Review of Software and Computing

Maxim Potekhin   Brookhaven National Laboratory

potekhin@bnl.gov

Goals of this meeting

▸ The LBNE software coordinators are requested to form the LBNE and FNAL-SCD Joint Assessment of Software/Computing needs for the Long Baseline experimental program.

▸ FNAL-SCD and the LBNE collaboration will form a standing team to advise the LBNE spokespeople and FNAL SCD regarding computing plan, software tools, and man power resources needed in the short and long term.

▸ The LBNE software team is requested to provide a summary presentation of the computing plan presented during the May 13-15 DOE software/computing review. The summary should be followed by a description of the team and the completed tasks since May.

▸ The outcome of the first meeting should be a joint very brief report outlining the needs in bulleted form for the 35 ton, requested by December 25. The brief report should address the process for resolving longer term issues, and place the short term requests in the context of these long term issues.

▸ *My quick summary of the goals: to define the context, status and priorities of our work.*

Overview

▸ Interim org chart and structure of the effort

▸ Events and Milestones of the past 18 months

▸ Documentation: Computing Plan and Requirements

▸ Key areas where progress was made – in no particular order:

    ▸ Core software infrastructure – progress made with improved software build procedures

    ▸ Improved system for geometry description

    ▸ Software Release Management

    ▸ Distributed computing (CVMFS+Grid)

    ▸ Initial quantitative parameters for the Computing Model

▸ 35t readiness (physics tools) will be covered separately (Tom, Michelle).

▸ Coordination with the DAQ/Online team.

▸

Software and Computing/ Physics Tools (Fall'14)
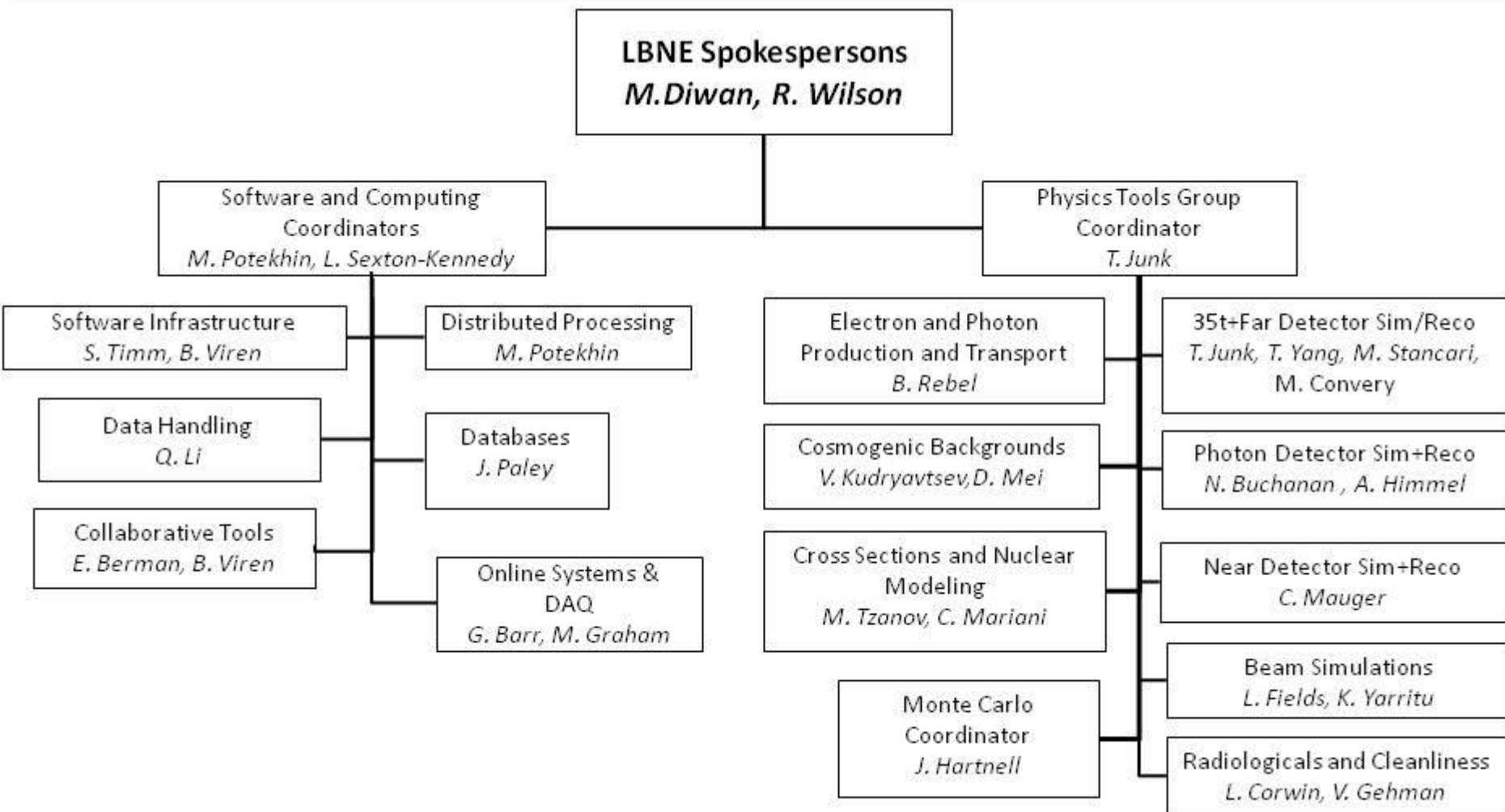
Events and Milestones

