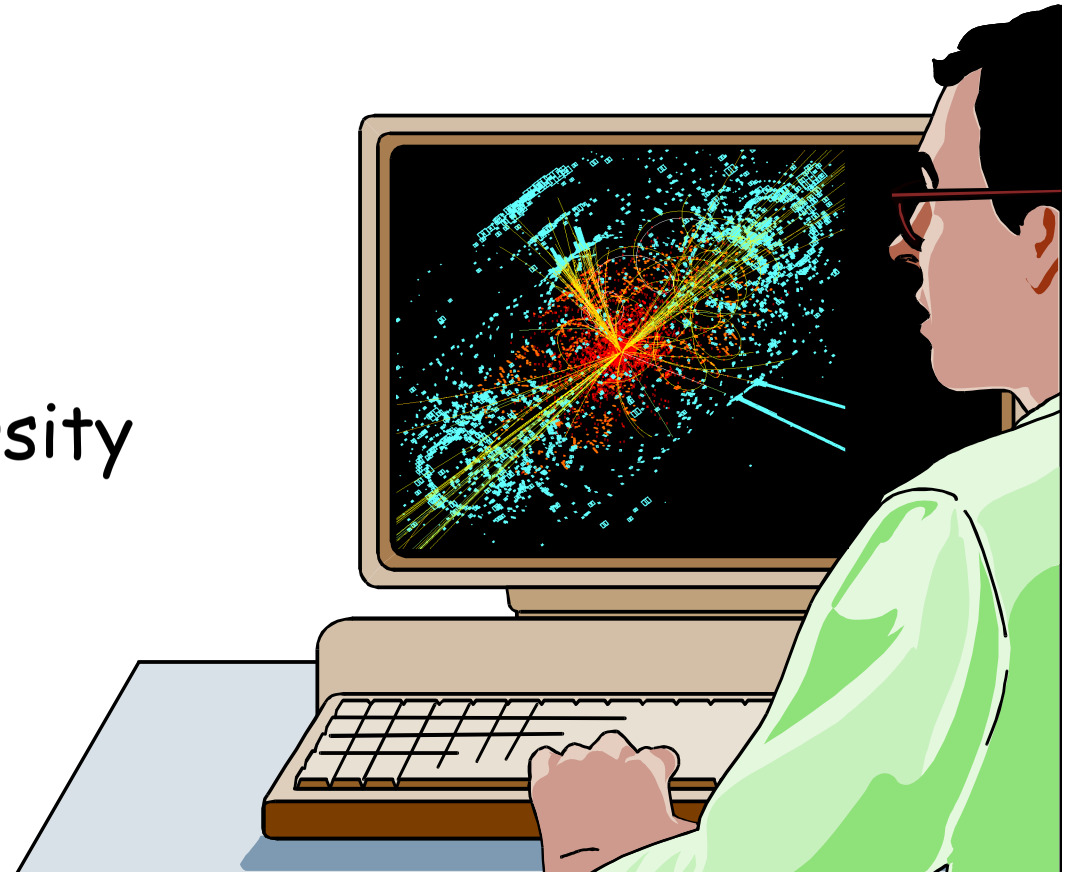
DOE / NSF Review of US CMS Software & Computing

Brookhaven National Laboratory, 14 November 2000

Core Applications Software

Lucas Taylor
Northeastern University



CMS Core Software Scope

- **Architecture of CMS software**
- **Software process and development environment**
- **Software framework, persistency services, utilities,…**
- **Tools for distributed data access / processing**
- **Analysis environment and toolkits**
- **etc…**

…requires **professional software engineering expertise**

- In other words, all that is required to support the closely related (mostly PRS) activities of:
  - event generation, detector and trigger simulation, reconstruction, data selection (online & offline), physics analysis, test-beams, etc….

...requires **professional physicist expertise**

CMS Core Software: Schedule

Functional prototype phase is now complete

CMS MILESTONES	1998	1999	2000	2001	2002	2003	2004	2005
CORE SOFTWARE								
End of Fortran development	Jun-98							
GEANT4 simulation of CMS	1 Jun-98	2	Dec-99	3 Jun-01		4	Dec-03	
Reconstruction/analysis framework	1 Jun-98	2	Dec-99	3	Dec-01	4	Dec-03	
Detector reconstruction	1	Dec-98 2	Dec-99		3	Jun-02	4	Jun-04
Physics object reconstruction		1 Mar-99	2	Jun-00	3	Dec-02	4	Dec-04
User analysis environment	1	Dec-98	2	Jun-00	3	Dec-02	4	Dec-04
DATABASE								
Use of ODBMS for test-beam	Dec-97							
Event storage/retrieval from ODBMS	1 Jun-98	2	Dec-99	3	Dec-01	4	Dec-03	
Data organisation/access strategy		Dec-98						
Filling ODBMS at 100 MB/s			Jun-99					
Simulation of data access patterns				Dec-00				
Integration of ODBMS and MSS				Dec-00				
Choice of vendor for ODBMS					Dec-01			
Installation of ODBMS and MSS						Dec-02		

