

Status of the PDG Computing Upgrade

- Overview
- Public interface (demonstration)
- Encoder interface (demonstration)
- Editor interface
- Overall status and plans



Introduction

- While we were still able to produce the 2004 RPP using the old computing infrastructure, **the PDG computing upgrade is urgent:**
 - Only editor can enter data through Oracle/FORMS interface (Betty -> Piotr)
 - Old hardware without a readily available replacement system
 - Obsolete software – our only data entry interface is built on the no longer supported Oracle/FORMS without easy upgrade path to latest Oracle tools
 - Lack of automation for many tasks
 - Only very basic online access to RPP
 - A number of annoying smaller problems
- Work on PDG computing upgrade started mid 1990s
- In this presentation, I will concentrate on
 - Work done during this year

- **Current status and deployment plans**



An Integral View of PDG Computing

- Currently must support the following activities:

- **Production of the printed RPP book and booklet (every other year)**

- **Web edition (every year)**

- Diary (every year)

- PDG product ordering system (for North/South America, Australia, Far East)

- High-volume, reliable web server ($O(10^7)$ hits per year) with several mirrors

- Several different outreach activities

- Institutional address/contacts database

- US HEPFOLK (directory and census of US particle physicists)

- Reporting and statistics

- General computing (local computing resources, backups, monitoring, ...)

Will only discuss computing related to these core activities - but everything else must be addressed as well as part of the computing upgrade



Goals for the Upgraded PDG Computing System

- **Continue to be able to produce the RPP, booklet and web editions with the highest quality in a timely fashion**
- Allow encoders to enter data directly through a sufficiently simple interface
- Improve web access to PDG data and provide cross-links e.g. to SPIRES
- Streamline the book production and automate as many tasks as possible
- Provide support for managing the workflow between encoders, overseers, coordinators, verifiers and the editor
- Provide a reliable state-of-the-art computing infrastructure
- Long-term (not discussed here): add additional features such as e.g. advanced searching of the RPP database, direct database access for external user applications (such as MC codes), ...



Strategy for the Present Upgrade (I)

- So far, RPP data stored in relational database using about 60 different tables
 - Database structure developed and refined over more than 10 years
 - Addresses all current scientific needs
 - Not compatible with modern database design (compromises due to computer performance available when database structure was designed)
 - The database structure together with the auxiliary programs define a relatively rigid structure that limits changes to the appearance of the RPP to some extent
 - Right now, this is a relatively minor problem
- **Proper redesign of database structure would be difficult, and it would require an amount of work that is clearly not compatible with current resources**



Strategy for the Present Upgrade (II)

- **Carry over database structure into upgraded system (for now)**
 - Makes upgraded system “backward compatible” with the old system
 - Auxiliary programs continue to work
 - Add further database tables for additional functionality as needed by new interfaces (additional tables do not contain RPP content)
- **Develop new user interfaces for editor, encoders, and for public access**
 - The different interfaces should be viewed as separate components communicating only via the database
 - **They can be developed, improved (or replaced) independently**
 - May use common code libraries



Interfaces (I)

- **Editor interface:**
 - (Web based) application that mimicks the current Oracle/FORMS interface
 - Essentially a customized database editor
 - Full functionality for producing the RPP in the “old style”
 - Requires expert-level knowledge about RPP database structure
- **Encoder interface:**
 - Web based application that allows
 - encoders to enter new measurements,
 - overseers to check, correct and sign off on new entries,
 - and the editor to correct and sign off on new entries
 - Task driven
 - Intended to become the primary means of entering new data, while editor interface will eventually only be used for special tasks