▸ Computing Plan ("roadmap") created in 2013 (DocDB 7818) and presented to DOE (briefing in September 2013).

▸ The S&C/PT Requirements were developed with broad participation of the Collaboration and accepted by LBNE in early 2014 (DocDB 8546).

▸ Workshop at ANL was conducted in late March 2014 to prepare for the DOE review according to the charge (R&D + Computing).

▸ DOE Review in May 2014. Initial feedback (largely positive) was available immediately, final results received last month – we are now working on our official response as requested.

▸ P5 Report and events triggered by it.

▸ Formation of the US HEP Forum for Computational Excellence

▸ Formation of the HEP Software Foundation (CERN)

▸ "Open Meetings" at CERN and FNAL, December 2014 ("ELBNF")

The Computing Plan Summary

▸ DocDB 7818

▸ List of "technology areas" (next slide)

▸ Software and Computing Organization (S&C) and the Physics Tools Group (PT) as defined two components of the LBNE Computing effort (are previously illustrated in the org chart).

▸ S&C seen as providing software infrastructure services to PT.

▸ The Plan serves as a useful reference describing the "landscape" of LBNE computing effort in all its aspects (S&C and PT). In addition to S&C (more or less infrastructure related), the PT part itemizes major science software deliverables.

The Plan:
List of S&C Technology Areas

The following list contains items corresponding to varying volumes of effort and importance and is meant to be inclusive.

▸ Software Frameworks

▸ Distributed Code Management

▸ Build, Testing, Validation and Distribution

▸ Simulation Tools

▸ Event Display

▸ Geometry Model and Description

▸ Databases

▸ Messaging

▸ Data Storage, Access and Management

▸ Grid Tools and Distributed Computing

▸ Workload and Workflow Management

▸ Networks

▸ Information Services and Web Frameworks

▸ Security

▸ Collaborative Tools

Summary of Actionable Items Presented in DOE briefing (2013)

▸ A short list has been created. Some of the work areas have been combined in order to reflect commonality and/or take advantage of synergies between certain projects. This is the list in the order of significance, as we currently see it:

1. Build/Testing/Distribution of the code
2. Grid Tools/Workload Management
3. Data Storage and Management, Meta-Data infrastructure
4. Geometry Model/Event Display

▸ *Notes:* item 2 has been completed as a technology demonstrator using OSG resources, CVMFS software provisioning and augmented "jobsub" at FNAL, for now it's off the top priority list.

The DOE Review (May 2014)

Full scale official review of LBNE computing with formal charge letter.

▸ Two components: R&D and Computing, with emphasis on the latter.

▸ S&C and PT presented separately.

▸ S&C part contained a status report and a summary of the Requirements document.

▸ The S&C Requirements which we brought to the review were developed carefully and with lots of consultation within the Collaboration over a period of a few months, and represented a high-level view of what LBNE needs from its software and computing.

▸ Progress report and status were presented (we'll cover this later in this presentation).

