

# Microboone Report

FIFE Workshop  
June 4, 2013

H. Greenlee

# Outline

- GPCF facilities.
  - Interactive.
  - Batch.
  - Bluearc disks.
- Data Handling (sam/ifdh/fts/enstore).
- Offline framework.
  - art
  - larsoft
- Databases.
- Online.

# Interactive Computing

- Microboone currently has 3 interactive uboonegpvmXX machines.
  - Seems adequate for the moment.
- There is currently no easy way to install something like a “microboone computing environment” (include larsoft software releases, art, other ups products, etc.) on non-gpcf machines, including remote sites.
  - Some non-Fermilab sites have maintained larsoft releases, including art suite, etc., with considerable effort.
  - We would like for there to be a better way to do that.
    - Cvmfs?

# Batch Computing

- Batch system (farms, condor, jobsub\_tools) mostly just works (thanks SCD).
  - There are a few known problems and issues.
    - Authentication system (cron job on gpsn01) is clumsy, should be improved.
      - Ownership of files returned from batch jobs is often not correct.
    - Better access to opportunistic resources (including outside Fermilab).
      - Cvmfs.
  - Large memory slots will be needed (more for lbne than microboone).
  - We know that we will want sam support in jobsub\_tools (e.g. for starting an stopping sam projects).
    - Maybe this already exists (not sure)?

# Bluearc Issues

- Microboone currently has 40 Tb (/uboone/data) + 2 Tb (/uboone/app).
  - We have had several instances where either of the above disks have filled up.
  - To help prevent this, we are adopting a quota system.
  - We are moving toward using enstore (progress has been slow).
    - This may never be much used by “ordinary users,” though.
    - Production files are only a small fraction of total disk usage (so far).
  - Better disk management tools would help.
  - There must also be continual growth of total disk size.
- As everyone knows, bluearc’s fatal flaw is that it has inadequate ability to prioritize or partition bandwidth.

# Data Handling – SAM/ FTS/ Enstore

- Microboone currently has a sam database, sam server and sam station.
- Sam database is populated with metadata from recent MC generation, including dataset definitions and locations on bluearc.
  - Files are waiting for some time to be uploaded via fts to enstore.
  - Fts and enstore configuration designs are available (microboone docdb 2414 and 2494), which are waiting to be implemented by SCD.
  - We want FTS and enstore designs implemented and files uploaded asap.
- As yet, no one has yet tried to start a sam project or fetch files from sam.
  - Should be possible to do using bluearc locations.

# Data Handling - IFDH

- Microboone has updated its production scripts to use “ifdh cp” instead of “cpn.”
  - I believe “ifdh cp” is still using “cpn” under the covers.
- We are intending to integrate sam into our art framework, which will involve using ifdh\_art services (IFFileCatalog and IFFileTransfer).
  - Haven’t tried it yet.
  - Ifdh\_art ups product is not yet integrated into art or larsoft setups.
- There is a lot of overlap between sam\_web\_client and ifdhc sam clients. Sometimes not clear which one should be used.
  - I believe (someone correct me if wrong) that sam\_web\_client has more features, but ifdhc includes a promise to be grid-friendly.

# Art Framework

- Microboone (and larsoft) are using art for our offline framework and data persistence mechanism.
  - I like art. It is well-designed. It has lots of useful features.
  - Art has a steep learning curve. Documentation is not that great.
    - In larsoft, we direct people to the larsoft wiki to learn about art.
    - I find myself resorting to reading art source code rather often.
    - My preference is not for more tutorials or examples (we have those), but rather for a reference manual documenting, e.g., art-supplied modules, services, and public classes.



# Art and Sam

- Reading art files from sam via art has been mostly farmed out to ifdh (ups product ifdh\_art).
  - Microboone hasn't experimented with this much. We're still struggling to get files into sam.
- On the other hand, generating sam metadata is handled completely internally by art via FileCatalogMetadata service and RootOutput module.
  - Art still lacking some basic functionality.
    - Inability to generate all standard metadata (e.g. file parentage).
    - Inability to generate unique output file names.
    - No support for per-file experiment-specific parameters.