May 2000: L2/L3 ×10 reduction

Nov 2000: L2/L3 ×100 reduction

Dec 2000: Trigger TDR

May 2001: TriDAS: discovery at 100Hz

End 2001: DAQ TDR

End 2002: Software & Computing TDR

~2003: Physics TDR

LHC/CMS turn on

20% Mock Data Challenge

5% Mock Data Challenge

US CMS Core Applications Software (CAS) Subproject

- 2 sub-projects of US-CMS Software&Computing Project
  - **User Facilities** (focussed on Tier 1 and Tier 2 centres)
  - **Core Applications Software** (the subject of this talk)

Two main tasks for "Core Applications Software"

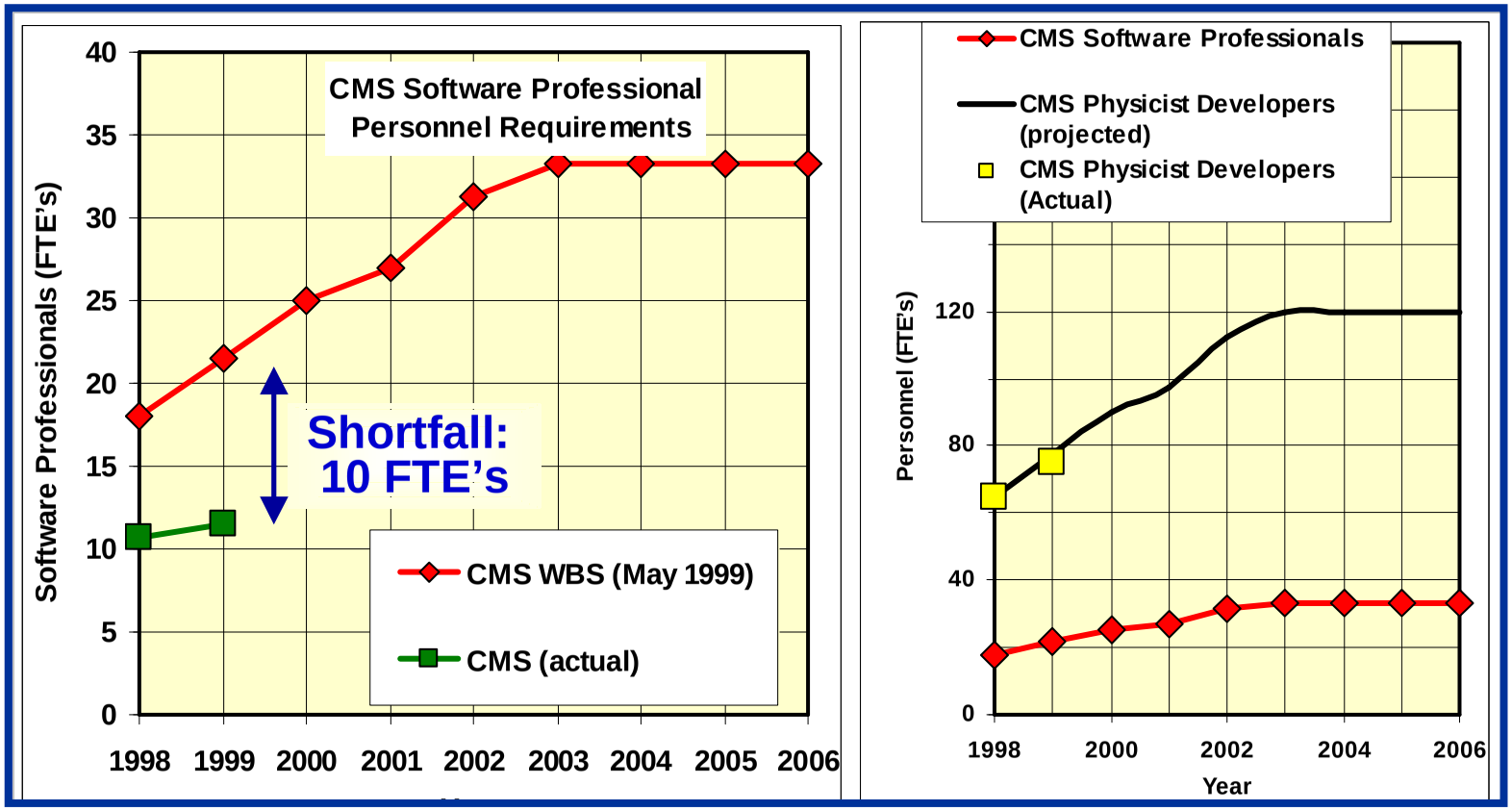
- **US contributions to CMS Core Software**
  - US responsibility to International CMS core software is **constrained to be a level-of-effort contribution of 25%**
  - US has responsibility & leadership for well-defined major tasks
- **Software support specifically for US-CMS physicists**
  - To enable them to meet their detector commitments
  - To enable them to fully exploit LHC physics
  - To enable them to do the above from their home institutes

