

A DOCUMENT DATABASE FOR THE GLUEX EXPERIMENT

Carolyn A. Ingram
Sparro
Dept. of Physics, University of Regina
Regina, SK, Canada
S4S 0A2

ABSTRACT

The GlueX Experiment is a project that will run at the Jefferson Laboratory in the near future (ca. 2010). Research and development work for this experiment began in 1998 and has been progressing swiftly with the addition of new groups to the collaboration and the awarding of Critical Decision 0 by the Department of Energy. Therefore, it is timely to introduce a centralized document database to store the collaboration's technical notes, images, drawings, videos, etc., and one that will cover our needs for the duration of the project (2010-2020). This application, termed GlueX DocDB, is described herein.

We realized the need for a method of storing and centralizing documents in order to better categorize and describe the various types of documents that are part of the GlueX project. What was found was the DocDB (Document DataBase) package (written in CGI/Perl) developed by Eric Vaandering (Vanderbilt University) for the BTeV Collaboration. In this case a document was defined as any piece of information that a person wants to save and share with others. It could include not only text files but also video, technical drawings, photos, etc. The former method we had of storing documents within the GlueX portal under 'Downloads' was far less versatile. One could include just a single file per document and only had fields to fill in for a Title, Description, a single Author, a Version number, the date the file was added, and the File size. This was not sufficient for our needs.

There are many features of the DocDB that make it very attractive:ⁱ

- All documents are kept on a web server, not personal machines. This way we don't suffer from link rot (linked-to files disappearing from the web).
- Each document can have multiple revisions and old revisions are still available.
- Each revision of each document can have multiple files. This accommodates multiple file types (source and presentable) and/or child files (especially useful for trees of HTML documents).
- The ability to upload documents by browser upload or forcing the document database to download via http.
- The ability to not only have public and private documents but also documents that are accessible to subgroups (ie. reader, writer, etc.) which have separate passwords.
- Users will only know that a document exists if they have privileges to view it.
- The ability to easily move documents, or just certain versions of documents, from restricted to public and vice-versa.
- The ability to search all documents according to very specific criteria including words in a document Title or Abstract, by Keywords, Filenames, etc. as well as by Requester, Author, Document Type (especially good for looking up documents specified as Hall D Notes, Minutes, etc.), Modification Dates, and by Topics.

ADDING INFORMATION TO THE DATABASE

The DocDB is accessible by clicking on the 'Documents' module either at the top or left hand side of the GlueX Portal home page (portal.glueX.org). In order to add or update information in the database you must click on the 'Private' button then log in as a 'writer'. Next click Create or change documents or metadata on the DocDB home page. From here there are five options to choose from:ⁱⁱ

- 1) *Create a new document*: This is for documents that have not been entered or reserved in the database. This is the way the first version of a document is entered.
- 2) *Reserve a document #*: This is used for documents that do not yet exist, but a number is needed. This allows the writer to know the document number before the real document is placed in the database.
- 3) *Update document*: This is the way that versions of existing documents are updated in the database. The document number remains the same, but the version number is incremented.
- 4) *Update DB information* : This is used when the document has not changed, but the information about it has. For example , it is now published so that information has to be added. Security settings on documents are also changed this way. The version number remains unchanged.
- 5) *Add files to a document*: This option is used to add new files to an existing version of a document, perhaps files that were forgotten or an additional presentation format of an old document. If the content of the document has changed at all, the Update, not the Add file, option should be used. The version number remains unchanged.

The Requester can upload up to three documents by default, unless otherwise specified. From here it is necessary to fill in the Title, and Abstract fields. The keywords are optional but help to classify the different documents. It is recommended that the Keyword Chooser is used. The file(s) can be uploaded using the browse button or typing in their location, and in the 'description field' the type of file should be specified (ie. pdf, ps, tar, etc.). Only the first box must contain a file, while the others may be left blank. Be sure to select which document type you are adding, choose a requester (ie. yourself), specify the author(s), and choose the security level for your document.

The security level should be set to public for Hall D notes, talks and the vast majority of our documents. Only minutes of meetings or conference calls and a few other documents should be set to a higher security level. It is also necessary that you specify the topic(s) under which your document would fall. You can select more than one topic by using Ctrl-click. If there are no References or other publication information that need to be entered simply click on [submit document](#) to finish. You can upload an archived file such as .tar, .tar.gz, or .zip by using the customized form and selecting archive upload under *Upload type*. This form should be filled out in the same manner as a regular document.

