Report on Collaborative Life Sciences Data Technical Exchange Meeting held at Ames Research Center May 31, 2018

Executive Summary

On May 31, 2018 Code TI and Code SC / Ames Life Sciences Data Archive sponsored an Ames "Collaborative Life Science Data and Biospecimens” Technical Exchange Meeting (TEM), the second in a series of SLSPR TEMs. There were interesting discussions held and much information exchanged. Presentations made can be found at <https://c3.nasa.gov/dashlink/resources/1001/>

1. **TEM Rationale and Vision**

The TEM Focus was on the following:

Have ARC-based participants in one or more of the four Space Life Sciences elements (Planetary Protection, Astrobiology, Space Biology and Human Research) share data-related knowledge about common goals, process challenges and plans, and help leverage the expanding LEO knowledgebase across Space Life Sciences toward path to Deep Space

* Foster a data archiving community at ARC, share knowledge by meeting periodically
* Identify common issues, leverage resources for long-term sustainability of archives
* Increase value to archive users
* Develop an archives roadmap?

Presenters were:

1. Richard Mains - Space Life Sciences Elements Supported at ARC
2. Kaushik Chakravarty - NMCC: NASA Microbial Culture Collection (under Planetary Protection)
3. Jason Duley – OCIO Open Data (SLS Open Science)
4. Sylvain Costes - Genelab Data Systems (under Space Biology)
5. Rebecca Klotz - Biospecimen Sharing Program (under Space Biology)
6. Alan Wood - ALSDA Data Tools/Repositories and Standards (under Space Biology)

The Astrobiology Data Archive Pilot Project was invited but not able to attend. Melissa Kirven-Brooks of the Astrobiology Institute recommended a recent online paper that gives a good idea of their data archiving challenges: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4060838/. She and other Astrobiology staff attended an early-June "Deep-time Data Driven Discovery" workshop in DC (<http://www.4d-workshop.net>)

that NASA co-sponsored. She will share her notes with us soon since there was good overlap with our TEM.

**Sid Sun - Welcome**

- Data/tissue archiving efforts are key to our future - we amplify results

- Collaboration between ALSDA, BSP and Genelab is paying off

- Supports TEM concept

**B.  Project/Program Reports**

**Richard Mains - Space Life Sciences Elements Supported at ARC**

* Big data makes reference to Data Engineers and Data Scientists but neglects to include archivists in the process, and the ‘people process’ is essential.
* Probably need a common SLS Data Dictionary.
* Need to use standards in order to confederate archives.
* Jeff Smith’s Translational book concept is a product of the archives (ALSDA). (<http://www.nasa.gov/sites/default/files/atoms/files/nasa-sp-2015-625.pdf>)
* Still a problem with people on the inside knowing how to find things but people on the outside having a difficult time, even to know what is there to look for. Easy access may be the biggest single issue.

**Kaushik Chakravarty - NASA Microbiology Culture Collection (NMCC)**

* David Smith, ARC Space Biosciences, is a NMCC member
* The NMCC has applications in Planetary Protection
* Starting to get a handle on the distributed collections of microbes, maybe ~7,000 isolates agency wide.
* Where are the extremophiles? They should be a part of this collection.
* What do other agencies (NIH, CDC for instance) have in the way of isolates?

Planning to link Genelab database to Isolate database? Genelab is receiving isolates from Venkat now, and could collaborate with NMCC. Could keep in Genelab and provide link to isolates too. ALSDA could be a core facility for NMCC members (ARC was suggested by E. Stoffan, former NASA Chief Scientist).

* Kevin Sato talked with David Smith when he was doing a survey for Craig Kundrot/SLPSRA. Various funding mechanisms are being considered and Kundrot has approved the Phase I effort (see slide)
* What is the status of cataloging the Isolates database? Need to query David Smith.

**Jason Duley on the Open Data NASA initiative**

* Open Data registry, open research access. Creating a top thin layer of directory information from which to navigate broadly.
* Creating a Research Data Commons. Capable of uploading 2GB data buckets.
* Is anyone addressing Preservation? Planetary Data System (PDS) probably doing the most in this area. Is Project Open Data addressing preservation?
* Digital Object Identifiers (DOI) and DOI architecture
* ‘Data Stewards Program’
* NIH is running multiple data programs but not DOI program.

