

Fall 2013 Meeting of the US Culture Collection Network  
October 9 - 10, 2013. National Center for Marine Algae and Microbiota, East Boothbay Harbor, Maine.

## Agenda

Wednesday October 9

9:00 – 12:00 Welcome and general presentations from participants

9:00 Kevin McCluskey, The USCCN: a research coordination network for a community of ex situ microbial germplasm repositories.

9:15 Micah Krichevsky, Collection data exchange experiences

9:30 Kyria Boundy-Mills, Can living collections learn from natural history collections

10:00 David Smith, The ECCO/ MIRRI solution to databases

10:30 Break

11:00 Juncai Ma, The World Data Centre for Microorganisms & Global Catalog of Microorganisms

11:30 Vincent Robert, biolomics, a total package for living microbe collections

12:00 Lunch

1:30 – 3:30 Presentations specifically on different lab/data management providers

1:30 Jim Beach, Specify, a system for taxonomic collections

2:00 Dave Labeda, The USDA NCAUR collection

2:30 Eric Perkins, The ADDGENE story

3:00 Richard Pearse, Harvard Eagle-I program for specimen management and identification

4:00 Travel to NCMA and Bigelow labs

4:30 Tour of NCMA

5:30 – 9:00 Working Dinner/ reception

Culture Collection website tours: Open projector to allow different collections to demonstrate online ordering systems

Thursday Oct 10

9 – 12 Brief presentations on software requirements

9:00 A. Rick Bennett, Welcome and perspective on a shared database

9:30 Kevin McCluskey, The evolution of software at the FGSC: Unique data does not require a unique solution

10:00 John Wertz, the E. coli collection: The collection history through growth of the database

10:30 Break

11:00 Seogchan Kang, Database challenges for living collections

11:30 Willie Wilson, infrastructure level databases: Selection and implementation

12:00 Lunch

1:30 – 3:30 Discussion of cyber infrastructure options

1:30 Specify model- open source, local install

2:00 Grin Global model- central install, remote support

2:30 Status quo- leaving well enough alone

3:00 – 4:30 Establishment of a working group to develop cyber infrastructure via grant proposal, etc.

5:00 Formal adjournment

## **Abstracts for the USCCN Meeting at the National Center for Marine Algae and Microbiota October 9 - 10, 2013**

**Unique data does not require unique solutions.** Kevin McCluskey and Aric Wiest, Fungal Genetics Stock Center, Kansas City, Missouri, USA.

While the Fungal Genetics Stock Center originally used paper deposit forms, it did not develop an electronic catalog until 1985 and then has upgraded its electronic capabilities several times. While it has been well served by its custom database, it is long overdue for a new update.

The data kept on strains in a genetic stock center is different from the data kept in a type or diversity focused collection. Most of the strains in the FGSC collection are of one species and differ primarily in their genotypes. Some wild-isolated strains, and genome reference strains are also kept at the FGSC.

To promote open data, the FGSC has been a collaborating collection at the straininfo.net database for many years. This requires a custom data export because the FGSC does not maintain data in traditional fields. This was also required for participation in the World Data Center for Microorganisms system. The FGSC is participating in a time when more and more genetically characterized or genetically modified strains are being entrusted to culture collections. Similarly, we are pioneering the resequencing of multiple isolates of the same species and will propose a number of new approaches to integrating whole genome sequence with culture collection databases.

**Database challenges for living collections.** Seogchan Kang, Department of Plant Pathology & Environmental Microbiology, Penn State University, University Park, PA 16802

To maximize the value of living microbial culture collections in advancing science and education, we should go beyond mere physical preservation of captured microbial diversity and associated data. Specimens and data derived from them should be archived in a format that is easily searchable and expandable. In addition, these collections must be connected intimately to active research communities and their associated expert personnel so as to ensure their continuous use and growth in content. To effectively address these needs, it is essential to establish a robust and user-friendly support cyberinfrastructure. Using lessons learned from developing online community platforms built around major *Fusarium* and *Phytophthora* culture collections, I will present challenges and opportunities in establishing such a cyberinfrastructure.

**The Addgene Story.** Eric J. Perkins, Addgene, Cambridge, MA

Since it was founded in 2004, Addgene, a plasmid repository, has been continuously updating its approach to database management. Internally, the organization's Lab Information Management System (LIMS) has needed to accommodate significant increases in both deposits and requests over the last five years. On the external website, the Scientist and Outreach teams are constantly curating the collection by updating data as needed and creating thematic education pages based on technologies and research areas. In order to illustrate our database strategy, I follow the timeline and status changes of a plasmid as it proceeds from deposit initiation, through storage and quality control, to eventual distribution.

**The CGSC: Managing the Collection and the Data Over Time.** John Wertz, The Coli Genetics Stock Center, Yale University, New Haven, CT

Over the past 50 years, the E. coli Genetic Stock Center has been the leading repository of genetically defined laboratory strains of E. coli. And since its inception, we have always understood that the information management aspect of the task was by far the greatest challenge. Over the decades, advances in technology have not only changed how we manage and make available our data, but also changed the nature of the data we have to manage. This trend of course continues as we enter the genomic era and grapple with the needs of a broadening user community. While some of the issues and examples may be unique to E. coli, I am sure that each collection has its idiosyncratic issues. This is perhaps the largest challenge for the development of a fully shared cyber infrastructure.

**Eagle-i: a better way to publish information about research resources.** Richard Pearse, Harvard Medical School  
Research labs around the world have developed extraordinary collections of resources that have been, or are currently being used to carry out experimental programs. These tools include animal models of human disease, antibodies, stem cells, microorganisms, plasmids, and custom software to name a few. While the ability to develop these research tools has accelerated dramatically in the past 30 years, the ability to share information about what we've created has plateaued. A relatively new technology for publishing information online allows people to use query information based on its meaning rather than simply based on character matching. The eagle-i project has used this semantic web (web 3.0) technology develop an open source platform for publishing information about research tools as linked open data. The eagle-i web is a federated network of servers each serving an eagle-i node consisting of an RDF repository for data storage, Semantic Web Entry and Editing Tool (SWEET) for data entry, and a local Search application, all of which are dynamically configured by a centralized ontology that describes the resources. This eagle-i framework allows the capability to browse resources using logical connections to related resources, diseases, genes, location, and other relevant properties. This presentation will demonstrate the eagle-i architecture, functionality, and relevance to specific resource collections.

## Can Living Collections Learn from Natural History Collections?

Kyria Boundy-Mills, PhD

Curator, Phaff Yeast Culture Collection

University of California Davis

The vast majority of living organisms on earth are unicellular. Microbes provide priceless ecosystem services such as nitrogen fixation and decomposition of organic matter. Microbes also have significant economic and social impacts in fields ranging from disease to biotechnology to fermented foods and beverages.

If UC Davis is a representative example, academic microbial collections far outnumber natural history collections. However, natural history collections garner proportionally more federal support through programs such as the NSF Collections in Support of Biological Research and the NSF Advancing Digitization of Biological Collections programs. How can microbial collections learn from the examples of natural history museums to generate greater support, public familiarity and positive image?

Forty years ago, the US was one of the leaders of the effort to establish the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). This international treaty brought positive publicity to the role that natural history museums play in conserving biodiversity worldwide. Each country participates in the Scientific Authority that meets yearly to decide which species are declared endangered or threatened. International transport of these species is regulated. Natural history museums, and any US researchers conducting field work that may involve species listed in CITES appendices, must apply and receive a CITES Protocol Number. NSF supports digitization of NHM databases, in part to document site of origin and date the specimen was collected, two critical pieces of information needed to enforce CITES.

In contrast, the US has failed to ratify the Convention on Biological Diversity, which aims to ensure conservation of biodiversity, including microbial biodiversity, and traditional knowledge to provide for sustainable use with appropriate access and benefit sharing. The Nagoya Protocol, completed in 2011 and to be enacted soon, will impact that international transport of ALL microbes, not just the endangered ones. The US is far behind in setting up national procedures and this lag may severely impact the operation of US culture collections. There are no federal support programs to build databases to document the country of origin and date the microbe was collected, two pieces of data that determine whether CBD terms apply to a given strain. No information is being circulated widely to US microbiologists and to culture collections regarding the National Focal Point and Competent National Authority responsible for determining national procedures. US microbiologists, including culture collection curators, may not be aware that international peer-to-peer microbe exchanges may soon not be allowed. Federal funding must significantly increase to support development of microbial culture collection databases and establishment of national protocols, or the international standing of US culture collections will continue to erode.

## **The development of the MIRRI ITC infrastructure for microbial resources**

Paolo Romano<sup>a</sup>, Boyke Bunk<sup>b</sup>, Anna Klindworth<sup>c</sup>, Alexander Vasilenko<sup>d</sup>, Frank Oliver Glöckner<sup>c</sup> and David Smith<sup>e</sup>  
<sup>a</sup>IRCCS AOU San Martino IST, Genova, Italy. <sup>b</sup>Leibniz-Institut DSMZ, Braunschweig, Germany. <sup>c</sup>Jacobs University Bremen, Bremen, Germany. <sup>d</sup>Russian Academy of Sciences, Pushchino, Russia. <sup>e</sup>CABI, Egham, United Kingdom. E-mail: d.smith@cabi.org

Microbial resource centers have offered fundamental services to the scientific research community for centuries. In the Internet era, many efforts have been carried out to support the integration of their data and services. Not least The World Data Centre for Microorganisms (WDCM) which continues to bring together culture collection data. In Europe the Common Access to Biological Resources and Information (CABRI), a project funded by the EU from 1996 to 1999, implemented unified access to culture collection catalogues, also guaranteeing a common level of quality of material and related information. These efforts are being reviewed to get a clear picture of what tools are already available and where the gaps are with a view to deliver an information system strategy for the Microbial Resource Research Infrastructure (MIRRI). MIRRI started its preparatory phase in 2012 aiming to provide a wealth of microbial resources, associated data, taxonomic methods, and expertise to serve users' needs. This European Forum on Research Infrastructures (ESFRI) project will collaborate with other European Research Infrastructures and link out globally to projects and initiatives to avoid re-inventing the wheel in this complex arena. The MIRRI "Data Resources Management" activity serves to improve the quantity, quality, interoperability, and usage of data associated with biological material for stakeholders in academia and industry. Current protocols presently available have resulted in heterogeneous and incomplete datasets. However, as an internationally accepted data management standard the CABRI guidelines are leading the way. This is a formative document to the Organization for Economic Cooperation and Development (OECD) best practices for Biological Resource Centres (the modern day culture collection) A MIRRI work package will address this problem by compiling and developing concepts, minimal and recommended data sets and standards for data acquisition. Common strategies for evaluation, curation, integration, and interoperability of existing and future data across MRCs will be developed. Moreover, the requirements for data access will be investigated. This includes modules for a user-friendly interface and restricted user access assuring maximum data security. An assessment on existing tools, platforms, standards, and many projects, including CABRI is part of the activities to avoid redundancy and duplication and to harness existing know-how.

The review to date shows that desk top applications remain the dominant systems to access and manage collection's data. They are easy to use and fast but installations and software maintenance can be challenging, especially in collections with multiple curators or users (technicians, researchers, etc.) using different operating systems. Some culture collections have moved from desk top applications to web based but the majority of them are still using desktop for the management/curation of their databases. Web applications remain too slow and limited in their functionalities and capacities to handle some specific data. Technological advances (Java, Silverlight, HTML 5, etc.) might resolve some of the issues but it remains doubtful that significant changes will be made in the near future. A number of large culture collections have developed their own systems to manage their data. While for a very small culture collection with one or two users, this can be seen as a viable solution provided that the system to be developed remains simple, it is certainly not an advisable solution for most culture collections that are anyway lacking serious teams of developers.

Using open-source or free software is really common among culture collections due to the lack of financial resources to buy commercial solutions. Many tools have been developed to manage, analyze and publish data. Some are easy to use and propose very interesting functionalities. While some of the solutions are extremely efficient in their field, there is no open-source or free solution that can handle all the operations that are needed for curators. However, some solutions are quite interesting such as ScratchPads (SP) and the newly developed WDCM workbench (WB) created by the Chinese Academy of Sciences (CAS). There are a few commercial software tools that could be used to manage all the operations associated with a CC. Some of which are, BioloMICS, Bionumerics, FileMaker Pro, Geneious, KE Emu, LabCollector LIMS, MS-Access, MuseumPlus, Oracle and others.

In this paper the most wanted functions for curators managing their websites are discussed including direct access to published data, easy release of new strains and associated data and restricted data access to Internet users/clients if needed. However, the user is most important and they need to find several pieces of information related to organism use such as specific properties or microbial associations. End-users most wanted functions include the need for an easy searching system for as many features as possible, an advanced query system allowing the combination of searches etc. It is clear that culture collections do not hold all the data about the strains they hold and are probably never likely to; so a key function of a MIRRI information system will be interoperability with data held by other communities and disciplines. Most culture collections will not have the resources both expertise and technologies to screen their entire holdings. Linking to other data sets at various levels that cover, habitats, ecosystems, metabolism, chemistry, sequence data, taxonomic hierarchy etc. becomes essential. You can identify known organisms with the potential for the right chemistry and identify relatives that might have similar properties; you can identify substrates and ecosystems that may require their associated organisms or inhabitants to have the right chemistry then either

select other organisms that inhabit such environments or discover where these environments are and isolate more potential candidates. Can the MIRRI information system achieve all this?

Reference: OECD Best Practice Guidelines for Biological Resource Centres (June 2007),

[http://www.oecd.org/document/36/0,3343,en\\_2649\\_34537\\_38777060\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/36/0,3343,en_2649_34537_38777060_1_1_1_1,00.html)

Acknowledgements: Colleagues in work package 8 of the preparatory phase of MIRRI which has been financially supported by the European Commission (grant agreement no. 312251).

## **Nomenclature Variation as a Component in Assessing the Consistency of Identification of Pseudomonads and Related Organisms**

*Micah I. Krichevsky<sup>1</sup>, Paul De Vos<sup>2</sup>, Mark Segal<sup>3</sup>, Elaine J. Krichevsky<sup>1</sup>, Surang Dejsirilert<sup>4</sup>, Patrick Gillevet<sup>5</sup>, Deborah Henry<sup>6</sup>, Jorge Lalucat<sup>7</sup>, Edward Moore<sup>8</sup>, Jane Tang<sup>9</sup>, Sue Whitehead<sup>10</sup>, Yuguang Zhou<sup>11</sup>, Hans Yu<sup>12</sup>, George Arvanitakis<sup>12</sup>*

<sup>1</sup>Bionomics International, Wheaton, MD 20902; <sup>2</sup>University of Gent, B-9000 Gent, Belgium; <sup>3</sup>Environmental Protection Agency, Washington, DC, <sup>4</sup>National Institute of Health, Nonthaburi 11000 Thailand; <sup>5</sup>George Mason University, Manassas, Virginia, 20110; <sup>6</sup>University of British Columbia, Vancouver, British Columbia V5Z 4H4 Canada; <sup>7</sup>Universitat de les Illes Balears, E-07071 Palma de Mallorca, Spain; <sup>8</sup>The Macaulay Research Institute, AB15 8QH Aberdeen, Scotland, UK; Washington, DC 20460; <sup>9</sup>American Type Culture Collection, Manassas, Virginia, 20110; <sup>10</sup>Children's & Women's Health Centre of B.C., Vancouver, B.C. V6H 3V4 Canada; <sup>11</sup>Chinese Academy of Sciences, Beijing 100080 P.R. China; <sup>12</sup>Health Canada, Tunney's Pasture, Ontario K1A 0L2, Canada  
Background: Those wanting identification of unknown bacteria need to know the reproducibility of identifications, especially to determine possible pathogenicity and to support regulatory assessments of potential hazard. Sponsored by Health Canada, 300 isolates (Phase I, 2000-2003) 250 isolates (Phase II, 2003-2006) and 25 isolates with published whole genome sequences, (Phase III, 2007-2010) of *Pseudomonas*, *Burkholderia* and related genera were sent to 8 laboratories (I and II) and 4 laboratories (III) for identification. The aim is to evaluate reproducibility. The laboratories varied in mission and methodologies. A serious issue in identification is the underlying nomenclatures used by taxonomists, commercial or personal identification methods, public data banks, etc. We wished to separate the inconsistencies in identification due to nomenclature from variations in the laboratory data as well as inconsistency due to methodology and names assigned by the original collections.

Methods: A distributing laboratory sent randomly encoded cultures to each laboratory. Using their own methods, each laboratory submitted best estimates of the isolates' identities with supporting data. We created databases of genera, species, subspecific identifications, identification failures and primary data. To normalize the taxon names we used the name reported by the contributor of the initial culture and the name in the Approved Lists of Bacterial Names for each Phase. We tabulated the consistency of identification across laboratories using the original names assigned and adjusted according to the official names for each of the dates certain.

# USCCN Fall 2013 meeting at the National Center for Marine Algae and Microbiota

## Post meeting survey results:

1) Did you participate in the Fall 2013 US Culture Collection Network meeting at the Bigelow Labs in Boothbay Harbor Maine?

**y 13 n 0**

2) Have you participated in previous USCCN activities

**y 9 n 4**

3) Did the Fall 2013 USCCN Meeting at the Bigelow labs and National Center for Marine Algae provide information that you were not aware of prior to the workshop?

**y 13 n 0**

4) Did you meet colleagues with shared interests and objectives but whom you had never met before?

**y 13 n 0**

5) Did you find the Fall 2013 USCCN Workshop at Bigelow Labs/ NCMA to be a valuable opportunity to meet and share information with colleagues from the living collection community?

**y 13 n 0**

6) Please describe your response regarding the amount of time spent on formal presentations

**Too much 4 just right 9 not enough 0**

7) Please describe your opinion regarding the amount of time available for informal interactions

**Too much 0 just right 10 not enough 3**

8) Please describe your opinion of the amount of time spent on touring the NCMA Culture Collection and related facilities

**Too much 0 just right 13 not enough 0**

9) What actions will you undertake as a result of this workshop? (select as many as you wish)

Activity	# responses
Participate in additional USCCN Activities	11
Register/update my collection with the WFCC	4
Post my collection database on the GCM for searching via <a href="http://www.usccn.org">www.usccn.org</a>	4
Explore the eagle-i data system for my collection	6
Encourage a colleague to participate in USCCN Activities	9

Responses added in reply to question 9)

- My collection is already registered with the WDCM (A WFCC affiliated collection) and has provided its data to the GCM (CABI (formely IMI, UK))
- Look at the potential use of Specify for the support of culture collections



## Invited feedback

- thought this was a well organized meeting. Despite a few people not being present due to the shut down, it was still a relatively diverse group capable of producing a useful discussion.
- This was one of the best meetings I've had the pleasure of attending in recent years. It was small enough that I had the chance to personally meet each representative yet large enough to thoroughly represent the much larger community. Kudos to the organizers and to the culture collection community for supporting this forum.
- Phenomenal meeting location and timing was perfect. Whoever planned that time and place is genius.
- The meeting has provided the opportunity for MIRRI to collaborate with USCCN and steps towards this are being taken
- Nice meeting. I would like to spend more time talking about specific goals and plans and less time on talks in future meetings. Thanks!