La iniciativa "Barcode of Life" y las colecciones de historia natural





- The Consortium for the Barcode of Life (CBOL) is an international initiative dedicated to supporting the development of <u>DNA barcoding</u> as a global standard for species identification.
- Barcoding was proposed in 2003 by Prof. Paul Hebert of the University of Guelph in Ontario as a way of distinguishing and identifying species with a short standardized gene sequence















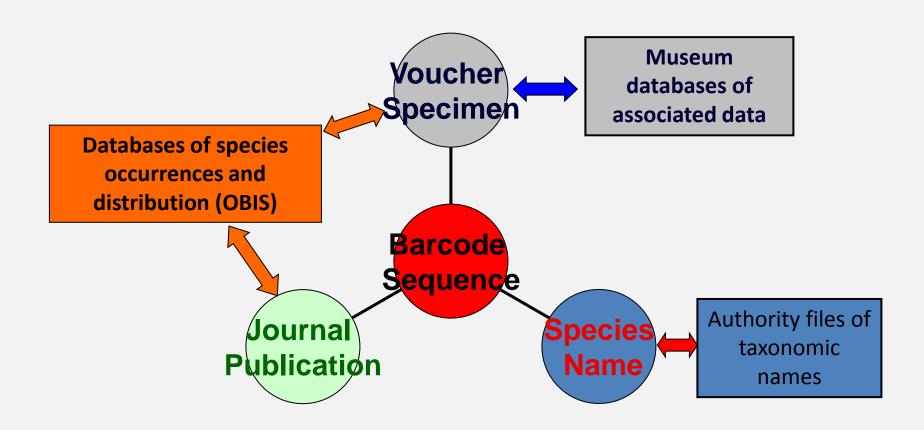
WIRED MAGAZINE: 16.10

SCIENCE : DISCOVERIES

A Simple Plan to ID Every Creature on Earth



How to do it





Isn't GenBank enough?

- Barcode records should be usable as authority references that connect DNA sequences to the names of species. To serve this function, each barcode record in GenBank should be linked to a voucher specimen, preserved and available for further study in a museum, herbarium, zoo, frozen tissue collection, or other repository of biological reference material.
- GenBank records can include references to such voucher specimens, but the data field is not structured and is therefore not easily searchable. Associating voucher specimens with GenBank records is not routine practice across all taxonomic groups.

"Linked (open) data"

¿Qué es Linked Data?

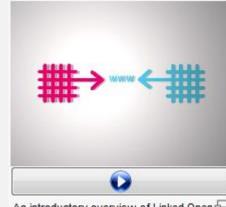
Los Datos Enlazados es la forma que tiene la <u>Web Semántica</u> de vincular los distintos datos que están distribuidos en la Web, de forma que se referencian de la misma forma que lo hacen los enlaces de las páginas web.

La Web Semántica no se trata únicamente de la publicación de datos en la Web, sino que éstos se pueden vincular a otros, de forma que las personas y las máquinas puedan explorar la web de los datos, pudiendo llegar a información relacionada que se hace referencia desde otros datos iniciales.

¿Cómo funciona?

Los Datos Enlazados, como parte de la Web Semántica, se basa en la aplicación de ciertos principios básicos y necesarios, que fomentarán el crecimiento de la Web, tanto a nivel de los documentos HTML (vista clásica de la Web), como a nivel de los datos expresados en RDF (vista de la Web Semántica).

- Usar URIs para identificar las cosas
- 2. Usar URIs HTTP
- 3. Ofrecer información sobre los recursos usando RDF
- Incluir enlaces a otros URIs



An introductory overview of Linked Open Data in the context of cultural institutions.

http://www.youtube.com/watch?v=uju4wT9uBIA

BOL & GenBank

 officials in GenBank, the US national repository for nucleotide sequence data, offered to act as an archival repository for DNA barcode records and to use "BARCODE" as a reserved keyword to identify these records. This keyword would be the flag on all BARCODE records in GenBank (BRGs).

S NCBI Resources ☑ How To ☑			
Nucleotide	Nucleotide ▼		
		Limits	Advanced

<u>Display Settings:</u> ✓ GenBank

Send:

Cricula timorensis voucher barcode SNB 1553 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial

GenBank: HM416780.1

FASTA Graphics

Go to: ☑

LOCUS HM416780 658 bp DNA linear INV 29-MAY-2013

DEFINITION Cricula timorensis voucher barcode SNB 1553 cytochrome oxidase

subunit 1 (COI) gene, partial cds; mitochondrial.

ACCESSION HM416780

VERSION HM416780.1 21:300873646
DBLINK BioProj : PRJNA37833

KEYWORDS BARCODE.

