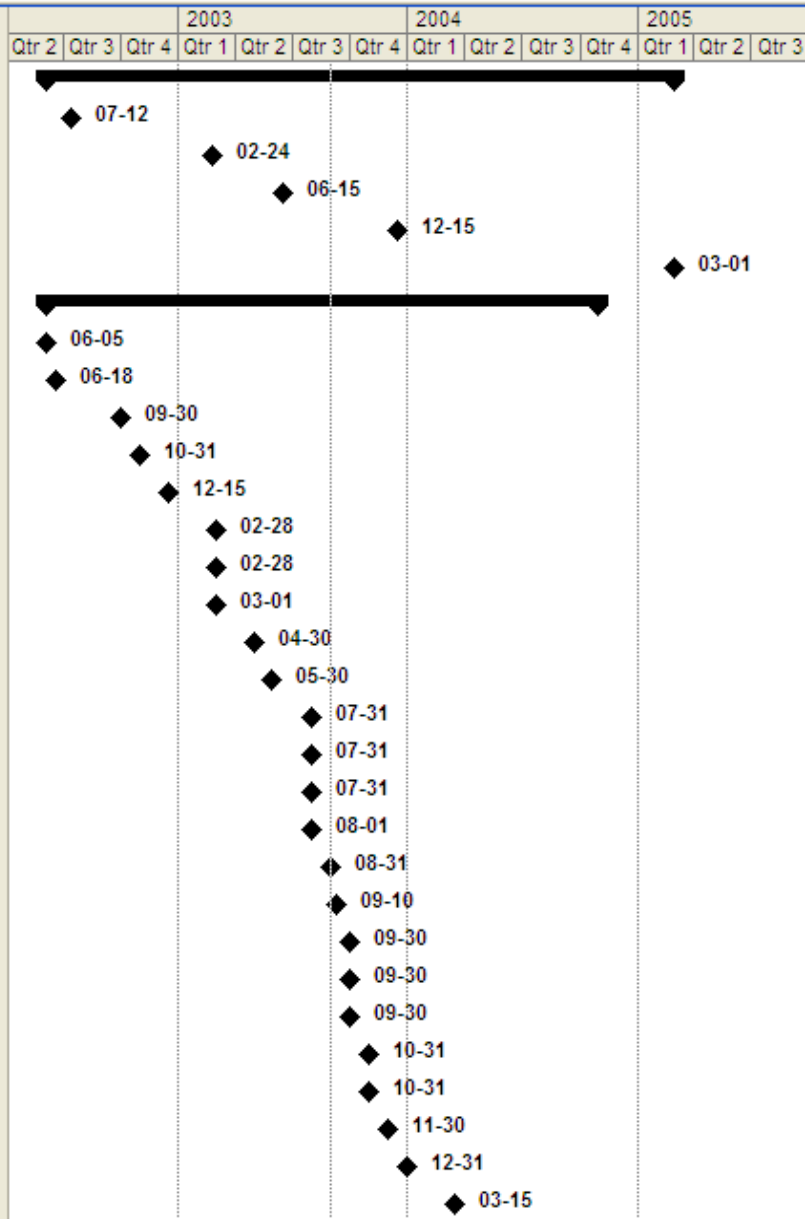
For expired items: Green = done, Red = not done. For completed milestones v=variance in days.
Milestones in **bold are level 1**. Milestones with IDs of the form 1.nnn are level 2. Level 3 milestones have IDs > 10000.