Title: _____

Abstract: _____

Keywords: (space separated) - [Keyword Chooser](#)

Notes and Changes:

Local file upload:

File: _____ Browse...

Description: _____ [Main?](#)

File: _____ Browse...

Description: _____ [Main?](#)

File: _____ Browse...

Description: _____ [Main?](#)

Make sure you supply a format everyone can view (e.g. PDF) along with source formats.

Document type:

- Minutes Journal Publications Audio
- Talk Technical Drawing Hall D Note
- Other Proceedings Technical Note
- Video Manual Figure/Photo
- Poster MOU

Requester:

- Adams, Gary
- Anumagalla, Ravi
- Barbosa, Fernando J.
- Barnes, Ted

Authors:

- Adams, Gary
- Anumagalla, Ravi
- Barbosa, Fernando J.
- Barnes, Ted

Security:

- Public
- DocDBAdm
- Executive
- Writer

Figure 1: Form for adding a new document.

COLLABORATION MEETINGS

Much important information is discussed in collaboration meetings, therefore keeping an organized and easy to navigate method of storing the talks that are presented is necessary. The task of creating a new meeting in which the talks will be organized will be generally left to the administrator.

To create a new meeting first click on [Create or Change](#) from the DocDB home page, then under *Database Maintenance* click on Organize a [new](#) meeting. It is necessary to select the topic *Collaboration Meetings* and to fill in both the long and short descriptions. You should also check off 'Show All Talks' and fill in the Location, URL (if there is one), and indicate the *Start* and *End* dates for the meeting. For the Sessions you will need to fill in the Title, Location, Start Date and Time. Check the Break box if your entry is a break between sessions such as a coffee break, a meal, or other activity. These Break sessions cannot have talks or documents associated with them. If you run out of space to fill in more sessions simply click on [Modify Meeting and Sessions](#) and another will appear

at the bottom of the page. To finish click on [Modify Meeting and Sessions](#) once more in order to save the changes. Once this is completed you will be able to start adding the talks that belong to the various sessions.

The method for adding a talk to a collaboration meeting is much the same as adding a regular document to the database. The first difference is that you must select the collaboration meeting your talk belongs to under *Collaboration Meetings* as a topic. The second difference is that you must place your talk in the correct session. To find the right location for your talk click on [Collaboration Meetings](#) on the DocDB home page under *List Documents* then click on the meeting that your talk belongs to. In order to place your talk in a session click on [update the agenda](#) under the session title your talk will belong to. The form you will be filling out is shown in Figure 2. Then you can enter the document number and/or the Talk title. Under *Time* you will need to select the duration of the (and not the time of day) talk. Under *Author* and *Topic Hints* you can select the authors and topics that correspond to the document that you wish to add to the session. This helps to make a match after you click on [Modify Sessions and Talks](#) at the bottom of the page. If you get a Match this means that the search engine has found a document that corresponds to the criteria you entered. If the match is correct, check off confirm on the left hand side of the page, then click [Modify Sessions and Talks](#) once again on the bottom of the page to finish.

Modify a Session

Session: [Plenary Session IVb](#) begins 22 May 2004 at 11:00
 Location: IUCF - Meeting Room 2
 ([Upload a document](#) or [update the agenda](#) for this session)