SOURCE mitochondrion Cricula timorensis

ORGANISM Cricula timorensis

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea; Saturniidae; Saturniinae; Saturniini;

Cricula.

REFERENCE 1 (bases 1 to 658)

CONSRTM International Barcode of Life (iBOL)

TITLE iBOL Data Release

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 658)

CONSRTM International Barcode of Life (iBOL)

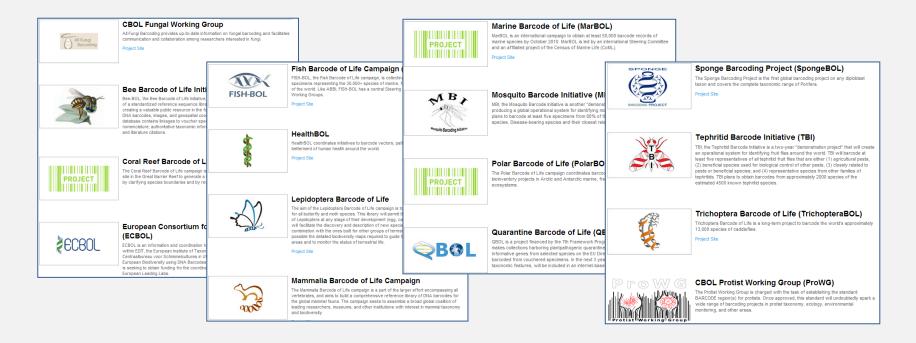
TITLE Direct Submission

JOURNAL Submitted (01-JUN-2010) Biodiversity Institute of Ontario,

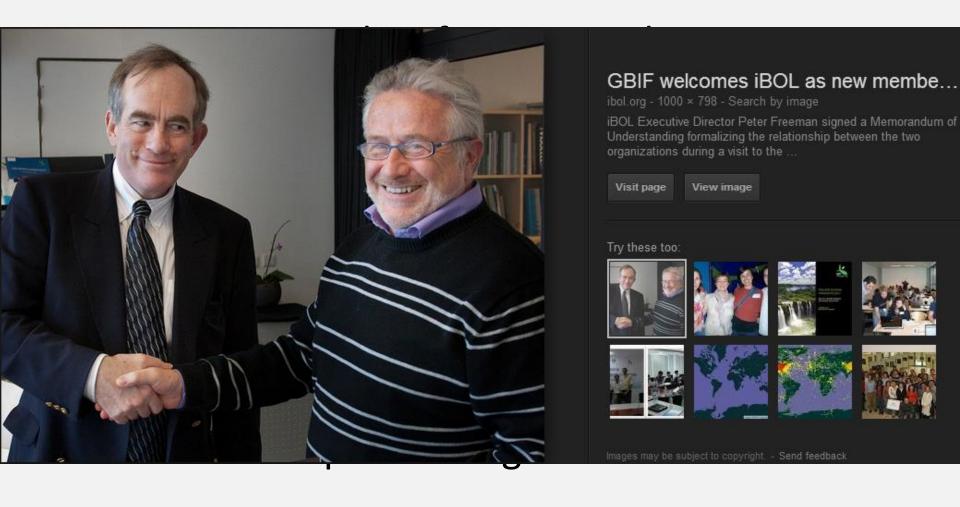
University of Guelph, 50 Stone Rd West, Guelph, Ontario N1G2W1,

BOL ecosystem

- Consortium for the Barcode of Life (CBOL
 - Agreed to support and promote the goals of CBOL and have pledged to invest resources in DNA barcoding activities
- GenBank,
 - the public archival repository for barcode data
- Barcode of Life Data Systems --www.barcodinglife.org
 - The global workbench for assembly, analysis and curation of barcode data
- iBOL (International Barcode of Life Project)
 - Formal activation in October 2010. Umbrella alliance for scientific projects



