Networking and Information Technology
Research and Development

***The government seeks individual input; attendees/participants may provide individual advice only*.**

Middleware and Grid Interagency Coordination (MAGIC) Meeting Minutes[1]
February 5, 2020, 12-2 pm ET
NCO, 490 L'Enfant Plaza, Ste. 8001
Washington, D.C. 20024

Participants (*In-Person Participants)

Rachana Ananthakrishnan (UChicago)	Glenn Lockwood (LBL)
Nick Balthaser (LBL)	Don Petravick (IUIC)
Richard Carlson (DOE/SC	Lavanya Ramakrishnan (LBL)
Dhruva Chakravorty (TAMU)	Birali Runesha (UChicago)
Kjiersten Fagnan (LBL)	Matt Selmici (UW Madison)
Sharon Broude Geva (UMich)	Alan Sill (TTU)
Devarshi Ghoshal (LBL)	Nathan Tallent (PNNL)
Wayne Hurlbert (LBL)	Kevin Thompson (NSF)
Margaret Johnson (IUIC)	Sudarshan Vazhkudai (ORNL)
Joyce Lee (NCO)*	Sean Wilkinson (ORNL)

Proceedings

This meeting was chaired by Richard Carlson (DOE/SC). December 2019 meeting minutes were approved.

MAGIC looks at middleware activities for networking. Meet first Wednesday of each month (12 -2 pm ET. Series of discussions on different topics. This month and next month, we will focus on data integrity.

Guest Speaker:

Kjiersten Fagnan, Chief Informatics Officer and Data Science & informatics Leader, National Microbiome Data Collaborative, Berkeley Lab

Fits well into data integrity theme: The Genome Institute, a user facility, produces interesting data.

New challenges: How wish to access data differs from what was previously sought. Significant amount of cataloging underway as researchers would like to do more with this data.

NMDC project started to make data more fair and accessible.

Cross-cutting nature of microbiome research Role of microbes, microbial communities, fungi, and viruses in health of ecosystems supporting our crops, clear air and clean water, not discussed as much. Thus,

[1] Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program.

broader applications than DOE mission research and other agencies seeking better understanding of microbiome (e.g., NIH repository)

Same challenges of enabling users to locating data in the repository.

Immense scale of Data (Slide 2, graph)

Increase in metagenomes:

- Complex data set to analyze – improved analytical tools, thus more interest in generating data

Growth in orders of magnitude of what's needed in scale up of this analysis (see graphic)

- Genetic, ecological and toxnomic scales - Important for creating holistic picture of activities of microbial organisms

Big data challenge: Holistic understanding and catalogue of microbes

- Important to preserve data integrity for strong foundation for conclusions
- Able to generate data, but need data management infrastructure for collecting, managing and making data available to scientific community

NMDC Pilot (Through Sept 2021) (Slide 3, goals)

Need abstracted identifiers for data objects; thus allowing metadata change over time (enables continuity in discussions of microbial communities):

- Context: Data bases of known information: constantly changing and getting updated. Naming conventions can change in 6 months

Need close connections with community to ensure solving real problems; providing needed access

Guiding princniples for data (Slide 5)

Focused on Standards

- Leveraging existing standards for metadata, and ontologies and
- Taking data already created and mapping to existing ontological structures and enabling search onto curated sets of metadata

Quality control process-

- Reprocess raw data to ensure consistent processing (thus, interoperable and comparable data). Otherwise difficult to compare.
- Enables more powerful search as annotations are consistent. Can provide access to raw, quality or assembled data; and user can re-process data.

NMDC Pilot: 4 Aims (Slide 6) Align with Slide 5 objectives. Example:

Data facilitation and integration: Building underlying set of services allowing user to upload own data set, have it processed, indexed results and made available through programmatic API or graphical user interface

NMDC Pilot Phase I (Slide 7, diagram)

DOE resources relevant to NMDC Pilot and connectivity with external community (providing facets of metadata standards to ensure will be aligned with community)

- Data from Joint Genome Institute and EMSL
- Automated Workflows: currently established as best practices in field, but will need to change over time due to new tools – need to adapt to it so users know where data generated
    - Leads to creation of curated data and data products that can be pulled into analysis platforms (e.g., DOE KBase allowing data analysis).

NMDC will fundamentally be a powerful data catalogue; i.e., "Amazon" for microbiome data

- Explore data and see related data sets
- Select and download data (some not centrally hosted in NERSC) – send users to other sites

<u>Aim 1 : Design metadata standards</u> (Slide 8, diagram)

Building index and incorporating workflows

Aim 2: Design and deploy workflow- currently running in NERSC infrastructure (have allocation, project directory in new community file system). This data will be available through services in NERSC pilot

Phase 2 (and beyond)

- Expand engagements and think of new algorithms and methods for exploring data.
- Third party groups with robust APIs to data: build tools and infra that leverage NMDC data resources for building new visualization or analysis tools (not in scope for NMDC). Would like sufficiently robust data and infrastructure to help others to do it.
- Obtain microbiome data from other partner organizations for statistical analysis and learning
- Making data more FAIR could assist machines in helping with data discoverability and usability.

