Document Database Requirements

The requirements below are initial requirements for the near future to get us started and fulfill our most urgent needs. Others might be added as the DB evolves.

Types of Document

We want to store the following document types.

A. Reports (e.g. CDRs, TDRs, White Papers, NAS review etc.)

B. Journal papers (this includes papers from group member but also important papers from collaborators)

C. Conference Proceedings

D. Notes (informal documents with important information, e.g. kinematic, detector technologies, technical notes etc). This could be text files, spreadsheets, figures, etc

E. Conference/Workshop talks (these are mostly talks by group members but can also entail some important and relevant talks by others that we want to have readily available.

F.   Plots (more on this below)

For reports, papers, and proceedings it is important to be able to store not only the final report but the underlying source. In many cases it's the PDF plus the tex files and figures, or a Word file etc. In that sense the "document" is more than a simple file.

For talks the same applies. The original can be a PowerPoint, Keynote, or OpenOffice file and the final product a PDF file (easier for browsing). People might or might not decide to store both and the option should be available.

Note on conferences. Many documents (and types) are often related to one conference (talks, proceedings). It would be good to have all those that are related to a single conference link to an "Conference Event". It is not added above as (G) as it is not a document per se, but it would be good if that would be also searchable. For example, search for DIS2019 would lead to a "conference" entry with all f our talks and proceedings. Or a search for a presentation would allow to follow to the referring conference.

Formats

The most common formats available on macOS, Windows, Linux should be accepted.

For documents: all pdf formats, Word, Pages, tex, txt, OpenOffice etc.

For talks: PDF, PowerPoint, Keynote, OpenOffice, …

For figures: pdf, eps, png, gif, jpg, tiff, …

Note: some of them are rather rare these days (Open Office. Tiff etc) but it would be good to have this option just in case.

Access

There will be documents that are open to view and download for everyone but there will be also those that are of internal character and thus need to be protected. The latter should be a bit more refined.

1. Open documents (free to view and download for everyone)

2. Entry in DB visible to everyone but access is restricted.

3. Document is invisible to the outside

Why (3)? There might be documents that should be visible, but we would like to monitor/control access. That is, people can see (also in search) that there's an entry but before giving access we would like to check who wants to access it and then decide if we grant access or not. In category (2) are documents that concern the group alone and are (should not be) of anyone's interest but us.

At times documents will change from one category to the other, e.g. first intern and at some point later open and public. Moving a document from one access class to the other shuld be possible.

Keywords

All documents should be searchable using either file name, author (one of or combination of), date (range), title (parts), or keywords. The keywords should be defined when the document is added but it would be good to have always the ability to update and add keywords. May be a small abstract/readme for each document.

Versioning

All documents should be under version control, I.e. they can be replaced by newer version, but the older version should be still accessible. If possible there should be an ability to have different acces control for different version.

Plots & Figures

This is important as still many in the community use outdated (speak wrong) plots. Often plots are hard to find and at times people do not even know where the plot is from.

Plots can be entered as either:

• Individual stand-alone plots

• Plots that are part of a document, paper, proceeding, note, or report

The latter would require that for example if a document with corresponding tex and figures is stored that the figures are automatically entered/linked in the plot "area".

Each figure will need versioning, keyword search, possibly a note/readme, and the link/pointer to the source (document, paper, …) It would be absolutely helpful if there would be a thumbnail viewer. In a search that ends with 30 plots as a result a viewer would allow to see the right one in no time, otherwise one has to open every one to find the correct one. If it's a stand-alone plot the option for a accompanying readme plot is needed so one can give background info, explain what the plots shows in detail, cuts etc.

User Interface

Needless to say, that a good user interface would be incredibly helpful. Inspire and CERN's CDS are good example.

Storage/Size Limitations

While the average document is not too large there are occasions where talks and documents have many large figures (reports, large papers). It would be therefore desirable to have no limit on the size for individual documents.

———————————