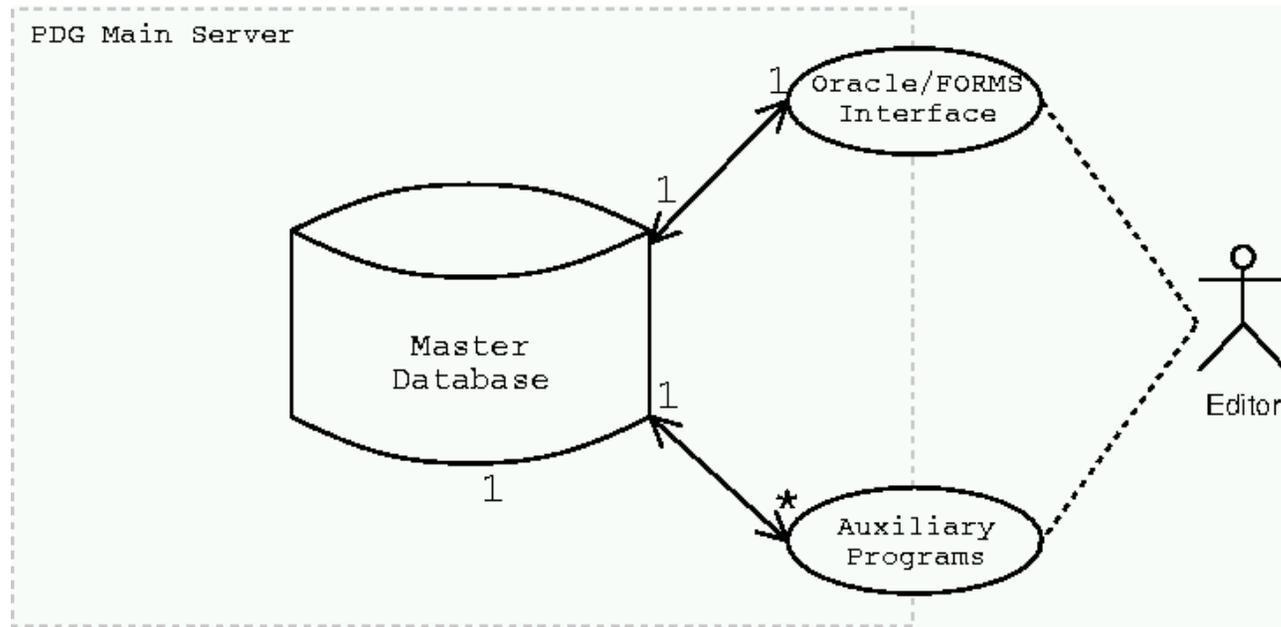


Interfaces (II)

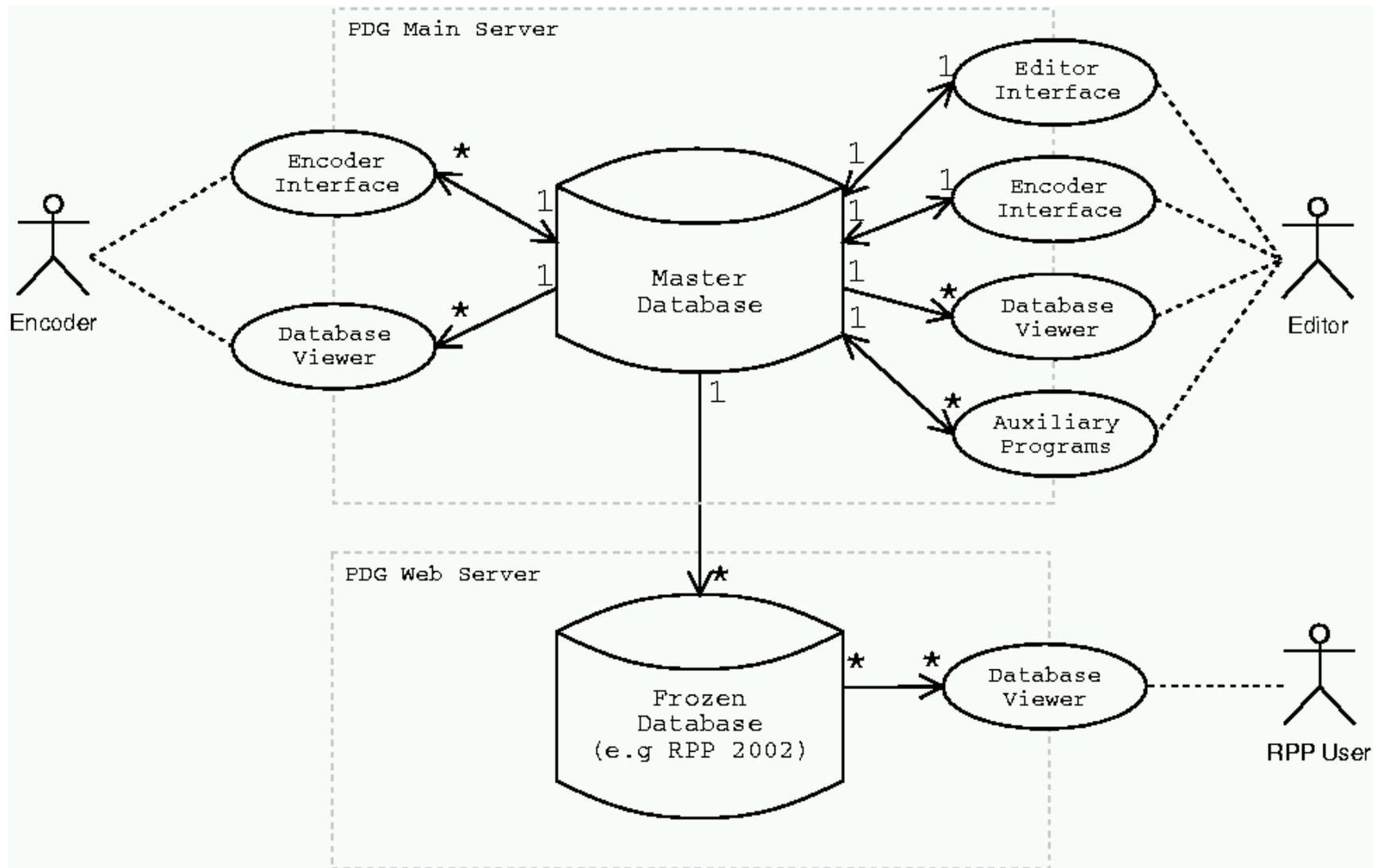
- **Auxiliary programs and scripts:**
 - Run to do fitting etc.
 - Processing of database contents in order to produce TeX file for book(let)
 - Directly interfaced to the database
 - Used only by editor
- **Database viewer (“public interface”):**
 - Web based application that allows browsing of the database contents
 - Aimed primarily at giving public access to frozen versions of the RPP (such as the 2002 edition), but it can run on any desired version of the database
 - Content is generated dynamically from the database (as opposed to making PDF files available on the web)
 - Provides direct links from RPP entries to SPIRES to the actual paper(s)
 - Advanced features will be added later