Relationship of CAS to CMS project

- Formal links through US CMS L1 PM
- **Significant US leadership in CMS project**
  - CMS Software&Computing Deputy PM & ORCA PM: **D. Stickland**
  - CMS Software& Computing Board Chair **H. Newman**
  - CMS IGUANA Interactive Analysis PM: **L. Taylor**
- Tasks for US CMS chosen in **full consultation with CMS**
- **US-CMS WBS** is distinct from the **CMS WBS**
  - ✔ Clarity regarding US roles and responsibilities
  - ✔ Inclusion of US-specific items (notably local support)
  - ✘ Requires ongoing integration with International CMS planning…
  - ✔ …but members of US are intimately involved in this planning
  - US responsibility to International CMS core software is **constrained to be a level-of-effort contribution of 25 %**

International CMS Core Software WBS

- International CMS WBS goes all the way to 2006
  - but only at a reasonably high level (typically level 3 / level 4)
- Filter for **only the software engineers** (see plots)
  - **i.e. do not include physicists** (~ 1 engineer/ 4 physicists)



CAS Resource Planning

- **Software Engineers for US deliverables to CMS**
  - **International CMS** profile is determined from the **CMS WBS**
  - **US responsibility to International CMS core software is constrained to be a level-of-effort contribution of 25%**
  - N.B. by definition software engineers do not work on tasks which could be done by physicists
- **US-specific software support for physicists**
  - **Each engineer spends 25% of their time on software support**
    - i.e. only 75% is dedicated to US deliverables for CMS
- **Peripheral items to support engineers**
  - Desktops, travel, training, documentation, licenses,…
  - Less than 10% of the total
- **Management reserve**
  - 10% globally for US project
  - Not contingency: flexibility to cope with unforeseen problems

CAS WBS rolling planning

- Adopt a **rolling approach to software planning**
  - More detail (deliverables, milestones, etc.) in short-term
  - Longer-term resources according to "level-of-effort" scaling
    - Optimal use of resources in a changing environment
- Don't panic…
  - there is **always a complete plan**
  - we roll forward the **level-of-detail and responsibilities**
  - the required **resources are constrained by the CMS envelope**
- Need to **ensure consensus of US-CMS & agencies**. E.g.
  - Minor changes (~L5) quarterly review (ASCB / JOG)
  - Major changes (~L4) annual review (CB, DOE/NSF)
- Need to **ensure consensus of CMS**
  - Increase formality of reporting to CMS project
    - CMS Software & Computing Technical Board (6 times / year)
    - CMS Software & Computing Board (4 times / year)

CAS WBS rolling planning

Recall, e.g.
L2 = CAS
L3 = IGUANA
L4 = Interactive graphics
L5 = GUI Extensions
L6 = Tree widget

- **t → t+1year** (i.e. **FY2001** for t=today)
  - Define US-CMS tasks to typically **level 5 / 6**
  - Associate deliverables, milestones
  - Assign resources to each task
  - Ensure that, by definition, the rolled-up resources assigned in the US-CMS WBS are the sum of:
    - *canonical 25% scaling of full CMS software project*
    - *US-specific support for physicists 25% of US-CMS total*
  - Consult repeatedly with CMS
- **t+1year → t+2year** (i.e. **FY2002** for t= today)
  - As above but only define tasks to typically **level 4 / 5**
- **beyond t+2years** (i.e. **FY2003,4,5,...** for t= today)
  - As above but define tasks to ~ **level 3** (~ ongoing resolution)
  - US-CMS responsibilities are essentially **level-of-effort**