Date		Description	WBS	ID#
2003/6/15	**Done v=15**	**General release of POOL hybrid data store**	lcg:1.1.2	**1.2**
2003/6/30	Done v=10	Nightly builds deployed in SEAL	lcg:1.1.3	1.128
2003/6/30	Done v=18	SEAL V1 release	lcg:1.1.3	1.130
2003/6/30	Done v=0	Generator librarian and alpha version of support infrastructure in place	lcg:1.1.5.6	1.138
2003/7/1	Done v=3	Physicist interface (PI) workplan completed	lcg:1.1.4	1.153
2003/7/31	Done v=0	CMS POOL integration: POOL persistency of CMS event	lcg:1.1.2.1	1.176
2003/7/31	Late	Math library workplan in place	lcg:1.1.3	1.184
2003/8/15	Late	SPI support for Windows binary version of LCG software	lcg:1.1.1	1.170
2003/9/10	Done v=1	ATLAS POOL integration: POOL persistency in Release 7	lcg:1.1.2.1	1.177
2003/9/15	Done v=24	SEAL support for Windows binaries	lcg:1.1.3	1.187
2003/9/15	Done v=24	AIDA interface review (users) completed	lcg:1.1.4	1.171
2003/9/30	Late	POOL RDBMS independence layer in beta	lcg:1.1.2.1	1.116
2003/9/30	Late	POOL support for Windows binaries	lcg:1.1.2.1	1.181
2003/9/30	Done v=0	First cycle of EM physics validation complete	lcg:1.1.5.4	1.143
2003/9/30	Late	Statement on GSL and NAG usage for math library	lcg:1.1.3.8	1.124
2003/10/31		ATLAS POOL validation with DC1 data	lcg:1.1.2.1	1.179
2003/10/31		CMS POOL validation with PCP data	lcg:1.1.2.1	1.180
2003/11/1		POOL implementation of conditions DB - beta	lcg:1.1.2.2	1.115
2003/11/15		Initial POOL deployment on LCG-1	lcg:1.1.2.1	1.114
2003/11/15		SPI-G4 collaborative infrastructure pilot	lcg:1.1.5.6	1.189
2003/11/30		SPI tools operational on IT CVS service	lcg:1.1.1	1.188
2003/12/15		2004-2005 persistency framework workplan complete	lcg:1.1.2	1.117
2003/12/31		Simulation physics requirements revisited	lcg:1.1.5.4	1.140
2003/12/31		Generic simulation framework prototype available (G4 and FLUKA)	lcg:1.1.5.1	1.144
2004/1/31		First cycle of hadronic physics validation complete	lcg:1.1.5.4	1.145

MS Project Integration – POOL Milestones

Gantt Chart

	Task Name	Duration	Start
47	**Physics data management**	**714 days**	**Wed 02-06-05**
48	Persistency framework workplan to SC2	0 days	Fri 02-07-12
49	2003 persistency framework workplan complete	0 days	Mon 03-02-24
50	General release of POOL hybrid data store	0 days	Sun 03-06-15
51	2004-2005 persistency framework workplan complete	0 days	Mon 03-12-15
52	Full function release of persistency framework	0 days	Tue 05-03-01
53	**POOL Persistency Framework**	**628 days**	**Wed 02-06-05**
54	Hybrid event store workshop	0 days	Wed 02-06-05
55	Persistency framework status report	0 days	Tue 02-06-18
56	V0.1 POOL internal release	0 days	Mon 02-09-30
57	V0.2 POOL internal release	0 days	Thu 02-10-31
58	First prototype (V0.3) release of POOL hybrid data	0 days	Sun 02-12-15
59	SPI compliance of POOL	0 days	Fri 03-02-28
60	POOL V0.4 'interface-complete' release	0 days	Fri 03-02-28
61	Define feature set of first general release of POOL	0 days	Sat 03-03-01
62	POOL V1.0 pre-production release	0 days	Wed 03-04-30
63	Nightly builds deployed in POOL	0 days	Fri 03-05-30
64	Initial POOL deployment on LCG-1	0 days	Thu 03-07-31
65	CMS POOL integration: POOL persistency of CMS	0 days	Thu 03-07-31
66	CMS POOL acceptance for PCP	0 days	Thu 03-07-31
67	POOL V1.2 release	0 days	Fri 03-08-01
68	POOL icc test build support	0 days	Sun 03-08-31
69	ATLAS POOL integration: POOL persistency in Rel	0 days	Wed 03-09-10
70	POOL RDBMS independence layer in beta	0 days	Tue 03-09-30
71	POOL ecc test build support	0 days	Tue 03-09-30
72	POOL support for Windows binaries	0 days	Tue 03-09-30
73	ATLAS POOL validation with DC1 data	0 days	Fri 03-10-31
74	CMS POOL validation with PCP data	0 days	Fri 03-10-31
75	POOL hierarchical cataloging alpha release	0 days	Sun 03-11-30
76	POOL fully distributed	0 days	Wed 03-12-31
77	POOL hierarchical cataloging production release	0 days	Mon 04-03-15



