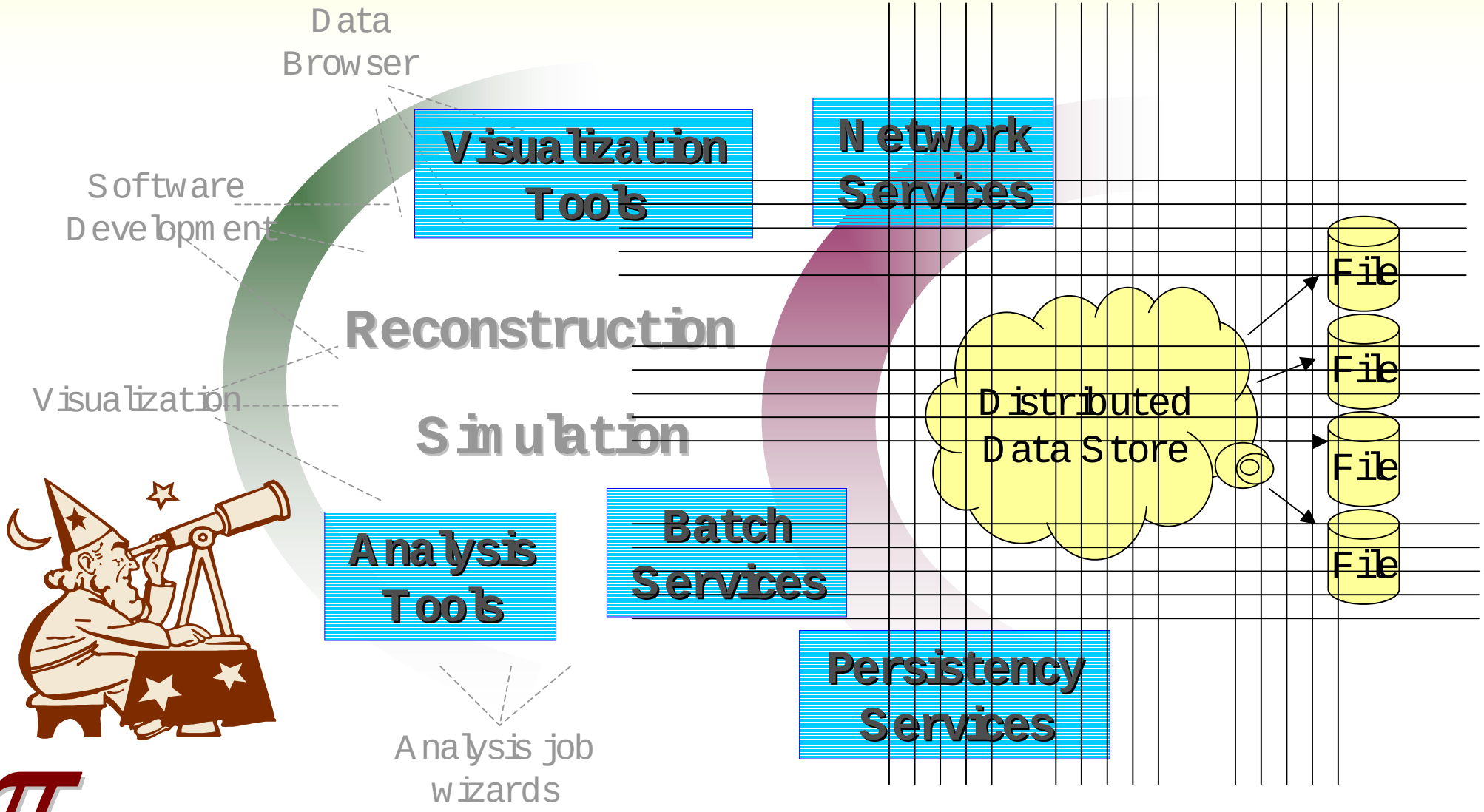
# Physicist Interface: Project Proposal & Status

## AAM, 12 March 2003

**Vincenzo Innocente**

CERN/EP

# Coherent Analysis Environment
## (chep01)



# “Blueprint” PI-Breakdown

- **Interactivity**
  - Physicist’s desktop
- **Analysis Tools**
  - Binned and unbinned statistical analysis
    - Data manipulation and transformation
    - Analysis (fitting) engine
    - Inspection of results
- **Visualization**
  - Data and results of statistical analysis
  - Event simulation, reconstruction
- **Distributed analysis**
  - Transparent access to “grid” resources
    - “bsub” on the wan…
    - “root” on the wan…
  - Tools to implement experiment’s analysis workflow
- **Grid Portals**
  - Transparent access to “grid” resources
  - Custom solution tailored to LHC experiment needs
    - Event collection handing
    - Metadata browsing
    - VDP definition
    - …

