



Contribution ID: 405

Type: **Poster presentation**

## Automatic Tools for Enhancing the Collaborative Experience in Large Projects

*Monday, 14 October 2013 15:00 (45 minutes)*

With the explosion of big data in many fields, the efficient management of knowledge about all aspects of the data analysis gains in importance. A key feature of collaboration in large scale projects is keeping a log of what and how is being done - for private use and reuse and for sharing selected parts with collaborators and peers, often distributed geographically on an increasingly global scale. Even better if this log is automatic, created on the fly while a scientist or software developer is working in a habitual way, without the need for extra efforts. This saves human time and enables a team to do more with the same resources. The CODESH - Collaborative DEvelopment SHell - and CAVES - Collaborative Analysis Versioning Environment System projects address this problem in a novel way. They build on the concepts of virtual states and transitions to enhance the collaborative experience by providing automatic persistent virtual logbooks. CAVES is designed for sessions of distributed data analysis using the popular ROOT framework, while CODESH generalizes the same approach for any type of work on the command line in typical UNIX shells like bash or tcsh. Repositories of sessions can be configured dynamically to record and make available the knowledge accumulated in the course of a scientific or software endeavor. Access can be controlled to define logbooks of private sessions or sessions shared within or between collaborating groups. A typical use case is building working scalable systems for analysis of Petascale volumes of data as encountered in the LHC experiments. Our approach is general enough to find applications in many fields.

### Summary

**Primary authors:** Dr BOURILKOV, Dimitri (University of Florida (US)); Dr RODRIGUEZ, Jorge Luis (UNIVERSITY OF FLORIDA)

**Presenter:** Dr RODRIGUEZ, Jorge Luis (UNIVERSITY OF FLORIDA)

**Session Classification:** Poster presentations

**Track Classification:** Facilities, Production Infrastructures, Networking and Collaborative Tools