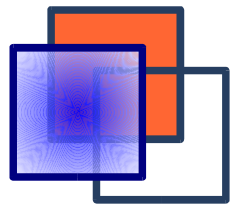


DocDB Overview for ILC EDMS

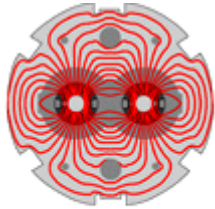
Eric W. Vaandering
(ewv@fnal.gov)

Vanderbilt University

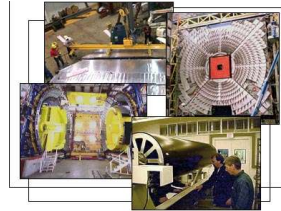
9 February 2006



DocDB Usage



LARP



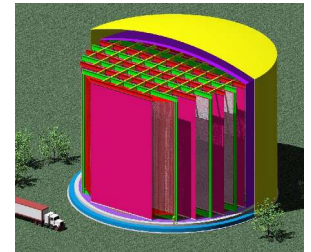
Pixels
Simulations



- Exact numbers unknown, but...



- About 20 groups using DocDB
- Over 1000 authors/users
- At least 10,000 documents managed



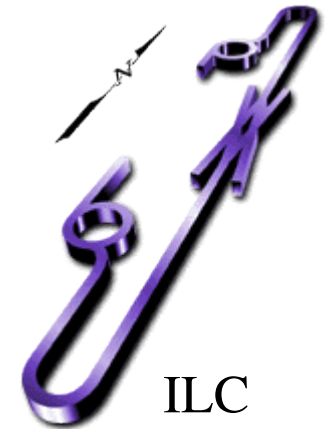
LArTPC



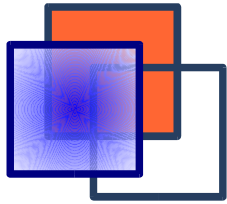
OSG



DARK ENERGY
Survey

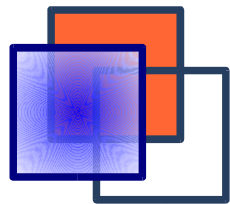


ILC



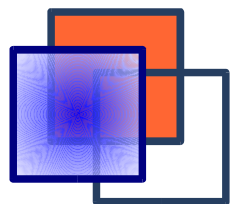
DocDB History

- First demo of protoDocDB: Nov 2001:
“What's missing is making the HTML pretty, being able to update a document, multiple file updates, security, and a search and display function.”
- V2 – Keywords, Search: Jan 2002
- V3 - Email Notification: Jun 2002
- V4 - Meeting Organizer: Aug 2003
- V5 - Signoffs, Notes, Meeting Security: Apr 2004
- V6 - Certificates, CSS: Jan 2005
- V7 – Aug 2005: Better events, calendar, XREFs
- Dec 2002: MI (later AD) adopts, Harry Cheung: “Still you are talking about upgrades....”
- Admin interface: Feb 2003
- Mar 2003: DocDB shown at CHEP
- Jan 2004: Gretina adopts, GPL chosen
- Mar 2004: Computing Division adopts
- Jun 2004: CD Contracts for certificates and other improvements
- BTeV Canceled: 7 Feb 2005
- MINOS, Minerva, EPP, LArTPC, and more begin using DocDB



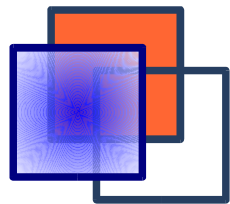
Some Numbers

- Current version
 - 29,000 lines of code including comments
 - From SLOCCount
 - 19,500 lines of code (no comments)
 - Estimated 4.5 person-years to develop
 - Estimated cost to develop: \$610K
 - Clearly overestimated, but interesting



Design Philosophy

- In BTeV one of our post-docs related his legal clerking experience: “Everything is a Document.” Implications:
 - Use a document system for all documents. If you wrote it and you want to share it with just a few people, put it in the document system
 - A document system should be trivial to use. The lower the barrier to entry, the better. If people can't easily share or find information, they won't
 - All documents in one place; no distinction between talks, papers, musings
- If people in a collaboration are routinely e-mailing files to more than 3 people, the system is not adequate.