<u>Discussion</u>

Metadata standards – broadly based, beyond microbiome community. Request list of standards developing organization in this area.

- Environmental ontology for environmental sciences area uses similar terms
- Genome Standards Consortium – genomic objects
- Lack good standards: metabolized, metabolome, proteomic analysis.
- Generation of sequencing data developed by Genome Standards Consortium and Gene Ontology group. Some targeted at data NMDC generating or provide links to more general ontological standards.

Sudharshan S. Vazhkudai, Director, Hyper-Scale Data Center Program, Distinguished Scientist, Computing and Computational Science Directorate,Oak Ridge National Laboratory; *Data Integrity and Quality: A Facility Perspective*

OCLF: 1 of 2 leadership computing facilities for DOE Office of Science, Advanced Scientific Computing Research (ASCR) mission. Delivers leadership class resources for grand challenge problems (astrophysics, to earth systems, biology, physics). Storage systems must cater to diverse workload producing tens of petabytes of datasets.

Have short term, medium and long term data storage - focus on long term. Caters to other programs at lab as well.

<u>Integrity at different levels of the data stack</u> (Slide 1)

From data holdings, metadata layer to data publishing

Data Quality Assurance: HPSS Integrity Crawler (Slide 2)

Data holdings level in ORNL long term data archive only.

- Storing around 77 PB over 25 years in over 140M files from diverse science domains.
- Early 1990s: ORNL, along with NERSC, is one of partner institutions developing archival solutions – set up in.
- Lab industry collaboration: IBM with ORNL, LBL, Livermore, Los Alamos and Sandia. Worldwide deployments of roughly 3.5 exabytes.
- 2013-14: ORNL built integrity crawler infrastructure to build confidence in data holdings. Framework: randomly crawl over 140M files and create statistically representative sample of data holdings at every instance (not overly invasive) along dimensions of projects hosted. Came up with series of plug-ins for representative list (e.g., checks and verification).
- Large multi-Terabyte files: random spot check of file segment? Begin creating metadata to check every now and then.
- Disk tape infrastructure (e.g., HPSS) – can have multiple copies (disk, tape).
- Stored the way want it stored? Examine metadata to ensure that desire and intent matched with reality on system.
- Example of plug-in: all migrations and purge records serviced in timely fashion.
- Began amassing low level data integrity

The above are part of data curation and services that any large scale data management infrastructure should provide.

Metadata Layer: Need for metadata capture to improve the quality of data holdings (Slide 3)

Conducted survey of 10 years worth of HPSS logs

- Workload: more reads (40%)
- Access: opportunities regarding files beginning to store are in well-known formats (e.g., PDF) – opportunity to do something in metadata space
- Domain agnostic approach (diverse data)

Goal: how to improve quality of data holdings for multiple domains?

- To enable data discovery and make it more searchable, target 3 types of metadata:
  - Stat metadata (creation times, users, projects, etc.).
  - User-defined attributes (tags)- Provide easily used tools for tagging?
  - Contents and keywords from document files

System Architecture for Metadata Capture and Indexing (Slide 4)

Data with well-known format: Creating more meaningful metadata

- Part of HPSS platform: periodically scan databases, build indexes and extract extended attributes - created inverted indexes and make available for searches.
- Challenges: 140M, obtaining timely results, building scalable, more responsive system to user base
- Beginning meta data extraction: domain agnostic approach (e.g, Apache Tika), enable answering basic queries on data holdings

Data Publishing Workflow (Slide 5, Diagram)

DOI framework: allow users to publish, etc.

Here, introduced ability to put dataset through scientific peer review (internal domain experts look at data set and meta data and see if valuable to community); instead of random publishing.

- Helps identify data valuable to community
- Attempting to sense of data holdings

Motivation: Not have sufficient meta data about data holdings. Need to get ahold of data holdings if wish to provide meaningful services to our users.

Discussion

- MD5 hash: for use checks. Less intensive and quick: to build confidence. In addition to resorting to something less computationally intensive and time saving, can do random spot checking (hash some segments of file).

Nick Balthaser, Storage Systems Administrator, Wayne Hurlbert, Archival Storage Team Lead; and Glenn Lockwood, Group Leader, Storage Systems Group (NERSC), NERSC, *Long-term Data Management in the NERSC Archive*

Introduction

LBL/NERSC: Serve diverse community of scientific users. Scientific computing facility focused on research (unclassified). Over 40 years of archived data.

