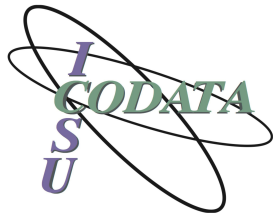# One Repository, Two Implementations, & A World of Legal Interoperability Opportunities and Challenges

Donat Agosti, Gail Clement,Willi Egloff, and Tom Morrell

Legal Interoperability of Research Interest Group

Session Outline

- Introduction to Invenio repository platform
- Caltech Data landscape
- Plazi Data landscape
- Comparison of Legal Interop Issues & Challenges
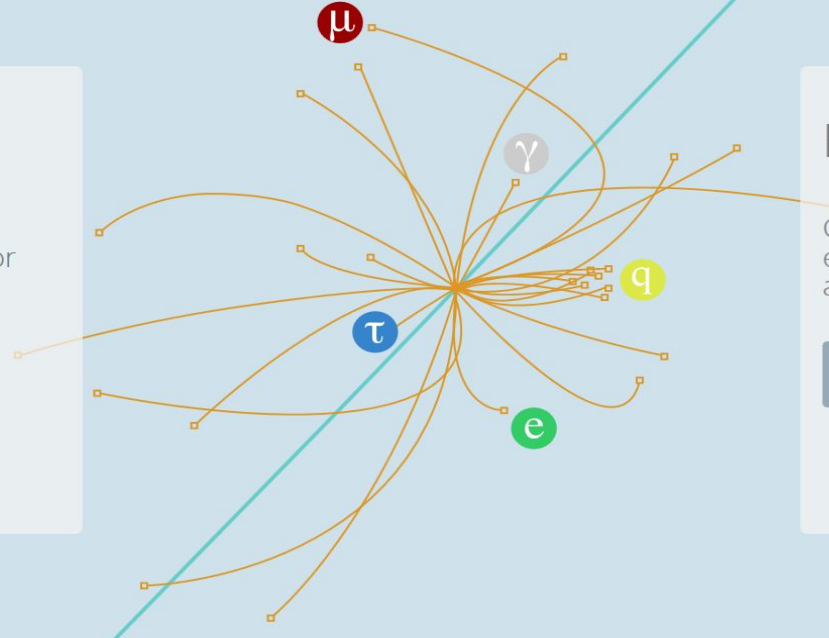- Discussion

One Repository: Invenio

1

INVENIO

Digital Library Framework

INVENIO

Flexible data model

Use JSON Schema to describe articles, books, photos, videos, data, and software. Several popular master metadata formats are supported, such as MARC21 with BibTeX, DataCite, Dublin Core, EndNote, RefWorks.

Powerful search engine

Very fast response for repositories of up to several million records. Customisable query language and second-order search operators. Configurable UI and facets. Combined metadata, fulltext and reference search in one go. Citation networks.

Configurable workflows

Organise document corpus in community collections. Configure user and robot ingestion workflows. Attribute community moderators.

Collaborative communities

Organise users in groups and teams. Share documents of interest in annotable baskets. Configure automated email and RSS notification alerts.

Extensible packages

Invenio is composed of hundreds of independent pluggable packages that collaborate via rich APIs. Pick the packages you need and use the full power of Python to extend their capabilities.

Open standards

We love open access, open source, and open standards. DOI, JSON Schema, Memento, OAI-PMH, ORCID, OpenAIRE, REST, XML... you name it.

Invenio Repository

- Free open source digital repository software
- Harvest datasets, analysis code, virtual machine environment, configuration and knowledge information.
- Visualise data in the browser.
- Rerun preserved code on the cloud.

INVENIO

The DataCite Metadata Schema is a list of core metadata properties chosen for the accurate and consistent identification of a resource for citation and retrieval purposes, along with recommended use instructions.