▸ Assessment of the LBNE computing effort was positive, in terms of its structure, management, attention to detail and in-depth understanding of issues.

S&C Mission Statement at the Review

The goal of the LBNE Software and Computing Organization is to make all aspects of computing technology, software tools and resources easily accessible to the largest number of Collaboration members possible, in a most efficient manner.

"Our Vision"

- In order to fulfill the mission of the Software and Computing Organization, we plan to do the following

  - Create a robust software infrastructure in which the barriers for the collaborators to join scientific studies and the R&D efforts are greatly reduced.

  - Provide solid support of scientific endeavor of LBNE by adequate QA, validation of its software, proper software process and administration, thus enhancing reliability and accuracy of its science tools.

  - Create opportunities for both the US and international LBNE participants to contribute to its core computing infrastructure by implementing systems for resource sharing and federation.

  - Maximize computing resources available to the Collaboration and increase their robustness by implementing a fully distributed system which combines a network of computing centers and data services combined with transparent monitoring and agile data and workload distribution mechanisms.

  - Foster cooperation and shorten learning curves by maintaining and evolving the Collaborative Tools.

- The Software and Computing Requirements are designed to promote this vision and act on it.

Purpose, origin and scope of the S&C Requirements

These Requirements are developed based on "Physics Tools and Software and Computing Organization of the Long Baseline Neutrino Experiment: the goals, the structure and the plan", FNAL DocDB 7818. They serve as a step towards the LBNE Computing Model, which will be considered a normative basis for policies to be implemented by the LBNE S&C Organization, in order to evolve the LBNE computing platform in the direction optimal for achieving the scientific goals of the experiment.

An effort was made to maintain a relatively high-level view of the S&C issues, and to not go into smaller details which are more likely to change as the project moves forward. In cases where it was impossible to establish concrete metrics or parameters for a specific requirement, it is still listed as an item the S&C Organization will need to address in the future.

The Requirements for the Online/DAQ work area are not included in this document and will be created separately. At the same time, we define principles used in establishing interface with the DAQ group and procedures for handling data produced by systems under its management.

What's in the Requirements

Section titles from the document:

▸ Software and Computing Organization of LBNE

▸ Data (storage, handling, distribution, access etc).

▸ Databases

▸ Software

▸ Distributed Computing

▸ Geometry

▸ Visualization

▸ Networks

▸ Collaborative Tools

▸ Cybersecurity

The DOE Review Recommendations

We received the final and official document in November'14. Summary:

▶ *Develop an initial computing model, based on quantitative assessments of the LBNE requirements, identifying all necessary tasks and tools, prioritizing efforts as well as identifying gaps in their current delivery plans and highlighting opportunities for new collaborators. This model should be completed before the next review.*

▶ *Establish a recurring review process to update computing requirements and monitor evolving hardware architectures and external software developments.*

▶ *Evaluate a project management and financial structure for the computing effort in LBNE, such as used by U.S. ATLAS / U.S. CMS that would be able to flexibly adapt to changing requirements, optimize resource usage across the collaboration and ensure on time delivery of critical capabilities.*

▶ *As the highest priority for S&C, produce the computing model document and direct resources as appropriate.*

▶ *Direct more scientific efforts towards the 35t prototype.*

Post-Review

▸ We are asked to provide a response to the review report, which is being done now.

▸ There was a recommendation to follow the LHC computing model.

▸ Since the review itself, we've been working to address the issues already outlined in the preliminary report:

  ▸ We started work to improve our quantitative understanding of the data rates, volumes and other parameters, in order to feed the Computing Model. Initial version of a "spreadsheet model" was created to help estimate data parameters in different scenarios.

  ▸ According to the report, the timeline for the Computing Model to be created is the *next review cycle*. This is TBD (see next slides).

  ▸ The 35t readiness item in the report (and all others) were acted upon during the months after the review (35t topics to be presented by Tom and Michelle in this meeting).

Other events impacting LBNE Software and Computing

- Challenge: ELBNF(?)
  - New international Collaboration is being formed *even as we speak*.
  - Open Forum at FNAL on Friday, an earlier one on 12/5/2014 at CERN.
  - Impact on computing is yet to be understood, too many unknowns at this point including the basic Far Detector configuration.
  - …however what is known is this – we'll need to **make extra effort to ensure portability and ease of use of the software components** that LBNE (in cooperation with FNAL SCD) will offer to the new Collaboration. This will be instrumental for the standing of the US research community in ELBNF.