CAS WBS: Task Overview

- Scope is determined by scaling laws from CMS WBS but **the US tasks are well defined in the US CMS WBS**

1	User Facility (UF)
2	⊟ Core Application Software (CAS)
2.1	⊟ Software Architecture
2.1.1	Overall Achitecture and Framework
2.1.2	⊞ Subsystem Architectures
2.1.3	⊞ CMS architecture forum and evaluation (CAFÉ)
2.2	⊟ Interactive Graphical User Analysis (IGUANA)
2.2.1	⊞ Graphical User Interface
2.2.2	⊞ Detector&Event Visualisation
2.2.3	⊞ Interactive Analysis and Presentation
2.2.4	⊞ IGUANA Infrastructure
2.3	⊟ Distributed Data Management and Processing
2.3.1	⊞ Distributed Process Management
2.3.2	⊞ Distributed Database Management
2.3.3	⊞ Load Balancing
2.3.4	⊞ Distributed Production Tools
2.3.5	⊞ MONARC
2.4	⊞ Support

Northeastern
UC Davis
Caltech
FNAL

Northeastern

Caltech
FNAL
Princeton

Canonical scaling of 25% of the time for all on-project personnel

CAS Engineers for FY 2000

All hires are new to CMS (not existing CMS personnel)

Institute	Main CAS Tasks and Responsibilities	Off-project Developers	CAS Engineers
Caltech	MONARC leadership, distributed system simulations, ODBMS test beds, grid R&D, ORCA development, videoconferencing, production and software support	J. Bunn, M. Hafeez, T. Hickey, K. Holtman, H. Newman, A. Samar, R. Wilkinson	**V. Litvin** **I. Legrand**
FNAL	Objectivity production DB, software support, training, development tools	J. Amundsen, M. Litmaath, S. Muzaffar	**H. Wenzel** **G. Graham**
Northeastern	IGUANA leadership, interactive graphics/analysis tools, detector and event visualisation, Café architecture, dependency analysis, documentation tools, software support	G. Alverson, T. Paul L. Taylor S. Villa	**Y. Gaponenko** **L. Tuura** **1 requested**
Princeton	ORCA leadership & calorimetry, architecture, distributed database and production management tools	D. Stickland C. Tully	**T. Wildish**
UC Davis	ORCA, detector description, G3/G4 simulations	R. Breedon T. Cox	**M. Case**

- Detailed description of WBS tasks
  - **Printed documents**
  - **Breakout session**

Significant leverage of off-project resources

CAS WBS 2.1
Software Architecture

1	User Facility (UF)
2	⊟ Core Application Software (CAS)
2.1	⊟ Software Architecture
2.1.1	Overall Achitecture and Framework
2.1.2	⊞ Subsystem Architectures
2.1.3	⊞ CMS architecture forum and evaluation (CAFÉ)
2.2	⊞ Interactive Graphical User Analysis (IGUANA)
2.3	⊞ Distributed Data Management and Processing
2.4	⊞ Support

Detector Description architecture (UC Davis)

Calorimetery architecture (Caltech)

Tracking / Muon architecture (FNAL)

Analysis architecture (Northeastern)

Architectural Documentation (Northeastern)

CAS WBS 2.1 Software Architecture

- **Café project (CMS Architecture Forum and Evaluation)**
  - Newly started CMS project with US leadership
  - Aims to document current CMS software architecture
    - Help to understand it and communicate to developers/users
    - Starting point for future evolution
  - Set of tools to manage a complicated suite of documents
    - use-cases, requirements, technologies, choices made, designs, plans, lessons learnt, …
    - Different "views" or slices through the information
    - Complex hyper-linking problem solved with XML technologies
- **First version of the Café documents and associated system have already been developed by US engineer**

CAS WBS 2.2 Interactive Graphical User Analysis (IGUANA)

1	User Facility (UF)
2	⊟ Core Application Software (CAS)
2.1	⊞ Software Architecture
2.2	⊟ Interactive Graphical User Analysis (IGUANA)
2.2.1	⊞ Graphical User Interface
2.2.2	⊞ Detector&Event Visualisation
2.2.3	⊞ Interactive Analysis and Presentation
2.2.4	⊞ IGUANA Infrastructure
2.3	⊞ Distributed Data Management and Processing
2.4	⊞ Support