Table 1: DataCite Mandatory Properties

ID	Property	Obligation
1	Identifier (with type sub-property)	M
2	Creator (with name identifier and affiliation sub-properties)	M
3	Title (with optional type sub-properties)	M
4	Publisher	M
5	PublicationYear	M

Table 2: DataCite Recommended and Optional Properties

ID	Property	Obligation
6	Subject (with scheme sub-property)	R
7	Contributor (with type, name identifier, and affiliation sub-property)	R
8	Date (with type sub-property)	R
9	Language	O
10	ResourceType (with general type description sub-property)	R
11	AlternateIdentifier (with type sub-property)	O
12	RelatedIdentifier (with type and relation type sub-properties)	R
13	Size	O
14	Format	O
15	Version	O
16	Rights	O
17	Description (with type sub-property)	R
18	GeoLocation (with point and box sub-properties)	R

ID	DataCite-Property	Occ	Definition	Allowed values, examples, other constraints
7.1	contributorType	1	The type of contributor of the resource.	If Contributor is used, then contributorType is mandatory. *Controlled List Values:* ContactPerson, DataCollector, DataCurator, DataManager, Distributor, Editor, Funder, HostingInstitution, Producer, ProjectLeader, ProjectManager, ProjectMember, RegistrationAgency, RegistrationAuthority, RelatedPerson, Researcher, RightsHolder, Sponsor, Supervisor

				individual stewards need to determine which are major vs. minor versions[21]. May be used in conjunction with properties 11 and 12 (AlternateIdentifier and RelatedIdentifier) to indicate various information updates. May be used in conjunction with property 17 (Description) to indicate the nature and ...
16	Rights	0-n	Any rights information for this resource.	Free text. *** Provide a rights management statement for the resource or reference a service providing such information. Include embargo information if applicable. Use the complete title of a license and include version information if applicable. Example: Creative Commons Attribution 3.0 Germany License

Legal Interop support in Invenio

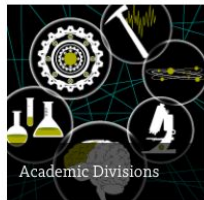## 2 Two Implementations
Caltech & Plazi

Caltech

Inventing Tools for Detecting Life Elsewhere

At Caltech, in the Exoplanet Technology Laboratory, or ET Lab, researchers have been developing new ways to search exoplanets for biosignatures.

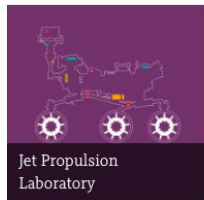## Discover More

Academic Divisions

Research

Admissions

Leadership

break**through**
THE CALTECH CAMPAIGN

The Caltech Campaign

Jet Propulsion Laboratory

Campus Life

Library

Major Research Institutes

Caltech Today

Caltech is small but prizes excellence and ambition.

The contributions of Caltech's faculty and alumni have earned national and international recognition, including 35 Nobel Prizes.

The Institute manages the Jet Propulsion Laboratory (JPL) for NASA, sending probes to explore the planets of our solar system and quantify changes on our home planet; owns and operates large-scale research facilities such as the Seismological Laboratory and a global network of astronomical observatories, including the Palomar and W. M. Keck Observatories; and cofounded and comanages LIGO, which, in 2016, observed gravitational waves for the first time.

LIGO Open Science Center

LIGO is operated by California Institute of Technology and Massachusetts Institute of Technology and supported by the U.S. National Science Foundation.

Getting Started

Tutorials

Data

   Events

   Bulk Data

Timelines

My Sources

Software

GPS ↔ UTC

About LIGO

Data Analysis Projects

Acknowledgement

Data release for event GW150914

This page has been prepared by the LIGO Scientific Collaboration (LSC) and the Virgo Collaboration to inform the broader community about a confirmed astrophysical event observed by the gravitational-wave detectors, and to make the data around that time available for others to analyze. There is also a **technical details** page about the data linked below, and feel free to **contact us**. This dataset has the Digital Object Identifier (doi) http://dx.doi.org/10.7935/K5MW2F23

Summary of Observation

The event occurred at GPS time 1126259462 == September 14 2015, 09:50:45 UTC. The false alarm rate is estimated to be less than 1 event per **203,000 years**, equivalent to a significance of **5.1 sigma**. The event was detected in data from the LIGO Hanford and LIGO Livingston observatories.

- There are Science Summaries, covering the information below in ordinary language.
- There is a one page factsheet about GW150914, summarizing the event.

How to Use this Page

- **Click on the section headings below to show available data files.**
  - (click to Open/Close all sections)
- There are lots of data files available in the sections below, look for the word **DATA**.
- Click on each thumbnail image for larger image.
- See the papers linked below for full information, references, and meaning.
- Many of the data files linked below have heterogeneous formatting; if you have any questions, please contact us.