# Survey (dec 2003)

- **Meeting with each experiment**
  - Blueprint contributors & readers (from computing and physics groups)
- **Meeting with partners**
  - EDG, US-Grid, LCG/GDA, Pool, Seal, Geant4, Anaphe, Root

- **Current practice and experience**
- **Short and medium term plans**
- **Requirements, constrains, expectations**
- **Possible common products and projects**
- **Specific Comments on:**
  - Analysis Model, Computing Model
  - Root, Aida, Python, Event Visualization
  - Grid middleware, experiment workflow management

# ALICE

- **Simple, consistent and coherent model:**
  - One application: AliRoot (Alice extension of Root)
  - Two independent base components: Root and Alien
    - All external resources are interfaced to them
  - Identical batch and interactive environment
  - Already deployed to physicists: 2003 just consolidation work
  - Full 3D interactive event display still missing
    - Non critical: one-way Root to OI gateway exists

- **Risks** *(in adopting this model for PI)*
  - Physicists interact only with Root and Alien: does it scale?
  - Rely heavily on the ability of Root and Alien to respond to
    - User requirements
    - Changing environment
    - New tools and computing paradigms

# LHCb

- **Clear coherent model**
  - One analysis application (DaVinci) based on Gaudi
    - Physicists expected to “code” in DaVinci using Gaudi paradigm
  - Gaudi Tag-collections with paw-ntuple interface and root-tree (not only) impl.
  - Gaudi Histogram-service (AIDA based) with native Root implementation
  - Python interactive environment with gateway to Root
  - Ganga as grid physicists interface (not used for production in the short term)
  - 3D event display (Panoramix) is a Gaudi extension based on OnX and HepVis
  - DaVinci deployed in batch mode, Panoramix deployed
  - Interactive environment and Ganga expected to be deployed early 2003
- **Risks**
  - Some scenarios seems to traverse too many independent components
  - Unclear role of Root in the interactive environment (just the best canvas around?)

# CMS

- **Striking similarities with LHCb approach (fewer layers)**
  - COBRA ⇦⇨ GAUDI
  - IGUANA ⇦⇨ OnX+HepVis
  - ORCA ⇦⇨ DaVinci
  - Very similar python environment
- **Major effort to integrate grid middleware**
  - Focus on production: physicist interface seen as its natural extension
  - Request for means of controlling resources
- **Analysis environment**
  - Root used (as-is, not integrated) but not satisfactory
  - Interactive version of ORCA is a requirement, mixed feeling on user interface
- **Risks**
  - High integration with middleware, will migration be always efficient?
  - What's the future of a stand-alone Root environment?

# ATLAS

- **Already Collaborating with LHCb on**
  - GAUDI
  - GANGA
- **Still in a transition phase**
  - Many past projects in the area of physics analysis abandoned or dormant
  - Emphasis on Abstract-Interfaces and implementation-independent descriptions
  - Physicists not yet really exposed to new framework and tools:
    - use Root, no request for a new analysis environment
- **Major involvement in Grid projects in Europe and US**
  - Lot of projects and products of potential interest (Grappa, Magda, …)
- **Risks**
  - Emphasis on being “Independent of everything” may tend to produce duplicate middleware and to slow down application code development
  - If Root is used as analysis environment, how analysis migrates back into “offline”?

# Summary of Survey

- **Consensus on the "BluePrint Architecture"**
- **Experiments have not finalized their physics and computing model**
  - Difficult to workout concrete use-cases and scenarios
  - No reason to have a "physics-analysis RTAG" prior to to this
- **All experiments have developed distributed production workflow managers**
  - Integrated with monitoring, book-keeping, error-recovery, user-interfaces
  - Often not matching with current architectural view of international Grid projects
  - Worries about robustness, readiness and stability of grid middle-ware
- **Role of Root as analysis tool confirmed: CMS, LHCb and in part ATLAS expressed:**
  - Root does not play a central role in their analysis software
  - Difference between root paradigm and experiment's computing paradigm
  - Difficulties in integrating it with their own software

# Working Items for PI

- **Abstract interface to analysis services**
  - GEANT4 and some experiments do not wish to depend on a specific implementation
  - One implementation must be Root
  - Request for a coherent LCG analysis tool-set
- **Interactive analysis environment**
  - Access to experiment objects (event-data, algorithms etc)
  - Access to high level Pool services (collections, metaData)
  - Transparent use of LCG and grid services
    - "escape" to debug and detail monitoring mode required
  - GUI (point&click) and scripting interface
- **Event and detector visualization**
  - Fully interactive (both "view" and "model")
  - Integrated with the analysis environment
  - Offering a large palette of 2D and 3D rendering

# WP1: Analysis Services

- **WP 1.1 - AIDA**
  - Review, Adjust, Extend
  - **Work started: first release March 19, AAM presentation April 9**
- **WP 1.2 - Root implementation of AIDA**
  - In part just of “typewriting”-work
  - requires design work in some area where object-relationship are involved
- **WP 1.3 – “AIDA” interface to Seal and Pool services**
  - whiteboard
  - collections
- **WP 1.4 - BluePrint compliant Analysis tool set**
  - ***Put on hold by SC2***
  - Join effort with Math project?