Applications Area Risk Analysis

Full analysis is at http://lcgapp.cern.ch/project/mgmt/apps-risks.html and includes for each risk:
- likelihood/impact/overall risk assessment, current process for managing risk,
future options for managing risk, crisis strategy, actions required

Key:

Likelihood: 1 - never expected to happen 2 - could happen but very unlikely
3 - could well happen at some point 4 - will probably happen

Impact: 1 - we can deal with it, no problem 2 - a bit of a hassle but not too bad
3 - clearly can be dealt with, but with significant effort 4 - crisis

Overall risk: 1 - very unlikely to have significant adverse impact 2 - unlikely to have...
3 - could have significant adverse impact 4 - would be a crisis and may happen

	Risk	Likelihood	Impact	Overall risk
1	Inadequate third party software	3	2	1
2	Third party software no longer available	3	2	1
3	Software obsolescence	4	2	1
4	Deployment failure	1	3	1
5	Platform incompatibility	1	2	1
6	Experiment discord	2	3	2
7	Experiment departure	2	3	1
8	Integration failure	1	2	1
9	Requirements/functionality mismatch	2	2	1
10	Product take-up failure	2	2	2
11	Licensing limitations	3	4	4
12	Loss of ROOT team	2	3	2
13	Missing personnel	3	3	3
14	Staffing continuity failure	3	3	3
15	Software delivery failure by work teams	4	3	2
16	Duplication of work	4	1	1
17	Grid middleware/infrastructure failure	3	3	3
18	POOL data storage performance failure	1	3	1
19	POOL catalog management performance failure	2	3	2
20	Overlooking a risk	4	2	1

Software Process and Infrastructure (SPI)

Alberto Aimar, CERN IT

Foster software quality "as least as good as and preferably better than that of any experiment"

- Full suite of tools and services in place, supporting developers and users
  - Adopt/adapt from experiments, community – minimize in-house development
- Testing frameworks for unit, integration and regression testing
- Software library offering LCG and third party software
- Software distribution/remote installation service recently deployed
- Strong QA activity: POOL and SEAL QA reports recently released
- Savannah development portal a great success with 70 projects, 395 users recently
- Policies on code standards and organization, testing, documentation in place
- Training program: ROOT; SCRAM; POOL and SEAL coming soon
- Additional platform porting underway: Windows; icc (new Intel); ecc (64-bit)
  - Not without difficulties!
- Manpower level (just) enough to support the essentials
  - 1-2 more would improve support and reduce 'firefighting mode'
  - Team made up of LCG, IT, EP, experiment participants
  - New LCG applications area arrivals contribute to SPI (a policy)

Core Libraries and Services (SEAL)

Pere Mato, CERN EP/SFT

- Provides foundation and utility libraries and tools, basic framework services, object dictionary, component infrastructure
  - Facilitate LCG software coherence and integration with non-LCG software
- Leverages experiment (e.g. Gaudi, Iguana) and community (e.g. Boost) sw
- Delivering, on schedule
  - Has successfully delivered POOL's needs, the top priority
  - Blueprint-driven component model and basic framework services, directed also at experiments, recently released
  - Also provides components important for analysis (e.g. PyROOT)
  - CLHEP now hosted by SEAL – also reflects well on SPI services
  - Doesn't always come easily: "Combining existing designs into a 'common' one is not trivial"
- Current focus areas: integration, new platforms (Windows, icc, ecc), component model feedback, object whiteboard
- Manpower is OK – team consists of LCG and experiment people