Currently the sole responsibility of Northeastern

CAS WBS 2.2 Interactive Graphical User Analysis (IGUANA)

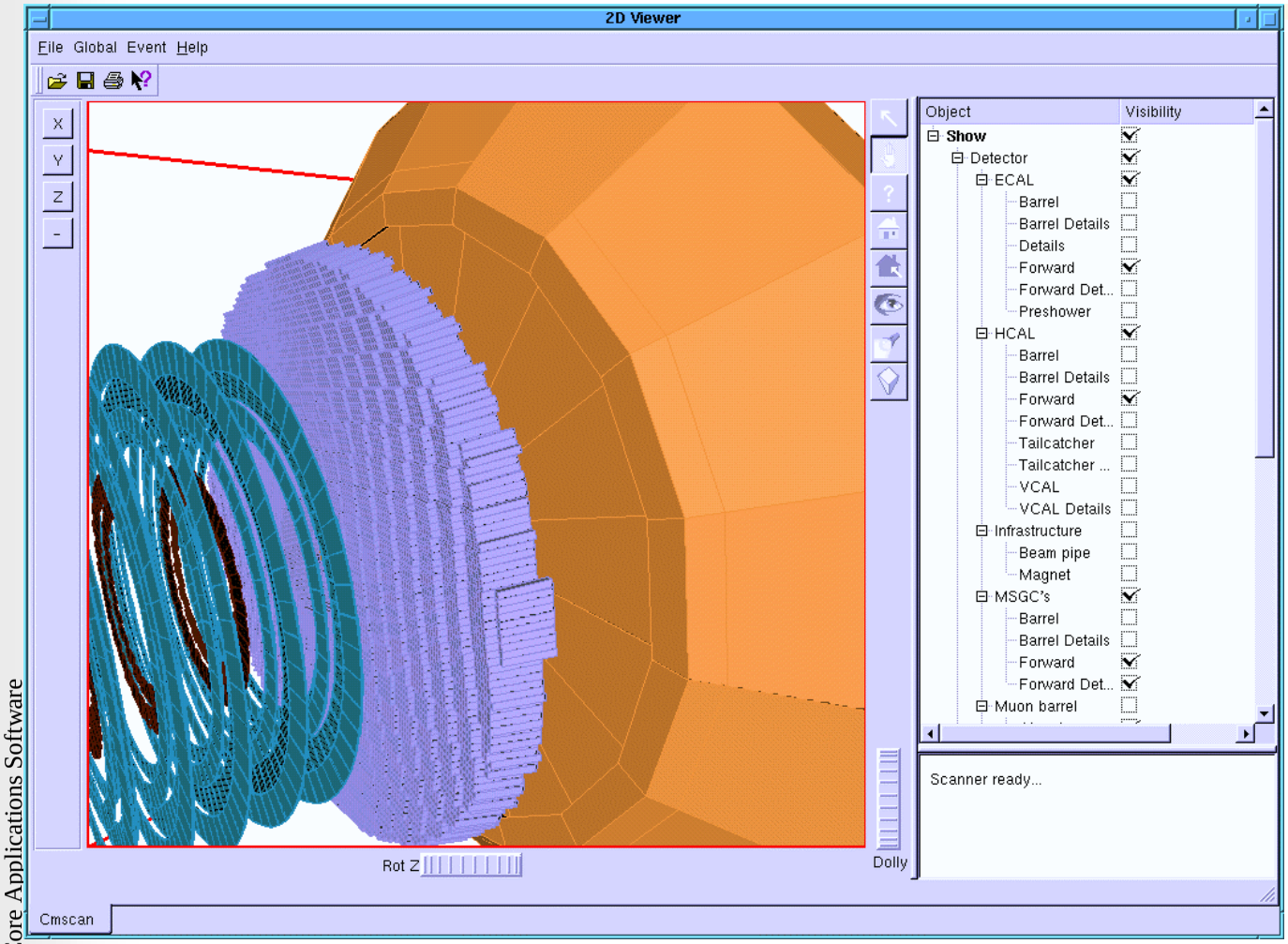
- **IGUANA Mission Statement:**
  - The IGUANA software project addresses interactive visualisation software needs for three domains:
    - **graphical user interfaces (GUI's);**
    - **interactive detector and event visualisation; and**
    - **interactive data analysis and presentation;**