- Positive: US HEP Forum for Computational Excellence (FCE)
  - "Facilitate collaboration by increasing awareness of the activities and resources within our field, promote awareness and use of open source tools and approaches within the HEP software and computing community etc…"
  - LBNE has formal representation on FCE (B.Viren as a group co-convener, M.Potekhin as one of the contributors)

- Positive: HEP Software Foundation (CERN) http://hepsoftwarefoundation.org/
  - LBNE is represented on this forum as well (B.Viren)

Summary of progress in 2014

▸ Priority "short list" remained fairly constant over time

- ▸ Core Software Infrastructure (build/testing/validation/distribution).
- ▸ Data Storage and Management
- ▸ Geometry Description and Model
- ▸ Distributed Computing

▸ Progress has been achieved in all of these areas and we'll provide a few detail in the following slides.

▸ In addition, we introduced Mediawiki as a promising collaborative tool and are working to collect and maintain documentation in this and other media, since we found pre-existing situation unsatisfactory (basically fragmented and hard to navigate information). The "landing page" for LBNE S&C is now in Mediawiki.

Software Infrastructure

- Build/validation/continuous integration
  - Issues encountered in 2013-2014 are outlined in the "white paper".
  - This has been top priority consistently over a year and a half.
  - This needs to be resolved for many reasons and becomes even more important in light of the new Collaboration begin formed – we want the existing LBNE software stack to remain relevant and to be used and appreciated by a large community of collaborators.
  - Methods to improve build procedures and tools have been identified.
  - The "*worch*" tool which helps orchestrate the build process (including CMake-based) has been developed and tested (Brett).
  - This is Work In Progress, ETA~O(weeks).
- Release Management
  - David Adams (BNL) has accepted the role of the LBNE release manager, the scope, policies and procedure are being worked out.
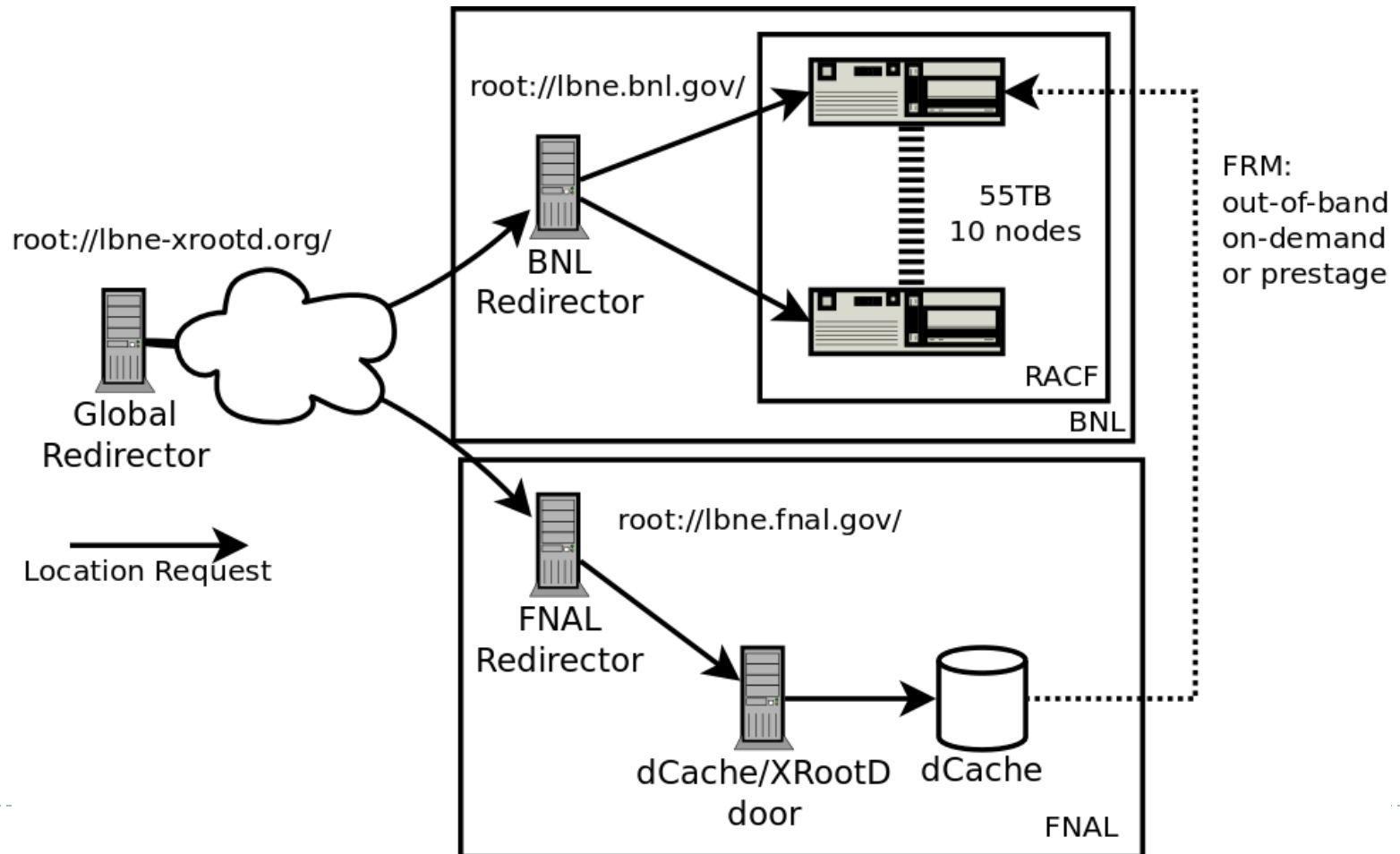
Data