Persistency Framework Project

Dirk Duellmann, CERN IT/DB

- To deliver
  - the physics data store (POOL) for ATLAS, CMS, LHCb
  - The common conditions DB software (a new work package)
- Production POOL release delivered on schedule in June – Level 1 milestone
  - Contains all functionality requested by the experiments for initial production usage in their data challenges
  - Leverages proven technologies: ROOT I/O, relational databases
  - Makes use of SEAL core components and SPI infrastructure
  - Provides stably supported (~18 mo) data format as requested
- Focus is now on supporting experiment integration (ATLAS, CMS, LHCb)
  - Responding to feedback, debugging, integration assistance, performance, documentation, process and infrastructure
  - Several such releases since the June production release
  - POOL team member assigned to each experiment to assist/liaison; examining how to further support integration
  - Successful take-up by the experiments will define POOL success
    - Going well so far with milestones met, though it took more time and effort than ATLAS, CMS expected
    - CMS writing ~1M POOL events/week: TB data volumes this fall
- Project manpower is OK
  - IT/DB, LCG, and experiments are all contributing substantial and vital manpower

Physicist Interface (PI) Project

Vincenzo Innocente, CERN EP/SFT

- Interfaces and tools by which a physicist (particularly a physicist doing analysis) will directly use the software
  - Interactivity (the "physicist's desktop"), analysis tools, visualization, distributed analysis, 'grid portals'
- Currently a small effort (<2 FTEs) focused on the limited scope opened so far
- Analysis Services – most of the activity so far
  - **Simplified, implementation-independent AIDA interface to histograms, tuples developed released for evaluation (little user feedback so far)**
  - **ROOT implementation of AIDA histograms completed**
  - Integration of analysis services with POOL, SEAL
- Analysis Environment
  - Interactivity and bridge to/from ROOT – joint work with SEAL
    - pyROOT Python interface to ROOT
  - Interoperability via abstract interfaces to fitters and other analysis components
  - End-user interfaces for tuples/event collections
- Future planning depends on what comes out of the ARDA RTAG

Simulation Project

Torre Wenaus et al

- **Generic simulation framework**
  - Generic interface to multiple simulation engines (G4, FLUKA), building on existing ALICE work (VMC)

Andrea Dell'Acqua

- Incorporates longstanding CERN/LHC **Geant4** work
  - Aligned with and responding to needs from LHC experiments, physics validation, generic framework