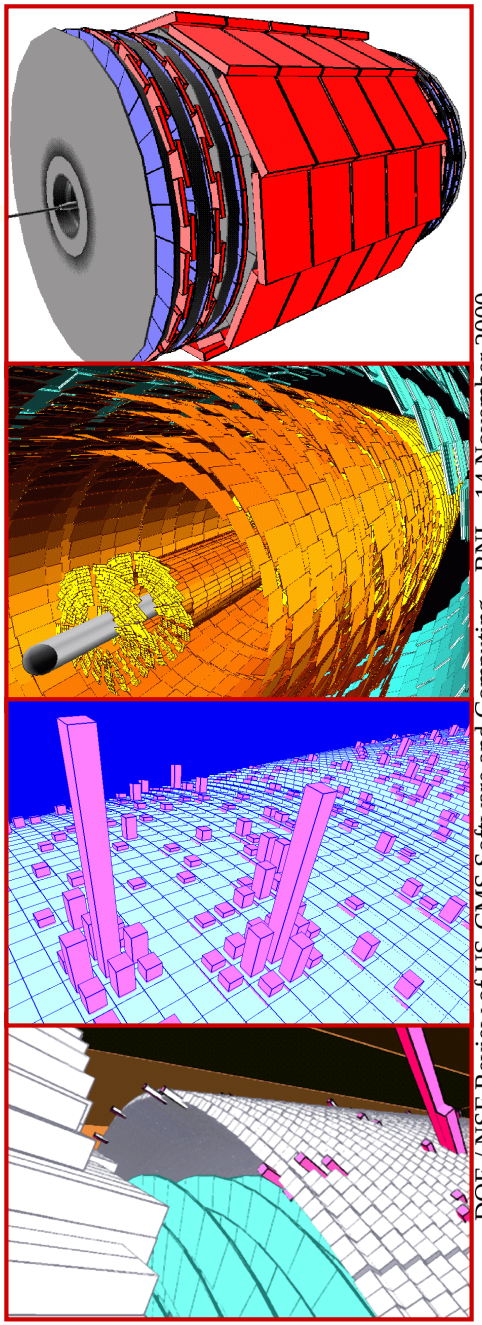
    for use in a variety of areas such as offline simulation and reconstruction, data analysis, and test beams.
  - Tasks include the assessment of use-cases and requirements and the **evaluation, integration, adaptation, verification, deployment, and support** in the CMS environment of visualization software from **HEP, academia, the public domain, and the commercial sector**.
  - Pre-existing software is exploited as much as possible to optimize the use of the resources available.

CAS WBS 2.2 Interactive Graphical User Analysis (IGUANA)