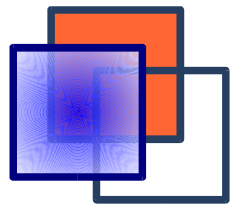
Introduction, cont.

With the files of a document, we store additional info:

- Title and Abstract
- Creation and modification times
- Authors, subject, who is allowed to view & modify
- Publication information
- Associated events

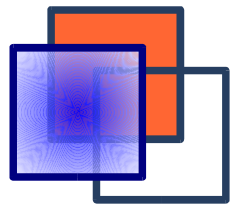
All this information is easily searched, allowing collaborators to find the information they need easily.

A reasonable expectation is that ILC will want to archive this information until ~2030.



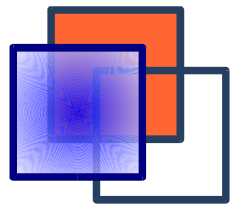
Archiving

- Each document can have many revisions, old revisions preserved.
 - Each version is independent, can make only latest public
 - Differs from other document systems
- A document is a collection of files, stored on central server (in file system)
 - Powerpoint, Word, PDF, etc. (any file format)
- Shouldn't be hard to migrate to a new system at a later date if needed



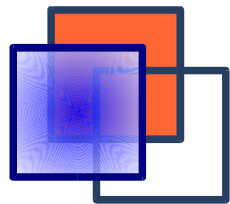
Security

- Access by group name and group password or by X-509 Certificate
 - Get certs from DOE or other CA
- Can have many groups share one DocDB
 - Reviewers have read-only access to selected documents
 - Support groups can upload documents, but don't have full access to all documents
 - Public has access to publicly available documents
 - Other groups need not share with the whole collaboration
 - “Viewers” and “modifiers” are different lists
 - Some documents only modifiable by management



Implementation

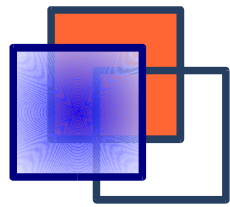
- Entire DB is managed and accessed via a web interface
- Users can easily upload documents from their PC or have them downloaded from the Web
- E-mail notifications
 - New or changed documents of interest
 - Either immediate e-mail or a daily/weekly digest
 - Also e-mail notification for sign-off procedures
 - Especially useful for managers, who can track changes to all their documents easily



DocDB for Reviews

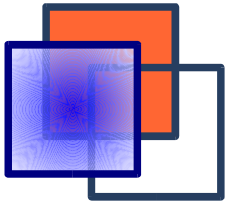
For BTeV there was a DocDB account for reviewers.

- Reviewers are not allowed to change anything, but a large number of documents were available to reviewers
- This allowed us to easily make information available that may not be printed and allows access before and after the review time



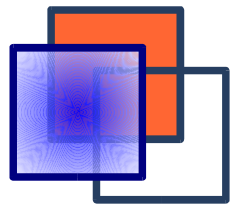
Sign-off and Change Control

- Aims to meet DOE requirements for CD2/3 as detailed by [SC-81](#) “Records” document
- Signatories sign by entering their password (or using certificate)
- Next signatory is notified by e-mail that document is ready
- Adding control to existing document is trivial



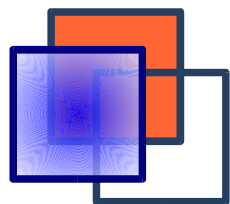
DocDB API

- You can add documents to DocDB without using DocDB
 - Currently a Perl API
 - Likely to be an XML wrapper around this API
 - Possibility to have various XML APIs to get data out of DocDB as well
- Currently in use by at least four groups
 - A bit of learning is needed, but already quite easy
- This should allow for interfacing with InDiCo and using DocDB as a backend document store
 - Other groups at FNAL are interested/pursuing this also



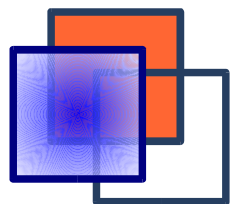
Other Comments

- Full search of meta-data built in, search of document contents by outside search engine easily added
- Time to upload a document: About 45 seconds
- Administration (adding topics, authors, institutions) all done via the web
- Visual appearance, text is easily customized in several different ways. (It will look like *your* DocDB.)