- ▸ **Evolution of the LHC Computing Models**
  - ▸ In 2013-2014, there is a trend to abandon the MONARC architecture and flatten the hierarchy of data and workload distribution. We started configuring and testing xrootd.
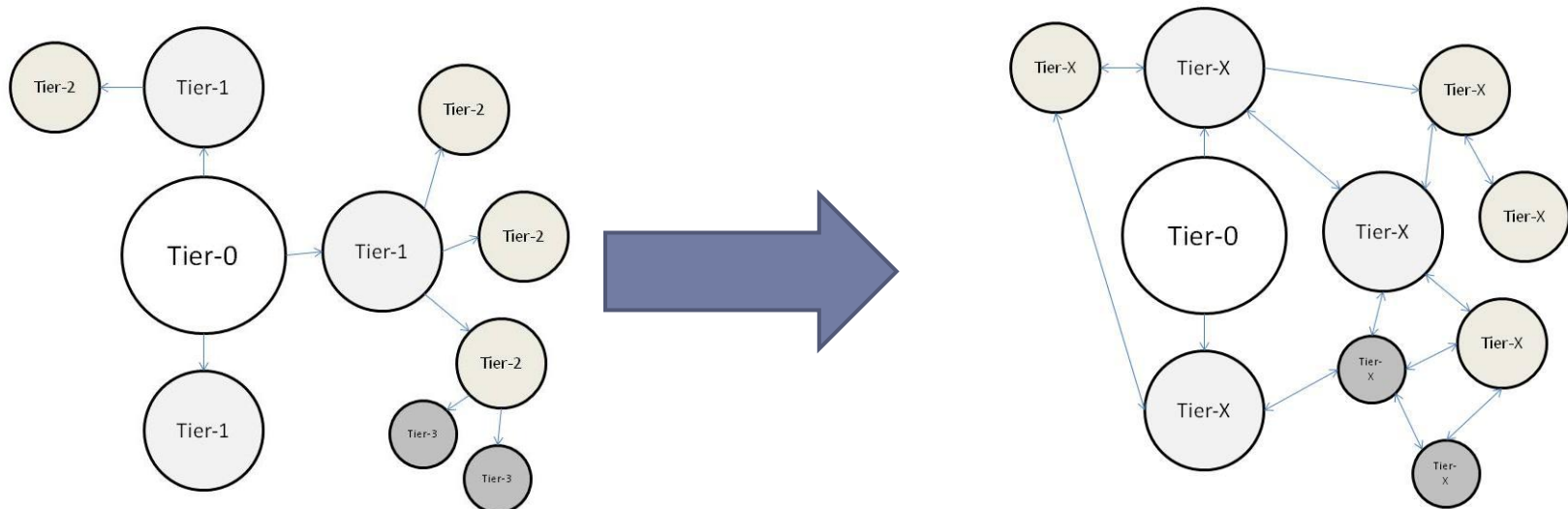
xrootd

- Conceptual diagram of xrootd@BNL and links to FNAL

Grid Computing

▸ No perceived shortage of "central" resources at FNAL, but Grid capability is a must have – people from other institutions will be able to do more if we do.