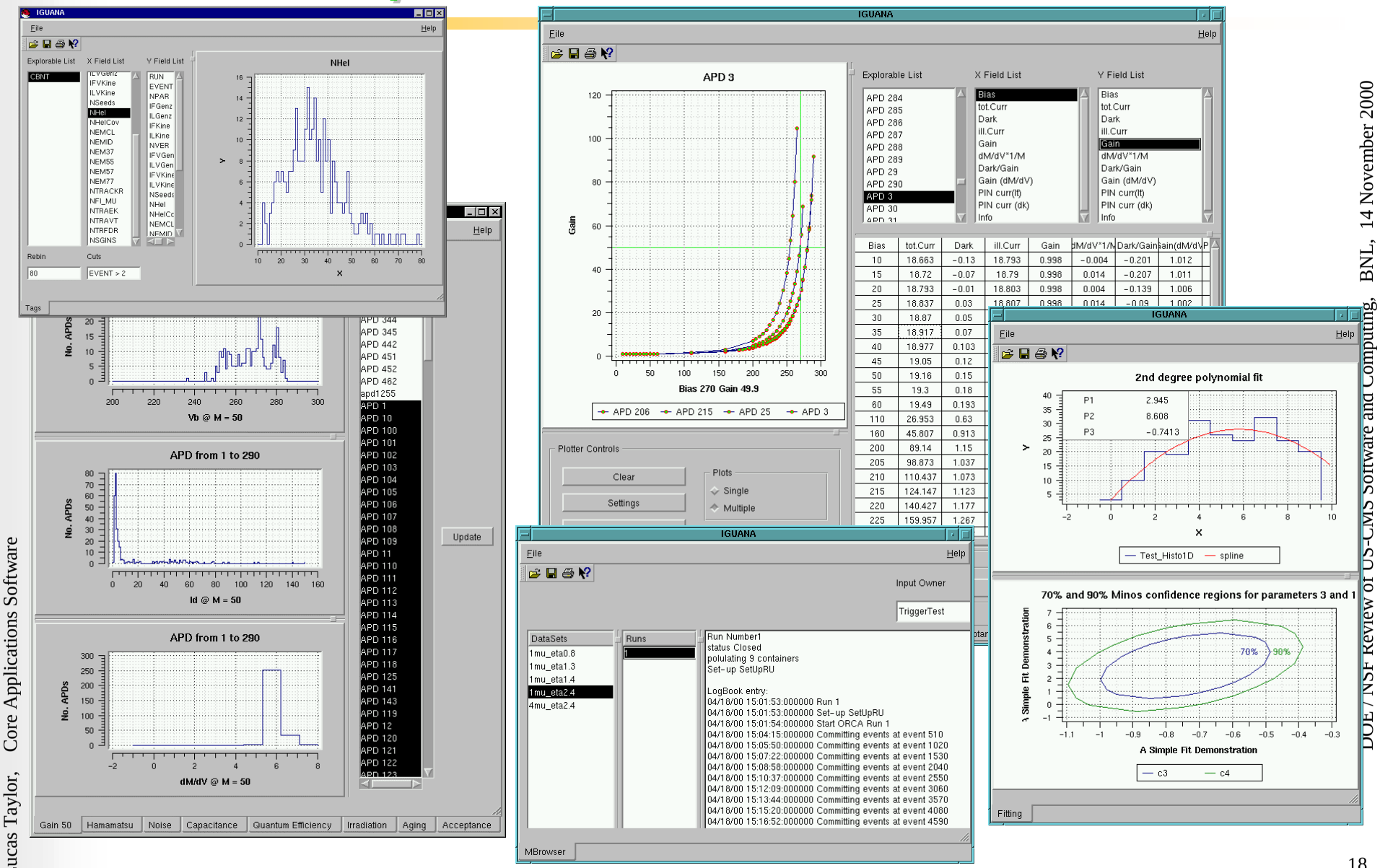
- Focus on a **sustainable medium to long-term strategy**
- Provide a **general set of tools** (not single application)
  - for a wide variety of applications and environments
  - for both experienced developers and non-expert users
- Strong emphasis on **modularity** and use of **standards**
  - to address the issues of scalability, maintenance, deployment, support, long-term evolution,…
- Pro-actively exploit **software developed elsewhere**
  - HEP, Public domain, Commercial
  - IGUANA work involves evaluation, integration, and support as well as developing extensions and CMS-specific software
- **IGUANA "Functional Prototype" just completed successfully (made possible by US CAS engineering)**

CAS WBS 2.2 User Analysis Environment (IGUANA)





CAS WBS 2.2
User Analysis Environment (IGUANA)



CAS WBS 2.3 Distributed Data Management and Processing

Caltech
Fermilab
Princeton

1	User Facility (UF)
2	⊟ Core Application Software (CAS)
2.1	⊞ Software Architecture
2.2	⊞ Interactive Graphical User Analysis (IGUANA)
2.3	⊟ Distributed Data Management and Processing
2.3.1	⊞ Distributed Process Management
2.3.2	⊞ Distributed Database Management
2.3.3	⊞ Load Balancing
2.3.4	⊞ Distributed Production Tools
2.3.5	⊞ MONARC
2.4	⊞ Support

CAS WBS 2.3 Distributed Data Management and Processing

- Prototype exists for Distributed Process Management
  - Allows selection of processors based on load, type, and data availability.
- Prototypes exists for Distributed Database Management.
  - Used to replicate databases to Fermilab and Padau.
  - Ready for use for world wide ORCA production this Fall.
- Load Balancing effort is beginning
- Production Support Tools
  - Tools to take advantage of existing production facilities.
  - Tools to archive and transfer results from centers worldwide.
- System Simulation
  - Successful validation of simulation package with simulation of Spring ORCA production.

CAS WBS 2.4
Support for US Physicists)

1	User Facility (UF)
2	⊟ Core Application Software (CAS)
2.1	⊞ Software Architecture
2.2	⊞ Interactive Graphical User Analysis (IGUANA)
2.3	⊟ Distributed Data Management and Processing
2.3.1	⊞ Distributed Process Management
2.3.2	⊞ Distributed Database Management
2.3.3	⊞ Load Balancing
2.3.4	⊞ Distributed Production Tools
2.3.5	⊞ MONARC
2.4	⊞ Support

Each engineer spends 25% of time supporting US CMS physicists

CAS WBS 2.4 Support

- CAS Engineers aid off-project developers in integrating, porting, optimizing their code.

