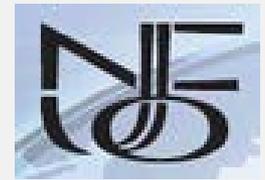


Discovery of Publications for the OLCF User Facility

Panel Facilitator

Panel participants

April XX, 2015



***NUFO 2015
Annual Meeting***

***Oak Ridge
National Laboratory***

April 21–23, 2015

What is the Oak Ridge Leadership Computing Facility?

- Collaborative, multi-lab, DOE/SC initiative ranked top domestic priority in *Facilities for the Future of Science: A Twenty-Year Outlook*.
- Mission: Provide the computational and data science resources required to solve the most important scientific & engineering problems in the world.
- Highly competitive user allocation program (INCITE, ALCC).
- Projects receive 100x more hours than at other generally available centers.
- LCF centers partner with users to enable science & engineering breakthroughs (Liaisons, Catalysts).



Titan

Cray XK7

Currently #2 HPC Top 500

27 petaflops

299,008 Processor Cores

18,688 GPUs



ASCR Facilities Division Publication Guidance for the Leadership Computing Facilities

The Facility tracks and reports the number of refereed publications written annually based on using (at least in part) the Facility's resources. For the LCFs, tracking is done for a period of five years following the project's use of the Facility. This number may include publications in press or accepted, but not submitted or in preparation.

To respond to this guidance we need an effective and efficient process to discover, curate, and report publications. What we have found is that no single collection method, if implemented in isolation, is sufficient to the task.



OLCF User

OLCF Users

- Average about 1,000 users per year
- Users access our facility from around the world
- Access period to use allocated resources may span months to years
- Users represent multiple science domains
- Users are from industry, academia, and government facilities

Publication Challenges

- Publications often occur long after the user's period of access ends
- OLCF users typically utilize other computing facilities
- User-reported lists of publications are error prone, and/or users do not always use proper citation
- Automated searches can be no more complete than the completeness of the database searched
- Due to diversity of users, there are numerous places to search

Common challenge for National Laboratories
and Academic Institutions



Previous Procedure

- Self-reports
 - Reports are periodic and lag behind the publications
 - Do not have a good mechanism to force users to provide reports after they leave
 - Users report erroneous results
 - Reports come from the PI and doesn't always capture all of the publications from the users on the project
- Web of Science search
 - WOS search conducted once a year using OLCF user list and keywords
 - Search took weeks to complete
 - Did not have access to all journals/proceedings in which OLCF users publish
 - Users do not always properly cite the user facility and the search misses OLCF publications
 - Search missed publications for the CY that get uploaded to WOS after the annual search is conducted

This process was manual in nature, required many staff hours to complete, and was only done once a year.



Operational Improvements

- Tools and procedures developed within ORNL research have advanced ability to track and report publication outcomes:
 - Discovery across multiple years
 - Ability to refresh information more frequently
 - Find publications not reported through other mechanisms, e.g., users' quarterly reports
- Implemented process of discovery integrating automated database searches with user-reported publications creating a validated repository of publications for institutional use
 - Well defined process provides high data quality
- Validated results can be communicated with confidence
 - Meeting reporting requirements of sponsor
 - More readily communicate outcomes to the full spectrum of stakeholders



Publications Discovery

- Three collection methods:
 - Automated Publication Discovery (ADP)
 - Self-Reports (SR) from user reporting to OLCF
 - PTS (Publication Tracking Service) for ORNL staff.
- Process:
 - Text Analytics: Intelligent Search for Key Words (e.g., User Name, OLCF, Titan and/or INCITE Acknowledgement, etc.), and
 - Validated and Curated Results (Manual integration with SR and PTS).
- We are refining process and tools to improve effectiveness:
 - Leverage automated method as primary discovery method
 - Minimize staff effort recognizing support will always be needed
 - Improving quality of self reports; terminating less productive manual library searches