- 1998- In production with HPSS. (2 HPSS systems: User-facing system (180PB on tape and smaller center backup system (~30PB). No data transfers via file system interface; rather, HSI and HTAR (CL interfaces and also Globus)
- HPSS: Presentation on how manage over 40 years' worth of data; maintaining its reliability and availability. Presentation loosely based on System management best practices article, by Wayne Hurlbert (2011)
- Data integrity- how manage data and promptly make it available to users

Archival Data Growth (Slide 3, graph)- Exponential growth

Preventing Data Loss: 3-Tiered Approach (Slide 4)

Data management practices: Approach to preserving data on long term basis on tape

3-tiered approach (Slide 4 -8)

1) Hardware/Software
    - Vendor and media diversity important to prevent corruption of data set by 1 vendor
    - Enterprise Tape: higher capacity and better performance; media reuse possible. Less diversity in tape environment (IBM for Enterprise drives and LTO drives)
    - Enterprise class servers and disk arrays – best quality can afford for reliability and come with better support contracts; reduces impact of any single points of failure from HPSS

Software

- HPSS: supports most diversity of hardware; has most features for tape storage; roadmaps used into future. Currently no equivalent software in market.
- Integration engineering performed in-house: choose best hardware and glue together; get better understanding of system and enhances troubleshooting ability

2) System and Data Management Practices:
   - Periodic drive/media refresh (every 3-4 years)
   - 2019 completed data center move – copied data onto new drives/media (over 120PB); lost small percentage of data (may lose small portion of cartridge; not entire cartridge). Refresh other servers and arrays
   - Planned vs. unplanned maintenance: periodic and planned incremental changes (minimizes impact while enhancing reliability). Very specific planning and recording; thus, repeatable.
   - Careful and conservative system management: test, monitor, address promptly
   - Dual copy media – second copies on separate media, where possible (currently, only small files)
   - Other: backup metadata daily. Use automated host provisioning.
   - Most systems controlled by configuration management. Can generally build servers in automated way; strive for complete automation in server builds.
3) Data Environment
   - Airborne particulate monitoring – alert if out of spec
   - Temperature and humidity control- use IBM Integrated Cooling TS4500 to control environment for tape

Incidents: Dust Incident (Slide 10)

Drywall dust contamination of Oracle /STK environment, but no data loss

Worked to mitigate media damage. Used 2 enterprise drives:

- 9840 drives unaffected);
- Library – needed extensive cleaning. Separated dirty media from clean media

Change practices: established particulate monitoring/alerting; shutdown library complex. Just retired Oracle/STK libraries with in-house filtrate system

"Dimple Syndrome" (Slide 11)

Vendor inspected cartridges in STK library complex; firmware change caused physical damage to media

Mitigation: "recovery" firmware.

Data loss: ~2K files, total of 200GB. Started dual-copying small files to separate media and try to keep all small files permanently resident in HPSS disk cache. Evaluating other small file solutions.

Future

- Expand dual copy operation
- RAIT (Redundant Array of Independent Tape) - HPSS feature
- Media verification and offsite Disaster Recovery

Future Enhancements (Slide 12)

- Full redundancy for all files in archive and permanent home for small files
- RAIT for large files – has parity written to tape
- Continuous media verification
- Offsite disaster recovery (DR) site for irreplaceable data
- No plans for HA systems for HPSS.
    - Replicating data bases adds complexity to system management. Difficulties in upgrading system or performing maintenance. Added benefit may not be worth the added cost.

Discussion

Statistics on read rate vs. what ORNL reported? About same as ORNL: 40%.

Expect more content search in archives?

- Not see a lot and don't expect it. Some things previously mentioned are critical. For decades, All In archival business wanted to track data, metadata-wise. Have to self-implement (e.g., use path names for metadata characterization).
- Some users implement own metadata searching and collection techniques. Would be good if we provided.

Additional speaker suggestions Session 2/Data integrity

So far Rick Ives (UPenn) and Rick Wagner (Globus)

Updates: Frank Wuerthwein, December 2019 speaker on Cloudburst : Conducted #2 of Cloudburst yesterday. Met all goals, but not know price yet.

Roundtable

CASC: Alan Sill and Sharon Broude Geva

CASC members and government entities welcome to participate in upcoming CASC meeting in April 1-3. If have timely, important announcement, Sharon Geva will put you in touch with CASC director. Alan Sill will discuss joint project between Open Grid Forum and Open Research Cloud Alliance to build foundational basis that organizations can use to build research clouds (best practices and standards).

Meetings:

April 1 – April 3, 2020: CASC meeting, Westin Crystal City, Virginia

April 29- May 1, 2020: Women in HPC Summit, Vancouver, BC

July 26 – 30, 2020, PEARC20 Meeting, Portland, OR (February 17, 2020 deadline for submissions)

Next Meeting: March 4, 2020 (12 noon ET)