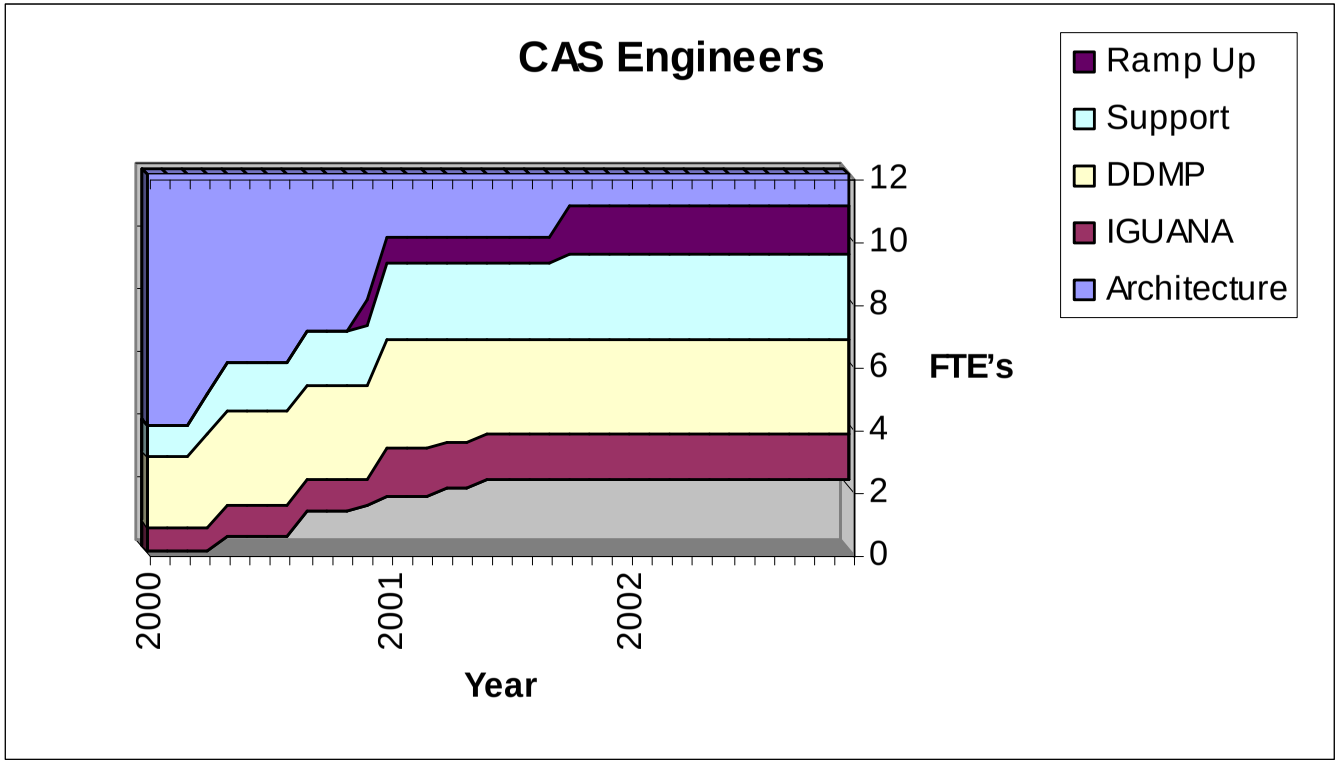
- User support
  - Documentation: Current and detailed documentation of US-CMS developed software
  - Examples: Working and useful examples to aid users learning the CMS analysis tools
  - Training: Ensuring that US-CMS physics are trained to work with the new software tools in the new environment.

- Software Support Tools
  - Guaranteeing that US-CMS released software is usable at remote sites by ensuring that it is configured and distributed with the agreed upon tools: scram and cvs.

CAS Personnel Needs

- Summary FTE profile

2001	2002	2003	2004	2005	2006	year
10	11	12	13	13	13	FTE



Summary

- **The US CMS "Core Applications Software" (CAS) is:**
  - well defined in terms of US responsibilities
  - well integrated with International CMS planning
  - well underway - a lot of solid work has already been done
- **Key CAS issues for FY2001 and FY2002**
  - Documentation, evolution, and consolidation of CMS architecture
  - Development and deployment of analysis software tools
  - Distributed data access / processing R&D and actual use
  - Solid support for US physicists
- **Sustained funding of software engineers is crucial**