▸ We received **excellent** help from OSG

  ▸ Software provisioning via CVMFS

  ▸ Ran LBNE payloads on ~20 Grid sites

▸ Declared a success, after completing a few runs on the Grid over a period of 2 months we decided to suspend this effort due to other priorities.

Geometry

- Brief history
  - Identified as a priority item a while ago, not much done in this area until recently due to lack of manpower.
  - Came to the fore due to issues with ongoing FD simulation work.
  - Current solution (a mix of Perl and XML) is not optimally structured and can benefit from factorization – see Brett's talk in the parallel session.
  - While we (Maxim and Brett) had experience with a more "holistic" approach (pure XML being the only source) it appears that from expediency point of view, combination of "builders" and "configuration" could deliver most benefits relatively quickly.
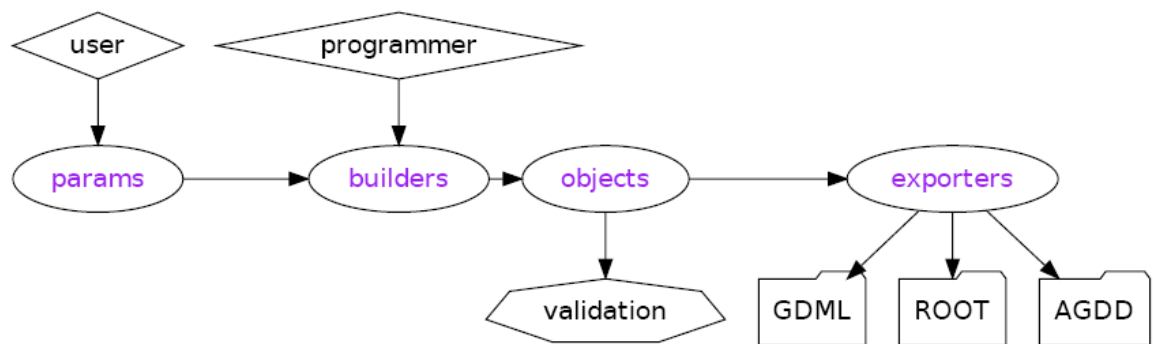
GGD

- Goals for general geo description (GGD)
  - Authoring
  - Persistence
  - Provenance
  - Visualization
  - Validation
  - Conversion
- Components of GGD
  - Parameters
  - Builders
  - Objects
  - Exporters
- See https://github.com/brettviren/gegede

Metadata

- Metadata
  - Always mentioned in the Requirements and plans but usually without reference to a concrete implementation which we'll need rather soon.
  - SAM was the prime candidate and we are now confident it will satisfy the requirements in near and mid-term.
  - Discussions under way with the DAQ team regarding Metadata for the 35t prototype run.
  - Setup for 35t close to complete (Q.Li)
- Area needs more attention in the long term. How do we use xrootd etc optimally in a way complementary to managed replicas etc?

Summary

▸ "Stay focused" and keep following our plans and vision as LBNE undergoes reformulation. There are deliverables that need to be completed within a year or less.

▸ It is not productive to focus on a detailed computing model right now since there is much uncertainty regarding the detector design and the new collaboration itself.

▸ We think that in following the LHC model and technology portfolio (as we did from the beginning) we are doing the right thing.

SCD&LBNE

- We need SCD support in facilitating use of the improved build procedures in and for LBNE

  - This was important even before reorganization, due to a potentially large and geographically dispersed organization – we want a reasonable build environment for developers on multiple sites. More important now due to more international representation in the project.

- Metadata (both conceptually and applied to technologies outside of SAM) is a difficult subject and if there is a source of expertise in SCD (a data scientist) that would be helpful.

- There are mostly technical issues e.g. design of data structures in art and larsoft where SCD could provide crucial assistance.