High-Level Architecture – Current System



High-Level Architecture – Upgraded System



Upgrade Plan

- **Phase 1: Switch to an upgraded system in time for the RPP 2006 edition**
 - Bare minimum requires new hardware (✓), stable database operation (✓), auxiliary programs (✓), and **editor interface (to be completed)**
 - Would like to deploy at least a first version of the encoder interface
 - Public interface is almost ready, so it will be included
 - This corresponds roughly to what was called “RPPs 1.5” previously
- **Phase 2: Improve the upgraded system in an incremental fashion**
 - Improvements resulting from phase 1 will free resources to work on phase 2
 - Streamline book production
 - Improve (or replace) individual interfaces
 - Add new interfaces for new functionality
- **Phase 3: If necessary, revise data model (ie. redesign database structure)**



Contributors

- From Protvino:
 - Vitaly Lugovsky (core libraries, database)
 - Kirill Lugovsky (web interfaces)
 - Slava Lugovsky (web interfaces)
 - Lyudmila Lugovskaya (documentation)
 - Vladimir Ezhela (group leader)
- From LBNL:
 - Juerg Beringer (project leader, since March 2004)
 - Orin Dahl (auxiliary programs, Oracle/FORMS related work)
 - Piotr Zyla (daily operation, production tasks, editors interface)
 - Jeff Anderson and Bryan Abshier (system management)
 - Mike Ronan (census, prototype editor interface)

These are all part-time contributors, mostly at the 10% to 50% level



Organization

- Visits of the Protvino team to LBNL:
 - Several visits over the last few years
 - Most recent visit: August–October 2004 by Kirill, Lyudmila and Vitaly
 - Next visit planned for spring 2005
- (Bi-)weekly PDG computing video conference meetings
- Frequent communication via e-mail
- Periodic reviews:
 - Last review by LBNL-internal review committee (I. Gaponenko, R. Jacobsen, S. Loken, M. Shapiro)
 - Next review planned for end of November or December



Recent Progress

- Initial version of Oracle/FORMS like editor interface
- Reverse-engineering and documentation of constraints in old Oracle/FORMS interface. Started implementation of same constraints in new editor interface.
- LBNL-internal encoding exercise using the new encoder interface
- Implementation of workflow management in encoder interface
- Revised public interface
- Progress with documentation
- Purchased hardware of new servers and started installation
- Clean up and installation of development environment on new main server
- Performance analysis and code optimization (in progress)



Public Interface

The screenshot shows a Netscape browser window titled "RPP Browser - Netscape". The address bar contains "http://pdg.lbl.gov/". The page features the PDG logo and navigation tabs for "Particle Listings" and "Reviews". A citation instruction reads: "Please use this CITATION: [S. Eidelman et al.](#), Phys. Lett. B 592, 1 (2004) ([bibtex](#))". The main content is organized into a grid of categories, each with a dropdown arrow and a list of sub-items:

GAUGE AND HIGGS BOSONS	LEPTONS	QUARKS
<ul style="list-style-type: none">Reviews on BosonsγgluongravitonWZHiggs BosonsHeavy BosonsAxions	<ul style="list-style-type: none">Reviews on the Leptonse, μ, τHeavy Charged Lepton Searchesν_e, ν_μ, ν_τNumber of Neutrino TypesDouble-β DecayNeutrino MixingHeavy Neutral Leptons, Searches for	<ul style="list-style-type: none">Reviews on QuarksLight quarks (u, d, s)cbtb' quark, searches forFree quark searches
MESONS	BARYONS	Other Searches
<ul style="list-style-type: none">Reviews on MesonsLight Unflavoured MesonsOther Light Unflavoured MesonsStrange MesonsCharmed MesonsCharmed, Strange MesonsBottom MesonsBottom, Strange MesonsBottom, Charmed Mesons$c\bar{c}$ Mesons$b\bar{b}$ MesonsNon $q\bar{q}$ Candidates	<ul style="list-style-type: none">Reviews on BaryonsN BaryonsΔ BaryonsExotic BaryonsΛ BaryonsΣ BaryonsΞ BaryonsΩ BaryonsCharmed BaryonsDoubly-Charmed BaryonsBottom Baryons	<ul style="list-style-type: none">Reviews on Other SearchesMagnetic Monopole SearchesSupersymmetric ParticlesTechnicolorSearches for Quark and Lepton CompositenessExtra DimensionsWIMPs and Other Particle Searches

Status of Public Interface

- Basic functionality is implemented
- **Performance bottlenecks** are understood (we think) and are being addressed
 - 60% of time is spent processing nested regular expressions
 - 40% of time is spent on string concatenation
- Work on improving the layout of individual pages
 - Font sizes
 - Graphical layout, alignment issues, etc.
- **Address browser-specific issues**
 - Would like to support reasonably current versions of all major browser
 - Support for browsers running under Linux, Win/XP and Mac
- **Testing, testing, testing ...**



Demonstration of the Public Interface



- We need a brave volunteer (who has never used the public interface) to test drive during our demonstration
- Suggestion:
 - use the public interface to find the best limit on B_s mixing ($\Delta m(B_s)$)
 - display the paper describing that measurement
 - find out if other results from that paper were included into RPP



Encoder Interface

The screenshot shows a Netscape browser window titled "PDG List of papers - Netscape". The page content includes a header with the PDG logo and "Encoding System" text. A navigation bar contains links for "Sign Out", "e-mail to Editor", "RPP Viewer", "Vocabularies", and "Help". Below this is a task management section with filters for status and particle, showing "Total tasks: 2 2 tasks listed". A table lists two tasks with columns for Status, Author, Year, Journal, Volume Number, Page Number, and Add'l Info. The first task is "Encode" by "JAMIN" in "EPJ" journal, volume "C24", page "237". The second task is "Encode" by "LOREDO" in "PR" journal, volume "D65", page "063002". A sidebar on the left has an "Add New Paper" button. The footer of the browser window shows "Page 1 of 1".

PDG
particle data group
Encoding System

You are logged as BERINGER

[Sign Out](#)

[e-mail to Editor](#)

[RPP Viewer](#)

[Vocabularies](#)

[Help](#)

Please choose a task from the list on the right, or click one of the options below:

[Add New Paper](#)

Filter tasks by status: Filter tasks by particle:

Total tasks: 2 2 tasks listed

Status	Author	Year	Journal	Volume Number	Page Number	Add'l Info
Encode	JAMIN	2002	EPJ	C24	237	
Encode	LOREDO	2002	PR	D65	063002	

Page 1 of 1



Status of the Encoder Interface

- Basic functionality is implemented
 - Need feedback on requirements from meson team
- LBNL-internal encoding exercise in September 2004 was mostly successful
 - Main problem: lack of sufficient workflow management (resolved)
 - Many small issues (mostly resolved)
- Ongoing effort to simplify interface
 - Find better solutions for entering “PDG macros” into database
- Improve and extend online help
- **Testing, testing, testing ...**



Demonstration of the Encoder Interface



- Demonstration of the most frequent use-case:
 - Literature searcher enters a new paper and assigns it to an encoder
 - Encoder enters a new measurement, checks it, and signs it off
 - Overseer checks the new measurement and signs it off
 - Editor signs off on the new measurement