# Art and Sam (cont.)

- I have been requesting changes on behalf of microboone at art stakeholders meetings, and via the art issue tracker. So far, none of our requested changes have been implemented.
  - Reasons given for delay (both valid):
    - Need to consult with other art stakeholders.
    - Inadequate manpower.
  - SCD can't do much to speed up consultation process. SCD could, in principle, devote more manpower to art development.

# Microboone on Art Issue Tracker

cet-is » art

Search:

[Overview](#) [Activity](#) [Roadmap](#) **[Issues](#)** [New issue](#) [Calendar](#) [News](#) [Documents](#) [Wiki](#) [Repository](#)

## Issues

▼ Filters

Status

open ▾

Add filter

Author

is ▾

Herbert Greenlee ▾

► Options

Apply

<input checked="" type="checkbox"/>	#	Tracker	Status	Priority	Subject	Assignee	Updated
<input type="checkbox"/>	3962	Bug	New	Normal	FileCatalogMetadata service improvements.		05/30/2013 09:45 am
<input type="checkbox"/>	3946	Feature	New	Normal	ActivityRegistry callbacks preCloseFile and postCloseFile should pass the file name as an argument.		05/28/2013 10:34 am
<input type="checkbox"/>	3945	Feature	New	Normal	ActivityRegistry should have callbacks for when output files are opened and closed.		05/28/2013 10:28 am
<input type="checkbox"/>	3906	Bug	New	High	Main parameter set not saved in RootFileDB		05/17/2013 09:57 pm
<input type="checkbox"/>	3776	Feature	New	Low	Script to generate service skeleton		04/26/2013 11:46 am
<input type="checkbox"/>	3774	Feature	New	Normal	Improved ability to generate output file names.		04/26/2013 11:16 am
<input type="checkbox"/>	3773	Feature	New	Normal	Improved ability to control when RootOutput will close and open output files.		04/26/2013 11:05 am
<input type="checkbox"/>	3746	Feature	New	Normal	Art should support generating per-file user-specified sam metadata.		04/23/2013 11:41 am
<input type="checkbox"/>	3745	Feature	New	Low	Art should not require metadata that sam considers to be optional or non-existent.		04/23/2013 11:41 am
<input type="checkbox"/>	3744	Feature	New	Normal	Missing sam metadata		04/23/2013 11:23 am
<input type="checkbox"/>	3743	Bug	New	Normal	FileCatalogMetadata service doesn't set file format.		04/23/2013 10:54 am

(1-11/11) | Per page: 25, 50, 100, 5000

# Larsoft

- Larsoft provides an offline framework and a basic set of data products and algorithms for liquid argon tpc experiments (argoneut, microboone, lbne).
  - Larsoft is also a great resource for test stand and test beam LAr tpcs (long bo, lariat, 35T).
- Larsoft leadership is transitioning from Brian Rebel and Eric Church to Rick Snider of SCD.
  - I think it is inevitable that algorithm development will tend to diverge somewhat for different experiments.
  - I also think it will be very useful to maintain a common code base and a common meeting to allow cross-fertilization.
  - Working out relations among different experiments/groups will be the hardest part of the transition.

# Databases

- Microboone currently has postgres databases up and running for offline and online.
- We are just starting to develop databases. We know that we will want to do the following things, which may need help from SCD (and CCD?).
  - Database design.
  - Database applications.
    - We will want line mode, python, and c++ database clients.
    - We already have a larsoft art service DatabaseUtil (by A. Szelc) for accessing databases from art programs.
  - Database servers.

# Online

- Online and daq mostly using same tools as offline.
  - Raw data will be converted to art format.
  - Raw data will be uploaded to sam/enstore.
  - Online system will have its own postgres database instance.
  - Online system will have its own larsoft software release.
  - Online system will have its own separate code base and repository.
- Online sys admins supplied by SCD/FEF.
- Obviously, we need all of above supported for high availability.

# Summary

- GPCF.
  - Things are generally working well. Some requests for improvements are given in this talk (not critical).
- Data handling.
  - We have made lot of progress with respect to defining sam metadata.
  - Need help and action to get FTS and enstore configured, and files uploaded. We want this soon.
  - Art needs to be updated to properly support sam (need action).
- Larsoft.
  - Still working out how larsoft will be organized going forward.
- Databases.
  - It's still early days. We need to make progress soon.