Summary of sessions, actions and wrap up

Order, Confirm or Delete	Break	Doc. #	Talk Title & Note	Time	Author Hints	Topic Hints
1 <input checked="" type="checkbox"/> Confirm <input type="checkbox"/> Delete	No	164	Preparing for the GlueX Detector Review	00:30	Cedeno, Joe Chung, Suh-Urk Cuevas, R. Chris Cummings, John P. Denisov, S. Dennis, Larry Dickey, J. Doughy, Dave Doyle, Tony Dzierba, Alex R.	Collaboration:Minutes Collaboration Meetings:Sept. 2004 Collaboration Meeting Collaboration Meetings:May 2004 Collaboration Meeting Computing:Code Management Computing:Compilation and Linking Computing:DAQ Computing:Databases Computing:General Computing:Grid Computing:HDDM
Move to new session? ▾			Match: Preparing for the GlueX Detector Review			
2 <input type="checkbox"/> Confirm	<input type="checkbox"/> Break			00:30	Ingram, Carolyn A. Jastrzemski, Ed Jones, Richard T. Kaditz, Jeffrey Keller, Lew Kelle, James D. Klein, Andreas Klimenko, A. Kovaltchouk, Vitaly D. Krahn, Zoharin	Collaboration:Minutes Collaboration Meetings:Sept. 2004 Collaboration Meeting Collaboration Meetings:May 2004 Collaboration Meeting Computing:Code Management Computing:Compilation and Linking Computing:DAQ Computing:Databases Computing:General Computing:Grid Computing:HDDM
3 <input type="checkbox"/> Confirm	<input type="checkbox"/> Break			00:30	Adams, Gary Anumagalla, Ravi Barbosa, Fernando	Collaboration:Minutes Collaboration Meetings:Sept. 2004 Collaboration Meeting Collaboration Meetings:May 2004 Collaboration Meeting

Figure 2: Form for adding a Talk to a Session within a collaboration meeting.

ADMINISTRATIVE TASKS

The administrator is responsible for several tasks that include adding, modifying or deleting major and minor topics, groups of users, keywords and keyword groups, authors, institutions, document types, journals as well as meetings and conferences. If a user should find that any one of these things needs changing they should contact the administrator.

In order to administer the aforementioned things a person needs first to be logged in as 'docdbadm'. The administrator should then click Create or Change documents or metadata. Under *Administrator only functions* you will find the needed options. If you click on Administer authors, institutions, topics you will also find the options to administer document types, journals, and meetings and conferences. The Administrator should then choose which action they wish to perform (Modify, Delete, New) and fill in all necessary fields. Lastly the administrator password must be entered and the button beside it clicked.

DATABASE (MY SQL)

The document database utilizes the same MySQL ⁱⁱⁱ server as the GlueX portal to store information. The phpMyAdmin tool^{iv}, currently being used, allows for easy access to a visual version of the files and information contained in the database. One can easily create and drop databases, create/drop/alter tables, delete/edit/add fields, manage privileges, and export data into various formats. Figure 3 shows the Database table structure of the DocDB.

When a document is entered into the database the majority of information about it is stored in three tables: *Document*, *DocumentRevision* and *DocumentFile*. *Document* contains information on the requester and document type. This is because most of the information about a document can change and therefore is stored instead within *DocumentRevision*. *DocumentFile* stores the information about the files associated with each document. It includes a field for the file name, a field called *RootFile* which is used to distinguish between primary and supporting files within a document (whether Main is checked off or not), as well as a field for the file description.

Other important tables include *RevisionSecurity*, *RevisionTopic*, and *RevisionAuthor* which are all used to link the settings for security, topics, and authors for a document to the document. A document is associated with its minor topics, not its major topics. The table *GroupHierarchy* displays the parent/child relationship between the various security groups in which the parent has access to all documents that its child has, but not vice versa. For example, "writer" is defined as a "parent" of "reader" and therefore for documents classified under reader security the parent, writer, is also able to view. The child and parent ID's are each associated with a group ID listed in the *SecurityGroup* table.