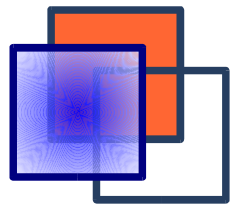
DocDB Version Numbers

- We use x.y.z version numbers
 - **X: Underlying DB format changes (also usually major changes)**
 - **Y: Major feature changes users are likely to notice**
 - **Z: Minor feature changes and bug fixes**
- Since DB format changes are disruptive, we want to keep these to a minimum



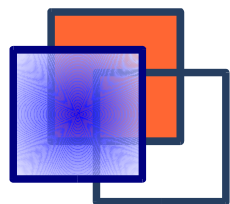
Possible V7 Improvements

- XML Insert/Show Document to complement API
 - Easier insertion of docs by external programs
 - Easy access to DocDB info by ext. programs
 - Would allow external DocDB “Related Docs” to show titles, not just numbers
- Web based configuration
 - Lots of options in DocDB/Project Globals
 - Most can be stored in DB (not security related)



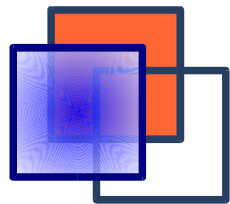
Version 8 Thoughts

- DocDB may not currently scale up very well for large, disparate groups (more than 300-400 people) producing thousands of documents/month
- I think a few changes can help with this
 - **This would touch on three major areas in DocDB**
 - Topics (primary way documents are organized)
 - Events
 - Groups (of several to hundreds of users)



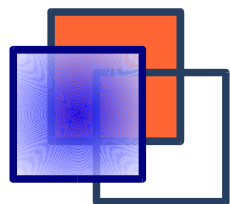
V8 Topic & Event Suggestions

- Allow infinite hierarchy of topics
 - Then can have SubSubSub...Topics
- In a large meeting, it can be difficult to find the sessions you want
 - Meetings and/or sessions within meetings can have optional topics
 - Lists of topics will also list events with those topics



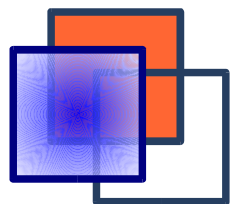
V8 Group Suggestions

- An old idea: Various groups do not see all options
 - e.g. certain groups may not be able to create documents with certain topics
 - or even see that a document HAS that topic
 - or have to take an extra step to add such a topic
 - same with events, document types, etc.
 - Allows moderators to designate “official” docs



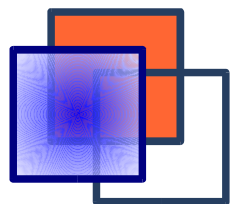
DocDB Support

- Support comes in three types
 - **Developer**
 - Currently only me under contract with FNAL. I think they are considering adding a lab person at some level as well
 - License is GPL
 - **Maintainer**
 - Installs and configures DocDB. Lynn is doing this for ILC test DocDB, Ruth Pordes's group does for many other experiments
 - **Administrator**
 - Adds topics, groups, other meta-data lists
 - You would want one or more ILC people in control of this



Getting a Change Made

- DocDB is actively developed, even after BTeV, but I don't work on it much during my “day job”
- So
 - Make the change yourself
 - Convince CD it's a good enough idea to ask for
 - DocDB is used by lots of groups, so changes should be mutually beneficial
 - DocDB will not incorporate other technology (e.g. Plone, InDiCo, etc.) but DocDB could be used as a backend to those systems with an XML (or other) API



Conclusions

- DocDB is different than other document systems
 - Emphasis is on collaboration rather than on forming only an official repository
 - Quick to share, view what others have done
 - File type agnostic
 - Can be used as an official and public repository as well
 - Modest changes can improve this functionality
 - Earliest drafts through final approved version
- Development after BTeV continues and adoption growing
- Code quality continues to improve (making future developments easier/possible)