Editor Interface

PDG
particle data group
Encoding System

You are logged as PIOTR

[Sign Out](#)
[e-mail to Editor](#)
[RPP Viewer](#)
[Vocabularies](#)
[Help](#)

Please choose a task from the list on the right, or click one of the options below:

Filter tasks by status: Filter tasks by particle: Filter tasks by encoder:

Total tasks: 466 466 tasks listed

Status	Author	Year	Journal	Volume Number	Page Number	Add'l Info
Reference/Measurement authors, institution, comments, verifiers, igipc, encoding, other related papers, reference also, measurements, footnotes	ETSHIN	2004B	PL	B580	119	
	ETSHIN	2004B	PL	B580	119	
	ETSHIN	2004	PL	B578	285	
Measurement/Datablock footnotes, measurement column header, units, average control, conservation laws, result summary	BERT	2004	PR	D69	011103	
	AI	2004A	PR	D69	012003	
	AI	2004A	PR	D69	012003	
Decay table footnotes, texts (clumps & decay headers), decay comments, conservation laws, result summary	AI	2004B	PRL	92	052001	
	AI	2004	PL	B578	16	
	WDER	2004	PL	B578	365	
	KANOVA	2004	PL	B578	33	

Page 1 of 47

Technical Editor Options

- ◀ Main Forms
- ◀ Other Forms
- ◀ Vocabularies
- Node Tree
- ◀ Programs
- ◀ Utilities
- home

Status of the Editor Interface

- Implementation of constraints present in Oracle/FORMS interface is in progress
 - Complete documentation of all required constraints available
 - Prototypes of two pages implemented
- Need tools for editing user information (name, password, roles, ...)
- Need tool to view and undo database transactions
- Need some monitoring facilities
- **Editor interface still requires substantial work (few months)**



Prototype Measurement/Reference Form

Decay/Decay comment form - Netscape

SOURCE_NAME	YR	OCC	PUBLICATION_NAME	FOREIGN	TM	PREFID		
Abe	2003						Search	reset
1 ABE	2003	B	PR D67 032003			49205	copy	delete
2 ABE	2003	C	PR D67 031102R			49208	copy	delete
3 ABE	2003	D	PRL 90 131803			49220	copy	delete
							copy	clear
							copy	clear
							copy	clear
							copy	clear

7
[NEXT](#) | [LAST](#)

save

NOD	MEASUREMENT	CL	EVTS	REFERENCE OCCURRENCE			TECN	CHG	INLINE COMMENT	MMYY	F	I	I	TYPE	H	I	Search	reset	
				Name	Year	Occ													
1	S042C 0.56+-0.41+-0.16			ABE	2003	C	BELL		Repl. by #ref{ABE 200:3	2003							N	copy	delete
2	S042C 0.40+-0.33+0.28-0.10			ABE	2003	C	BELL		Repl. by #ref{ABE 200:3	2003							N	copy	delete
3	S042S -0.73+-0.64+-0.22			ABE	2003	C	BELL		Repl. by #ref{ABE 200:3	2003							N	copy	delete
4	S042S -0.49+-0.43+-0.11			ABE	2003	C	BELL		Repl. by #ref{ABE 200:3	2003							N	copy	delete
5	S042Y -0.26+-0.22+-0.03			ABE	2003	C	BELL		Repl. by #ref{ABE 200:3	2003							N	copy	delete
6	S042Y 0.71+-0.37+0.05-0.06			ABE	2003	C	BELL		Repl. by #ref{ABE 200:3	2003							N	copy	delete
																		copy	clear
																		copy	clear
																		copy	clear
																		copy	clear



Overall Status and Deployment Plan

- Basic functionality of public and encoder interfaces is implemented, and a prototype implementation of parts of the editor interface is available
- Sufficient progress has been made to start preparing for deployment of the upgraded system (phase 1)
 - **We think that a deployment on the time scale of early next summer is not impossible,**
 - **Assuming significant contributions from Protvino team over next few months**
 - Aggressively pushing for deployment is the best way to transfer knowledge to the LBNL PDG group and to make this project succeed
 - We want to exploit the momentum gained during the last visit to LBNL
 - Major remaining issues are
 - **Completion of editor interface**
 - **Resolution of performance issues in public interface**
 - Planning migration of non-RPP parts over next few months (web server, ...)