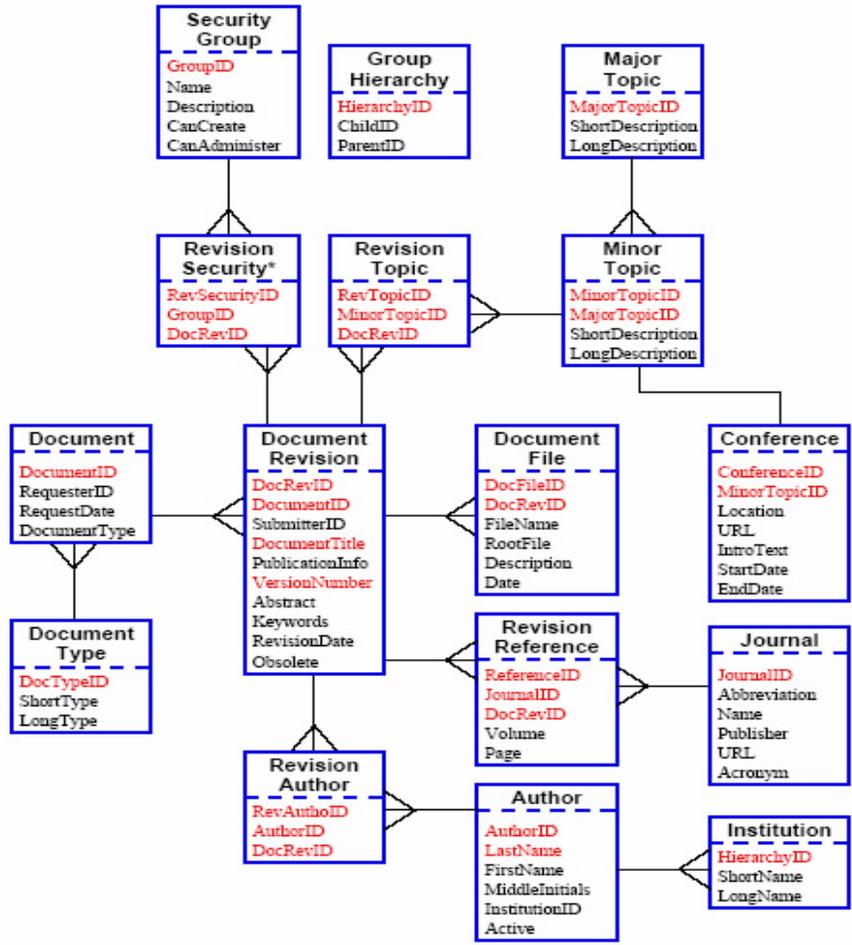


Figure 3: Database table structure of the DocDB. Fields in red are indexed for fast lookup.

SECURITY GROUPS

The security mechanism being used is based on http Basic Realm Authorization. The meta-information is protected by the Perl scripts, the actual files by `.htaccess` files. Within the DocDB we have Public and Private access. For public documents no username or password is required, but only documents designated as 'Public' are viewable. Within the Private portion of the DocDB there are five different groups.

- 1) *Reader*: This is a read-only-access account. This access can be granted to anyone who asks for it (presumably physicists) and can be used to view notes, talks etc, anything designated with 'Reader' or 'Public' access.
- 2) *Writer*: This is a read-write-access account. This access will only be granted to a Hall D Collaborator. As a writer one is able to also upload documents into the database and view any document designated as 'Writer', 'Reader', and 'Public'.

- 3) *Executive*. This access is restricted to the CB+EB group. As 'Executive' one is able to upload documents as well as view any document designated as 'Executive', 'Writer', 'Reader', and 'Public'.
- 4) *Review*. This access is restricted for project reviews by external reviewers. As 'Review' one is able to view any document designated as 'Review' and 'Public'.
- 5) *DocDBAdm*. This access is restricted to the administrator of the DocDB. As administrator one is able to view all documents and perform administrative duties such as adding institutions, topics, authors, etc.

In the future more groups can be added and the group hierarchy defined as needed.

BACKUPS

As the GlueX DocDB expands with the addition of more and more documents it is very important to keep backups of the database. Currently the database is backed up into SQL scripts manually. This is done by entering the MySQL database and exporting DocDB as a file. These SQL scripts plus all code for the portal and entire file flat system are backed up onto a second server hourly via the unix rsync command. Presently we are working on a method to have the databases backed up automatically and hope to have this in place in the near future.

ADDITIONAL FEATURES

[Set Preferences](#) (Email notification), and [Setup/manage a personal account](#) have not been activated yet.

CONCLUSIONS

The GlueX DocDB has proven to meet all our needs and provided the desired flexibility. We now are nearing 300 documents and have two collaboration meetings within the database. Although first populating the DocDB with all archive documents was a tedious task, the benefits we will reap from having the documents centralized and well organized under various categories was well worth the effort.

ⁱ Eric W. Vaandering, Vanderbilt University, Lynn Barren, Ferilab, *BTeV Document Database Design and Interface*, BTeV-doc-140, November 27, 2001, Revised: June 8, 2002

ⁱⁱ E. W. Vaandering, Vanderbilt University, *A Web Based Document Database*, March 2003

ⁱⁱⁱ <http://www.mysql.com>

^{iv} <http://www.phpMyAdmin.net>

^v E. W. Vaandering, Vanderbilt University, *A Web Based Document Database*, March 2003