John Apostolakis

- **FLUKA** team participating
  - Framework integration, physics validation

Alfredo Ferrari

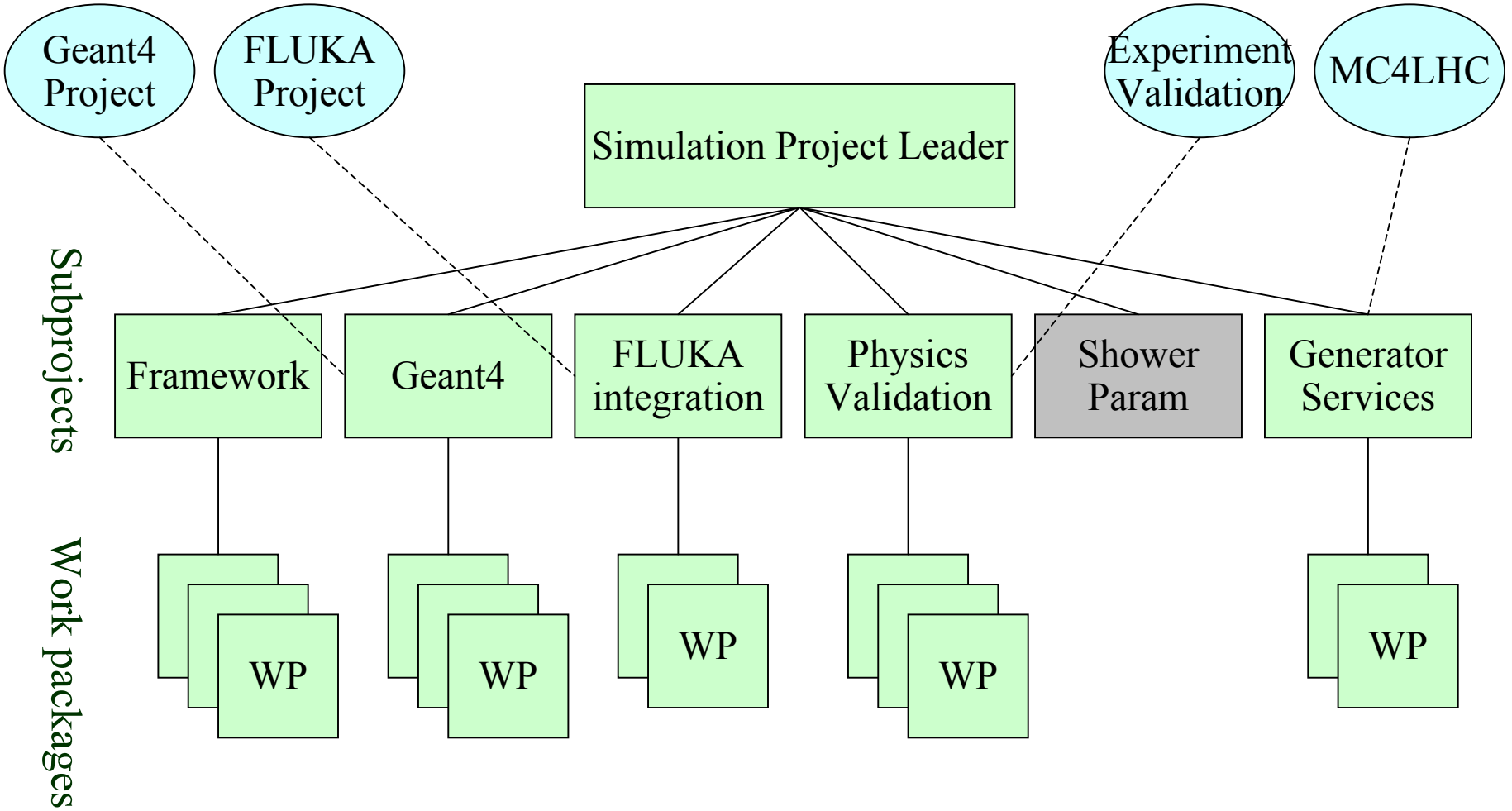
- Simulation **physics validation** subproject very active
  - Physics requirements; hadronic, em physics validation of G4, FLUKA; framework validation; monitoring non-LHC activity

Fabiola Gianotti

- **Generator services** subproject also very active
  - Generator librarian; common event files; validation/test suite; development when needed (HEPMC, etc.)