**Effort: 2.5 FTE**

# WP2: Analysis Environment

- **WP 2.1 - Basic Interactive LCG application**
  - Depends on SEAL
    - plug-in manager, dictionary, python binding, distributed applications
  - May start from Iguana/OnX
- **WP 2.2 - Core Visualization services**
  - 2D & 3D Canvas, basic primitives, basic operations
  - Model-View architecture (picking, etc)
  - Export to viewers and printers
- **WP 2.3 - Bridge to Root**
  - To use root as canvas and analysis engine
- **WP 2.4 - Bridge from Root**
  - To use PI Analysis Services from Root (Cint prompt)

**Effort: 2.5 FTE**

# WP3: Pool & Grid PI

- **WP 3.1 - Collection manipulation**
  - including full “MetaData” handling
- **WP 3.2 - Interface to low-level services**
  - catalog, replicas (local, grid)
- **WP 3.3 - Job wizard**
  - preparation, submission, validation
- **WP 3.4 - Job helpers**
  - monitoring, error-recovery, resource discovery

**Effort: 6 FTE**

**SC2 decided to start a RTAG on this item**

**May well go outside PI scope**

# WP4: Event and Detector Visualization

- **WP 4.1 - HepVis**
  - Review, adjust, extend to cover LHC & Geant4 needs
- **WP 4.2 - "Extend" WP2 to provide event and detector visualization and interactivity**
  - Start from Iguana/OrcaVis & OnX/Panoramix
- **WP 4.3 - Geant4 visualization application**
  - Out of the box application configured for Geant4

**Effort: 2 FTE**

**Put on Hold by SC2 till May**

# WP5: Infrastructure & Documentation

- **WP 5.1 - Liaison with SPI**
  - External Software
  - Savannah Portal
  - CVS repository
    - CVS view
  - Release area
    - Doxigen, LXR
  - “Daily” build
    - Doxigen, LXR
    - Web area
- **WP 5.2 - Liaison with Experiments**
- **WP 5.3 – Documentation**

**Effort: 1 FTE**

# Organization

- **For each work package**
  - Organize an inception workshop
    - Identify direct contributors (people, software)
    - Identify partners (other projects)
    - Identify users
    - Finalize work package breakdown
  - Get commitment from developers, experiments, institutions
    - As *producers* and as *users*
  - Produce timescale that fits experiments' schedule
  - Assign LCG person-power
  - First deployment of products after **Three months** provided that:
    - The basic infrastructures and services (grid, pool, seal, spi) are ready
    - The developers of the present products commit 100% to the project
      - Most of the code already exists!

# Timescale

- **Start work on Analysis Services and Environment now**
  - Work on Analysis Services started
  - Experiments contacted about experts on interactive environment: not a single answer yet
- **Collaborate with Pool and Seal to define and implement proper interfaces to their services and components**
- **Review with SC2 strategy how to keek-off “Interactive access and visualization of Events” in May**
- **Wait that EGEE and ITR proposals are out before planning grid related work**
  - Close collaboration with ITR and EGEE proponents established
  - SC2 will start a RTAG soon

# Collaboration

- **PI will re-use, reengineer and finally integrate in a coherent software-base a large, heterogeneous collection of products mostly already currently available**
- **In the short/medium time-scale many of these products, even those HEP specific, may stay "external" to LCG. Mechanisms should be put in place to guarantee:**
  - LCG architecture is respected (the exception cannot become the rule)
  - LHC experiments' requirements and priorities are taken into account
  - Eventual migration into LCG managed software-base in the long-term
- **The success of this project will depend, even more than other LCG projects, on the commitment of all involved parties:**
  - Product providers and developers
  - Application developers and users
  - LCG developers and management

# Concluding Remarks

- **Definite interest in a common LHC Physicist Interface**
  - Areas identified by the Blueprint RTAG confirmed
  - Priorities and timescale need to be verified with experiments
    - Analysis and computing models still in development
    - “Challenges” schedule and content not finalized
- **Role of components (root, aida, qt, python, grid-mw) clarified**
  - Enough to start working
  - Should be revised as project progress and user-base enlarge
- **Major concerns about quality of  middle-ware (grid, root, etc…)**
  - Stability, robustness, data-integrity, readiness
- **Collaboration mechanisms with external projects needs to be clarified**