Guidelines for Updating Hall Data Management Plans
Modified: November 22, 2024

It has been noted by both internal and external Reviews that JLab's Data Management Plans have fallen out of step with current and upcoming guidelines. The current Hall DMPs may be found at https://data.jlab.org. That site also includes a recently updated Data Management Plan Guidance document that can provide direction updating your Halls DMPs.

For this initial stage, we are requesting that you simply revise your existing Hall DMPs and bring the information in them up to date. Where possible, we ask that you add URLs, pathnames, or other directions to resources listed in the document as sub-bullets under the relevant category. There should be enough information for us (you, me, SciComp, IT) to locate them and assess backup/snapshot status.

A more comprehensive rewrite of JLab DMP policy will occur in 2025, with a completion goal of October 1, 2025. Your work on the present plans will be critical to that rewrite and will allow us to assess what tools and support software the Lab needs to meet our DMP requirements.

Please review your Hall's current DMP and provide an updated version following these guidelines:

- Author, Contacts, and last reviewed lines
- Raw / Processed Data
  - Please indicate where the data are stored (e.g. /mss/hallX/…) including any standard conventions baked into pathnames that would help researchers locate and identify classes of files for a run-group or experiment.
  - This does not need to be exhaustive.
    - An example for Hall C may be along these lines:
      - raw data: /mss/hallc/<experiment_id>/raw/
      - processed data: /mss/hallc/<experiment_id>/analysis/
    - An example for Hall D may describe their standard practices:
      - /mss/halld/
        - RunPeriod-2022-08/
          - calib/   - <brief description>
          - production/ - <brief description>
          - rawdata,   - <brief description>
          - recon/   - <brief description>

- Run Conditions / Databases

- ○ Please provide a list of databases critical to analysis:
  - ▪ This may include instances of RunDB, CCDB, and/or other more customized DBs as well as more centrally managed databases such as MYA (Accelerator EPICS archiver system).
  - ▪ For each, provide a description of where they are located (URL, hostname, git repo, CUE pathname, etc).
- Logbooks
  - ○ If collaborations within your Hall use logbooks other than those at logbooks.jlab.org for analysis and offline documentation, please provide links or directions to those resources.
- Analysis software source code and build systems
  - ○ Please provide URLs that indicate where/how critical software may be obtained. This should include analysis software and simulation codes that individual Collaborations/Experiments may fork and modify.
- Documentation
  - ○ Please provide URLs to documentation for standard Hall software (analysis and simulation) where possible. This can be a link to a top-level 'Documentation' page on a wiki or Hall website.

Historical Data Management Guidance (from 2013)

Graham Heyes, Brad Sawatzky

The material below was provided to the authors of the existing Hall DMPs in 2013. I'm including it only to provide context for those tasked with updating those documents. The new 2024 Guidance document supersedes these historic instructions.

Title and summary: The document title states that this is a plan for a hall or collaboration. Authors should decide which one it is stick to that throughout the document to avoid confusing the reader, for example is this the hall-D plan or the GLUEX plan? Similarly, who gets contacted if there are questions? Is that the author of the document or someone else? Do we need one name or two?

Responsibilities: I think this is fairly generic and probably doesn't need to be tailored to each hall. Again, we should decide if we delete either the word hall or collaboration to make it clear who is responsible for the data.

Experimental Nuclear Physics Data Management processes: If there is some form of data that I missed feel free to add it or to elaborate, for example I don't mention simulation output but that may be important.

- **Raw Data:** As far as I know all the halls do the same thing here so my text may be enough.

- **Processed Data:** Hall specific clarification of what process data is kept and where it is kept is good.

- **Run Conditions:** Same as above, if there is anything done that is special, please write about it.

- **Databases:** If your data bases are NOT managed by IT then they should be. If they are not then we need to know why, how the contents are managed and archived and by who.

  - **Log Books:** We are aiming for a common logbook in the 12 GeV era, but nobody has said that is mandatory (yet). As far as I know everyone is now using electronic logbooks. If they are backed up by IT, then that's probably all that we need to write here.

  - **Calibration and Geometry databases:** You know a lot more about this than I do so write about it please.

  - **Other databases:** Same as above.

  - **Analysis software source code and build systems:** The software support committee is working towards maintenance of common third-party packages so we can use that

to take care of most of the stuff we didn't write ourselves. At the moment each hall is doing something with code versioning, CVS, SVN or GIT. Please write about which one it is, who maintains it and how it is backed up. I think many people are moving to GIT but we can update these documents when that happens.

- **Documentation:** Write about your documentation system, give examples, also indicate how it is maintained, for example the web servers are maintained by IT but the content management system is maintained by someone in your collaboration/group.

- **Quality Assurance:** Not sure what else to say here, the IT plan says that overall oversight is from the Deputy Director for science who can and will call reviews. That is probably enough!