Hardware

- **Main server** (pdg0.lbl.gov):
 - Main “workhorse” for all PDG work, accessible only to PDG members
 - Mirrors web server and could take over its function immediately if needed
 - RPP production, user accounts, backups, ...
 - Dual-CPU 2.8GHz Intel Xeon, 2GB RAM, 512GB RAID mirrored disk
 - Status: used for development since August 2004, being reconfigured
- **Web server** (will become pdg.lbl.gov):
 - Publicly visible machine handling all external requests
 - PDG and outreach web access, ordering system, ...
 - Not necessary for book production
 - Dual-CPU 3.0GHz Intel Xeon, 2GB RAM, 256GB RAID mirrored disk
 - Status: currently being installed



Can We Maintain the New RPP System? (I)

- Technologies (all open-source):
 - PC/Linux - Mainstream hardware and operating system
 - PostgreSQL – mature relational database
 - BRL – A Scheme based system to embed dynamic content into web pages
 - Kawa – Java based Scheme system (alternative: Bigloo)
 - Tomcat – Servlet container used to run BRL / Kawa within the web server
 - Note: Kawa and BRL are not mainstream tools
 - Should be ok for a few years, but availability/support needs to be watched
- It will be a moderately complex system:
 - 6k lines of Scheme in core libraries
 - O(50k) lines of BRL with some embedded JavaScript
 - 55k / 4k lines of Fortran / C in auxiliary programs (excluding inline doc)



Can We Maintain the New RPP System? (II)

- Existing documentation of the current system:
 - “Red Book” (230p) - database structure and procedures
 - “Green Book” (30p) – same for diary and institution list
 - Several notes describing recent changes to database structure
 - “Forms Book” (100p) – constraints imposed by Oracle/FORMS interface
- Documentation of the upgraded system (in progress):
 - [“Data Handling Software Requirement Specification” (106p) – partly outdated requirements specification]
 - Installation and developer's guide (19p) – technical information
 - Encoding system description (37p) – technical information, to be updated
 - Online help for all interfaces
 - Wiki-based system log book (will include pointers to all documentation)



Can We Maintain the New RPP System? (III)

- Exercise 1: Can we install and run the encoding system without further help?
 - Using only the preliminary documentation, it took less than 1 day to install everything needed (starting from a standard RH 9.0 installation), and to import the 2004 RPP database and run the encoder interface
- Exercise 2: Can we write a BRL page displaying information from database?
 - To be done
- Conclusions (re maintaining the system w/o help from the developers):
 - Installation ✓ - proven by exercise 1
 - Daily operations (start, stop, backup, ...) ✓ - proven by exercise 1
 - Small bug fixes ✓ - to be proven by exercise 2
 - Small improvements probably yes
 - Adding major new features (might choose other technology)



Plans Up to Deployment ...

- Finish hardware installation
- Start migration of non-RPP parts of PDG computing while completing editor interface and addressing remaining issues with public and encoder interfaces
- Planning several intensive rounds of testing of public and encoder interface:
 - Within LBNL
 - By volunteers within PDG collaboration
 - Stress testing
- Extensive testing of editor interface and book production:
 - Reproduce (parts of) the TeX file of the 2004 edition with the new system and understand any differences
 - Go through a complete book production exercise with the new system
- Hope to switch to upgraded system in summer 2005 in time for the 2006 edition (some encoding will happen beforehand with old system)



... and Beyond

- Deploying the upgraded system will be a gradual process
 - Encoders will gradually start to work with new system
 - Expect bug fixes and improvements as required during this phase
 - Initially more work, before improvements will start to pay off
- Once we switch to the new system, streamlining of the book production process will begin immediately
 - Again, initially more work, before improvements will start to pay off
- After gaining some operating experience with the upgraded system will
 - Revisit plans for further upgrades (phase 2)
 - Decide if maintainability concerns warrant rewriting (parts of) the new system
 - Decide if we need to redesign the database structure (phase 3)



Conclusions

- The Protvino team has made a lot of progress during their last LBNL visit
- **We are aggressively pushing forward towards deployment of an upgraded system**
 - The earliest realistic date for deployment is sometimes early next summer
 - Achieving this implies significant contributions from the Protvino team over the next few months
 - We are planning another visit of the Protvino team to LBNL next spring to help us with the deployment
- Deployment will require a lot of work that must take place in parallel to the production of next year's web edition
 - **Manpower will likely be a problem - may delay the deployment schedule**
 - We'll do the best we can given the resources we have
- Stay tuned for good news sometimes next year!