Paolo Bartalini

Project Organization



Simulation Project Major Milestones

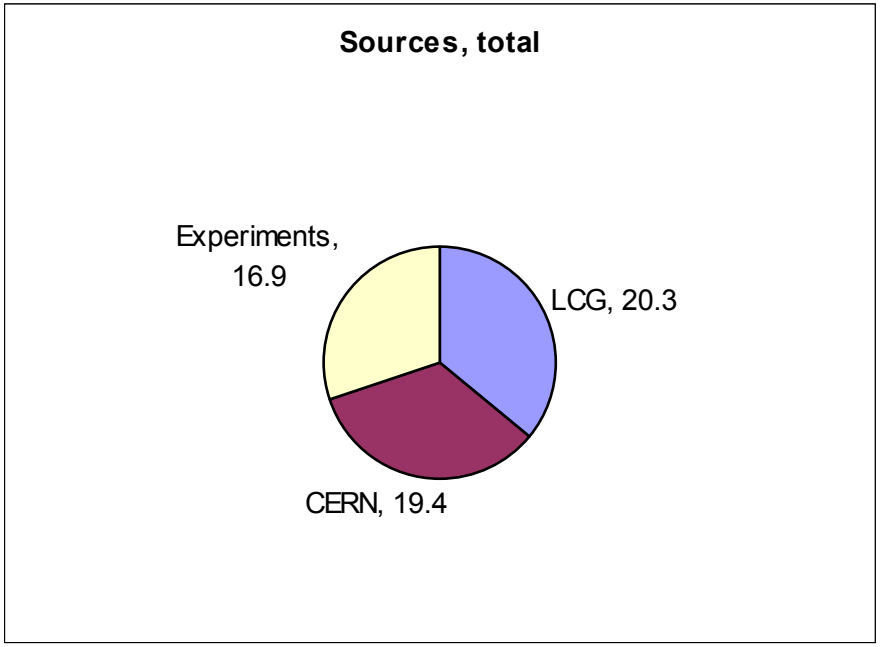
- 2003/6: **Generator librarian** and first library version in place
  - Completed
- 2003/7: Simulation **physics requirements** revisited
  - Completed for ATLAS; will be complete for all experiments by year-end
- 2003/9: 1st cycle of **EM physics** validation complete
  - Essentially complete; will continue to evaluate new data
- 2003/12: Generic **framework prototype** with G4, FLUKA engines
- 2004/1: 1st cycle of **hadronic physics** validation complete
- 2004/3: Simulation **test and benchmark suite** available
- 2004/9: First generic **simulation framework production** release
- 2004/12: Final **physics validation document** complete

Integration and Experiment Support

- A common thread: the experiments are finding their schedules and their manpower to be very stretched by LCG software integration
- "Release early and often" only works when complemented by "Integrate early and often"
- Prompt and successful integration in experiments, leading to early feedback, is essential to project success
- We are helping the experiments now, but we must help them even more
  - The need is visible in all development projects
- Will be exploring a tighter association between project and experiments at the developer level by assigning presently 'unaligned' developers with a particular experiment, to enhance liaison and integration support

Applications Area Personnel Resources

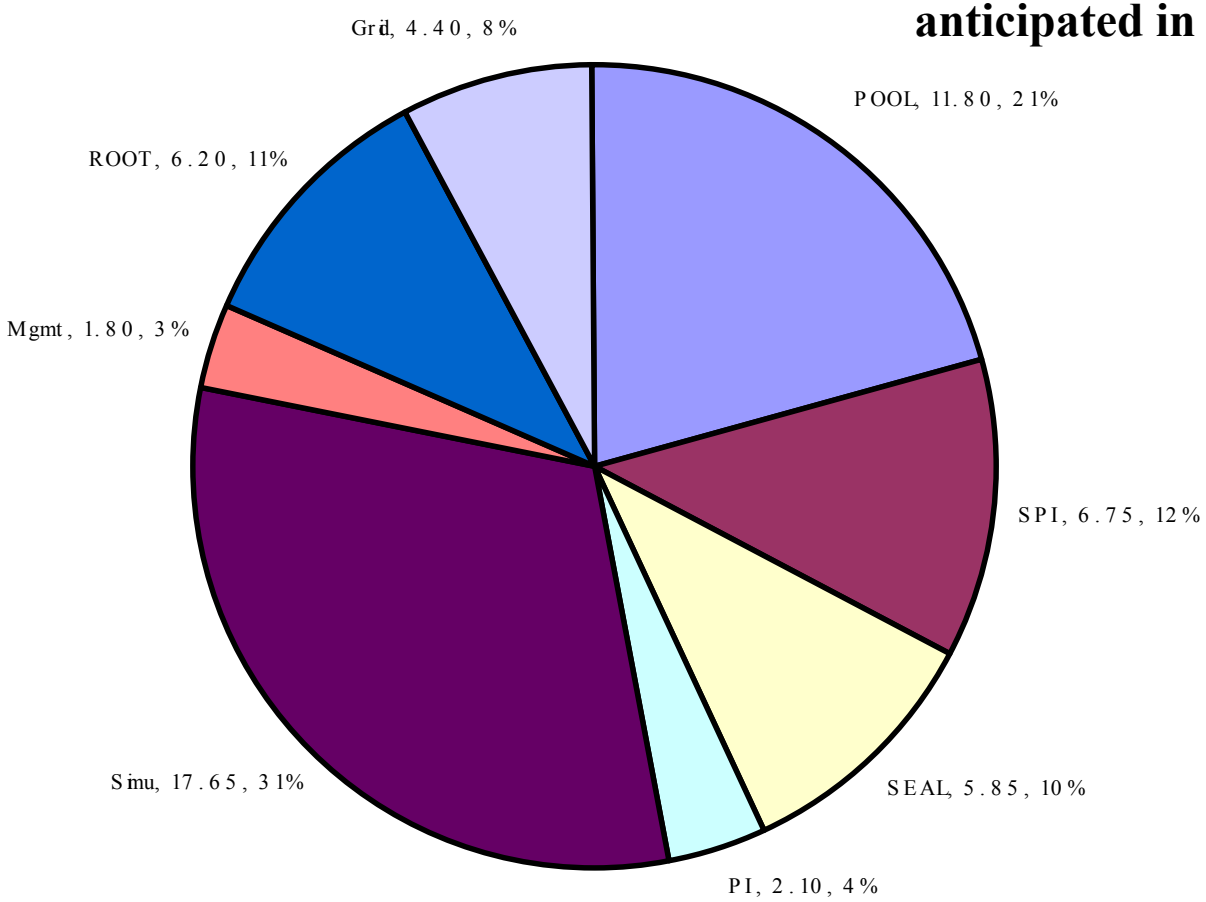
- LCG applications area hires essentially complete
  - 21 working; target in Sep 2001 LCG proposal was 23
  - Contributions from UK, Spain, Switzerland, Germany, Sweden, Israel, Portugal, US, India, and Russia
- Similar contribution levels from CERN, experiments



Experiment number includes CERN people working on experiments

Personnel Distribution

Effort levels match need anticipated in blueprint RTAG



Non-CERN Participation

- Not always easy to engage non-CERN participation but it is vital to the project and we try to foster and support it
- Examples:
  - POOL collections (US)
  - POOL RDBMS data storage back end (India)
  - POOL tests (UK)
  - POOL-driven ROOT I/O development & debugging (US)
  - SEAL scripting tools (US)
  - Generator services (Russia)
  - SPI tools (France, US)
  - Math libraries (India)

Concluding Remarks

- POOL, SEAL, and PI software is out there
  - Level 1 POOL milestone met, take-up in progress
    - Take-up the real measure of success, and signs are good so far
    - First round of POOL-SEAL validation milestones met
    - But, integration took more effort than expected, in manpower-starved experiments
- Project giving major attention to integration, but more attention still needed if it delays experiment take-up
  - "Release early and often" has to be complemented by "integrate early and often"
- Collaboration on joint projects doesn't always come easily: "Combining existing designs into a 'common' one is not trivial"
  - But it is being done, and is delivering
- Manpower is appropriate at the moment
  - Delivering what we are mandated to deliver, with the manpower that was estimated to be required
- **It is time for a technical evaluation of what we are doing – hence we are all here, and thanks very much to reviewers for giving their time**

Supplementary…

Personnel

	People	FTEs
LCG applications area personnel	21	20.25
Working directly for apps area projects	13	12.85
ROOT	2	2
Grid integration work with experiments	3	2.8
Distributed analysis	3	2.6
Contributions from		
IT	4	3.30
EP/SFT not experiment specific	21	16.10
EP/SFT experiment specific	7	4.35
Experiments outside EP/SFT	29	12.55
Total - direct project contributions	52	30.50
Total - indirect contributions (ROOT, ALICE VMC)	9	5.80
Total directly working on apps area projects	65	43.35
Overall total	82	56.55