**Sylvain Costes - GeneLab**

* Omics acquisition in space is now a reality.
* The challenge is more than just putting data on a hard drive, need to make something out of it.
* Hoping to create an international presence for GeneLab
* Genelab currently only academic focus but hopefully will include commercial (there is some CASIS data in GLDB)
* Discussion of storage methods and costs. Is cloud storage really cheaper?
* GL has produced two journal publications so far.

**Rebecca Klotz - Biospecimen Sharing Plan (BSP) Overview**

* There are 100’s of thousand of vials in N261 that few know about.
* Working on figuring out best overall preservation methods to make specimens most attractive to PI’s.
* Record management under the CIO (Chief Information Officer) is unclear.

**Alan Wood - ARC Biobank & Institutional Scientific Collection**

* Overview of the software systems in use by the ALSDA and ARC Biobank/ISC
* Provided some background on the Open Archival Information System (OAIS) that the LSDA contributed to and learned from.

**C. Wrap-up remarks and questions**

* Genelab is collecting images for immunocytochemistry and wondering where to store them? Jason Duley has some options to propose later.
* Information Management in our area may be too democratic and needs a more directed approach.
* How to link Omics data withPhenotype data from model organisms? This needs to be addressed by GL soon.
* How about a SLS Archives Roadmap to navigate? Gaps and risks.
* Cooperation needed in specimen descriptions.
* Cooperation needed to begin transition to machine readable vial and box labels.
* Cooperation would be useful in tissue sharing.
* Where will we be in 5-10 years?

**Actions**:

1. Del Alto continue these meetings as a series and the next meeting to be ARC-centric again to plan a cross center TEM.Maybe as soon as July 2018

2. Del Alto gather more specifics from Sylvain Costes on path of author/PI record

3. Stewart communicate to JSC LSDA about Jason Duley's OCIO PMP and DOI registry (done).

4. Costes gather more insight on linking omic and phenotype data for model organisms.

5. Duley provide process for utilizing NASA Scientific and Technical Information (NSTI) system.

**Invitee List** (draft): "P" = presentation solicited (initial list)

General:

- Jacob Cohen (ARC Science)

- Elizabeth Taylor (Space Biosciences)

- Richard Mains (SLS Communications) - **P**

- Kaushik Chakravarty (SLS, Genelab)

- Josh Alwood (Space Biosciences, SB/HRP Translational Res.)

Planetary Protection:

- David Smith (Space Biosciences) - **P**

- Jim Arnold (?)

- Craig Burkhart

Astrobiology Institute:

- Estelle Dodson (not available?)

- Melissa Kirvan-Brooks (SLS linkage)

- Jason Duley - **P**

Space Biology:

- Sylvain Costas (Genelab) - **P**

- Kevin Sato (SB Science) - **P**

- Alan Wood (ALSDA - Databases) - **P**

- Alison French (ALSDA - Data Entry)

Human Research:

- Chris Maese (ARC HRP Mgr)

- Laura Lewis - (ARC HHC Mgr) - **P**

- Helen Stewart - Transl. Integration - **P**

Attendee List

|  |  |
| --- | --- |
| **NAME** | **ROLE** |
| Sid Sun | ARC Biosciences Division Chief |
| Richard Mains | Translational Research & SLS History SME |
| Kaushik Chakravarty | ALSDA Lead Senior Scientist  |
| Helen Stewart | Ames LSDA Civil Servant Lead |
| Martha Del Alto | ALSDA Civil Servant Lead |
| Bill McDermott | ALSDA Civil Servant Lead Programmer |
| Daniel Berrios | IT Architecture/Bioinformatics, GeneLab |
| Chris Maese | ARC HRP Project Manager & SC Deputy Div Chief |
| April Gage | ALSDA Archivist |
| Alan Wood | ALSDA Information Management Lead |
| Edina Blazevic-Levic | History Office Support/Video Analyst |
| Sylvain Costes | GeneLab Project Manager |
| Mellissa Kervin-Brooks | Staff Scientist Astrobiology Institute |
| Rebecca Koltz | Rodent Research BSP Coodinator |
| Richard Papasin | Branch Chief (Collaborative Assistant Systems)  |
| Jason Duley | Program Manager, Open Data and Code Sharing Programs |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |