

**Inaugural Digital Data in Biodiversity Research Conference
GBIF Session Abstracts**

5-6 June 2017

**Co-sponsored by the University of Michigan and iDigBio
Hosted at Michigan League, University of Michigan**

Bruneau, Anne
Professor
Université de Montréal
anne.bruneau@umontreal.ca

Canadensys: revealing the biodiversity of Canada

Canadensys (www.canadensys.net), the largest Canadian GBIF node, currently publishes over 3 million records to GBIF and aggregates nearly 4 million occurrences on the Canadensys Explorer. This pan-Canadian initiative, which took its roots within universities and museum around the country, has published 52 datasets from 23 institutions since 2012 and continues to actively incorporate additional datasets. In order to provide better tools and more effective visualization to its users, Canadensys is moving to an Atlas of Living Australia model, customizing its own platform by using the structure developed by ALA. Canadensys is also developing new partnerships, locally and internationally to integrate new data types and to share the expertise gained through the years with a broad community of biodiversity managers. For example, to assess the impact of private land conservation on biodiversity, Canadensys is working with the Réseau des Milieux Naturels Protégés to link geospatial data of private protected areas in Québec with occurrence data published through Canadensys. Canadensys is also involved in the Bioblitz Canada 150 initiative, helping with the aggregation and visualization of the data collected, and is mentoring a digitization initiative in Haiti, sharing its expertise with the curators of the Jardin Botanique des Cayes.

Coddington, Jonathan
Executive Committee
Global Genome Biodiversity Network
Smithsonian Institution
coddington@si.edu

Global Genome Biodiversity Network – Infrastructure for Genomic Research

The Global Genome Biodiversity Network (GGBN) is an international collaboration of 60 institutions in 22 countries dedicated to “making genomic collections discoverable for research.” All institutions holding genetic resources are encouraged to join. GGBN built a rich, modern, genetic extension to DarwinCoreArchive (also to ABCD and MIX) so that information on genetic resources can be machine reasonable, compiled, and integrated into collections databases. GGBN supplies a “missing link” in biodiversity informatics: GBIF handles occurrence data, GenBank handles sequence data, but, prior to GGBN, no globally-approved data structure existed for tissues, DNA’s, RNA’s, etc. GGBN holdings have increased 460% in the last year, comprising over 600K samples representing 2800 families, 14,000 genera, and 39,000 species. GGBN offers many member benefits, notably standardized approaches to the new requirements of the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising From Their Utilization. Advances in genomic science are

revolutionizing biodiversity research and collections management, representing a growing user community that requires fast access to global genetic resources.

Guala, Gerald
Branch Chief for Eco-Science Synthesis
USGS
gguala@usgs.gov

Biodiversity Information Serving Our Nation (BISON): Connections and Cooperation

BISON: Biodiversity Information Serving Our Nation (bison.usgs.gov) is the primary application of the US Node of GBIF and currently contains more than 375 million records for species occurrences in the US and all of its territories and Exclusive Economic Zone (EEZ) marine areas, as well as records for Canada. BISON provides true integration of those records including taxonomic synonymy and hierarchical integration as well as geospatial binning of records for polygons relevant to US researchers and the public. It also provides a large array of high speed web services that power numerous other projects. It relies on ITIS: the Integrated Taxonomic Information System (www.itis.gov) for taxonomic amelioration and on mutual sharing with the Global Biodiversity Information Facility (GBIF) for a large number of records. ITIS and ITIS Canada have had a very close collaboration for many years that ensures continuity and new efforts on both sides of the border are encouraging a closer collaboration on occurrence data as well.

Hobern, Donald
Executive Secretary
GBIF Secretariat
dhobern@gbif.org

National and Regional Coordination Roles within GBIF

GBIF's mission is not only to mobilize and organize biodiversity data for use in science and policy, but also to facilitate and promote cooperation between agencies, institutions and other stakeholders with an interest in holding or using biodiversity information. The overall aim is to remove overlaps and frictions which reduce the effectiveness and benefits possible from limited investments in understanding biodiversity. GBIF Participant countries are expected to establish a Node to achieve these goals. In many countries, the role of the GBIF Node fits into an existing national coordinating activity. In others, there may be significant challenges around developing a suitable mechanism to support the necessary interactions. GBIF also promotes regional-scale coordination to assist national Nodes in developing complementary strategies and to deliver larger-scale datasets.

Kalfatovic, Martin
Associate Director, Smithsonian Libraries
Program Director, Biodiversity Heritage Library
KalfatovicM@si.edu

Overview of the Biodiversity Heritage Library Recent Activities

As the world's largest open access digital library for biodiversity literature, the Biodiversity Heritage Library (BHL) is an unparalleled resource that has forever changed the way researchers around the globe

understand, describe, and conserve life on Earth. BHL has become not only a model for digital libraries but also a fundamental resource for taxonomic literature aggregation, discovery, and presentation by engaging the taxonomic community and responding to user needs. To achieve this, BHL relies on many standards and tools, such as Digital Object Identifiers (DOIs), Application Programming Interfaces (APIs), and Global Names Architecture (GNA). These standards and tools help ensure that data in and about the literature matches the needs and expectations of the scientific community and is readily available for widespread reuse. To meet the evolving needs and expectations of researchers, we must continually innovate and adapt to the changing technological landscape. BHL is in the process of organizing widespread user needs analysis and an environmental scan of information resources to define requirements for a next generation digital library.

Macklin, James A.
Research Scientist, Botany and Biodiversity Informatics
Agriculture and Agri-Food Canada
James.Macklin@AGR.GC.CA

The Canadian Biodiversity Information Facility (CBIF)

The Canadian Biodiversity Information Facility (CBIF) has been Canada's federal gateway to Canada's biodiversity information since 2001. It has three core components, a biodiversity data search and mapping interface, the Integrated Taxonomic Information System (ITIS) for resolving taxonomic names, and a set of informational knowledge products (e.g., Butterflies of Canada, Brassicaceae [Mustards] of Canada). This national portal serves a broad audience and will be undergoing a major rework based on 30 million dollars in new funding to digitize and mobilize Agriculture and Agri-Food Canada (AAFC) natural science collections. This funding includes resources to support the government's Open Data and Open Science policies through refreshing the current CBIF portal and enhancing its functionality through adding molecular data, high-resolution images, and new knowledge products over the next 5 years. In addition, we will add millions of new specimen records to this resource and continue to work with our Canadian academic and museum partners and other stakeholders to mobilize more biodiversity data.

Orrell, Thomas
Research Biologist
Deputy Director IT IS
Informatics Branch Chief
Smithsonian Institution
Orrellt@si.edu

The Catalogue of Life: Infrastructure for Science

The Catalogue of Life is the most comprehensive expert knowledge base of the worlds described species. It is a major component of the underlying infrastructure of biodiversity informatics as it allows the exchange of information via a common taxonomic vocabulary. Currently, the Catalogue covers over 1.7 M described species and includes significant, additional coverage of synonymy. In order to meet the growing demands of science, a new approach is underway to: 1) close known taxonomic gaps in the Catalogue; 2) include the broadest coverage of synonymy including subsequent combinations and orthographic variants; 3) include robust nomenclatural elements and lexical reconciliation; and 4) include a broad suite of services – in conjunction with other large biodiversity initiatives - to serve, curate, and maintain the Catalogue.

Page, Larry
Director, iDigBio
Florida Museum of Natural History
lpag1@ufl.edu

iDigBio, National Coordinating Center for NSF's ADBC Program

The national effort to digitize information in biocollections as outlined in the NIBA Strategic Plan has been successfully catalyzed by funding from NSF and activities of iDigBio. Digitization priorities have been defined, best practices have been integrated into workflows, cyberinfrastructure resources including a search portal have been provided, working groups have focused on major challenges, and collaborations with data providers/users have been built. These community-driven activities have led to improved digitization practices, increased involvement in digitization and training, and adoption of informatics tools that improve the efficiency and scalability of digitization. In its 6th year, iDigBio is working with staff in 495 collections in 289 institutions. Communication among stakeholders to increase access to collections data has been established through workshops, webinars, and social media. iDigBio has sponsored attendance of 3,200 unique participants from 511 institutions to 85 workshops that targeted digitization. iDigBio has ingested 1,600 recordsets containing 95 million records and 20 million images. All data ingested are indexed so that queries and other types of index-based access are supported. Searches can be done through a Web-based graphical interface or through APIs. Search and analytical tools enable users to mine diverse data such as taxonomy, location, images, traits, and vocalizations.