BARCODE Data Standards

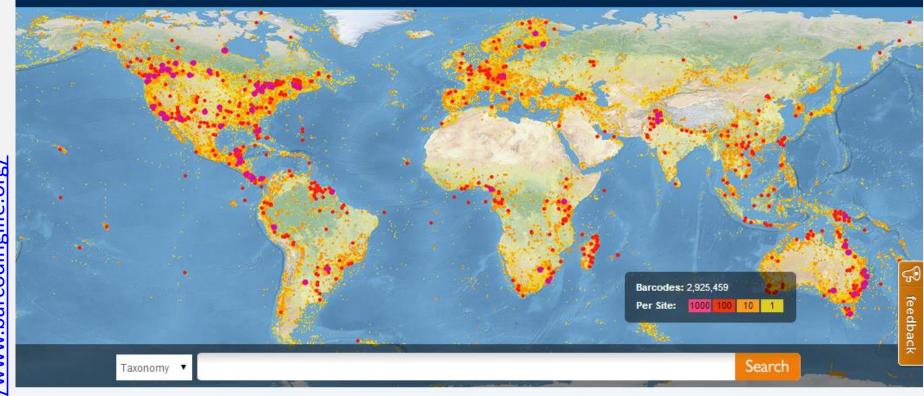


CBOL quality standards

- <u>Gold Standard.</u> Sources of taxonomic names that have been reviewed for their adherence to taxonomic standards, objective and subjective synonymy, and reflect expert opinions. Gold standard sources have the added value of being well maintained; that is, as a name is revised in subsequent releases of an index, new information on the status of that name will be retrievable through the GenBank record. The species name in the BRG will not have to be updated because the authority file will be kept current, and the history of the species name of a **voucher specimen** will be available to GenBank users through that record;
- <u>Silver Standard</u>. Sources of taxonomic names compiled in published nomenclators that may be reviewed for adherence to taxonomic standards of the nomenclatural Code and monotypic synonymy. These lists provide links back to the publication in which the name was proposed
- <u>Bronze Standard.</u> Sources of all published names (such as the proposed NameBank). This would include new names that have been recently published for taxonomic groups covered by gold and silver standard sources, but have not yet been incorporated through the normal compilation and review processes. In some cases, a submitter may want to attach a provisional name (e.g., species A) to a voucher specimen because it is a new species that is awaiting formal description. The Bronze Standard would include published provisional names. Linkage to the publication will ensure that the provisional name is unique and retrievable
- <u>Tin Standard.</u> In many cases, submitters will want to put a provisional species name on a BRG but will not publish the data. In these cases, the provisional names may not be retrievable and may not be globally unique within genus. In these cases, GenBank will add a unique string to the provisional species name at the time of submission

tgtaaaacgacggccagttctcaaccaaccacaaagacattgg, fwd name:

Databases





Public Data Portal:

A data retrieval interface that allows for searching over 1.7M public records in BOLD using multiple search criteria including, but not limited to, geography, taxonomy, and depository.



DNA Barcode Education Portal:

A custom platform for educators and students to explore barcode data and contribute novel barcodes to the BOLD database.



Barcode Index Numbers:

A searchable database of Barcode Index Numbers (BINs), sequence clusters that closely approximate species.

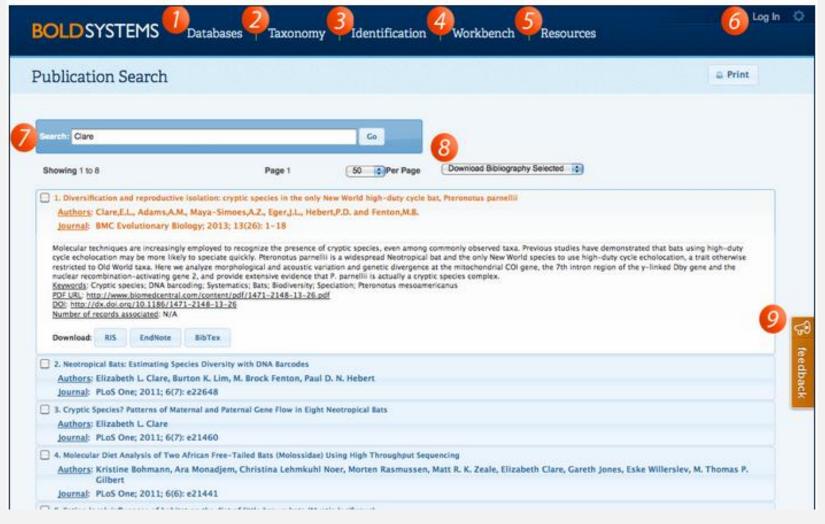


Workbench:

An integrated data collection and analysis environment that securely supports the assembly and validation of DNA barcodes and ancillary sequences.

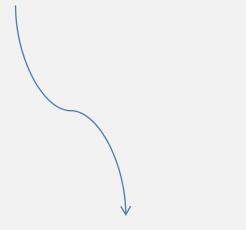
Navigating BOLD

http://www.boldsystems.org/index.php/resources/handbook?chapter=3_submissions.html



Ejercicio

http://www.ncbi.nlm.nih.gov/nuccore/JN029392.1



http://www.barcodinglife.org/index.php/IDS OpenIdEngine



http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0066213

Explore BINs

http://www.boldsystems.org/index.php/Public BarcodeCluster?clusteruri=BOLD:AAI1263

Francisco Pando

Unidad de coordinación, GBIF España Real Jardín Botánico - CSIC Claudio Moyano 1, 28014 Madrid, Spain pando@gbif.es www.gbif.es



http://creativecommons.org/licenses/by-sa/3.0/es/