The G150914 detection paper:

Observation of Gravitational Waves from a Binary Black Hole Merger

For full details see LIGO DCC, arXiv, or Phys. Rev. Letters
This paper and all the companion papers can also be found at papers.ligo.org.

Estimated source parameters

Quantity	Value	Upper/Lower error estimate	Unit
Primary black hole mass	36	+5 -4	M sun
Secondary black hole mass	29	+4 -4	M sun
Final black hole mass	62	+4 -4	M sun
Final black hole spin	0.67	+0.05 -0.07	
Luminosity distance	410	+160 -180	Mpc
Source redshift, z	0.09	+0.03 -0.04	
Energy radiated	3	+0.5 -0.5	M sun

TABLE I. Estimated source parameters for GW150914. We report the median value as well as the range of the 90% credible interval. Masses are measured in the source frame; to convert masses to detector frame, multiply by (1 + z). The source redshift assumes standard cosmology.

California Institute of Technology

Research Data Repository

Search 11 records

Data Sets

Software

Submit a Data Set

Terms >

data.caltech.edu

Long-tail datasets in CaltechData

Identifying and Quantifying Mineral Abundance through VSWIR Microimaging Spectroscopy: A Comparison to XRD and SEM

Dataset
2017-03-13
CaltechDATA

Details

Authors	Leask, Ellen K. Division of Geological & Planetary Sciences, California Institute of Technology Ehlmann, Bethany L. Division of Geological & Planetary Sciences, California Institute of Technology 0000-0002-2745-3240 ORCID
Description	Other: Sample: A rock was collected from an travertine conglomerate, containing ophiolite clasts. The rock was collected in Oman near the Samail ophiolite (collected Jan 2012; B. Ehlmann; analyzed in Leask & Ehlmann, 2016). Data included are: 1) UCIS (Ultra-Compact Imaging Spectrometer; B. Van Gorp et al) data cube. Data near ends of sensor may be suspect [e.g. under 500 nm, over 2500 nm]. {OM12L_001_UCIS_cube_masked} 2) EDS-SEM mosaic image cube (4x downsampled to despeckle, using a nearest neighbour algorithm). Acquired Caltech October 2015. Chemical data are in atomic %. {om12L_001_SEM_4x_masked} 3) EDS-SEM data warped to UCIS cube [degree 4 convolution, using built-in ENVI warping algorithm and hand-[ ] control points]. Three UCIS images were mosaicked together to build the [ ] quantitative mineralogy Mars analogue petrology
DOI	10.22002/D1.222
Version	1.0
Related Identifier(s)	IsIdenticalTo (URL): ftp://ftp.gps.caltech.edu/pub/ehlmann /Leask_and_Ehlmann_2016_OM12001_hyperspectralcube[ ] IsSupplementTo (URL): http://resolver.ca[ ] [ ]920758
License	cc-zero
Funding	NASA: NNX12AB42G JPL-RTD/PDF Rose Hills Foundation
Language	eng
Related publications	Identifying and Quantifying Mineral Abundance through VSWIR[ ] XRD and SEM Leask, Ellen K., Division of Geological & Planetary Sciences, Califo[ ] Ehlmann, Bethany L., Division of Geological & Planetary Sciences

Graduate Student

Faculty | JPL Researcher

Data copied from another server

US federal funder

US private funder

Data associated with published paper

Identifying and Quantifying Mineral Abundance through VSWIR Microimaging Spectroscopy: A Comparison to XRD and SEM

Dataset
2017-03-13
CaltechDATA

Variety of data types

Filename	Size	
Macro_OM12L_001.JPG	1.1 MB	⬇ Download
om12L_001_SEM_4x_masked.hdr	717 Bytes	⬇ Download
OM12L_001_smoothface_2.JPG	898.7 kB	⬇ Download
om12L_001_SEM_4x_masked	17.4 MB	⬇ Download
OM12L_001_UCIS_cube_masked.hdr	3.2 kB	⬇ Download
OM12L_001_UCIS_cube_masked	194.4 MB	⬇ Download
OM12L_001_warp_SEM_to_UCIS.hdr	15.7 kB	⬇ Download
om12L_001_XRD.csv	191 Bytes	⬇ Download
read_me.txt	2.2 kB	⬇ Download
OM12L_001_warp_SEM_to_UCIS	313.7 MB	⬇ Download

Big Science datasets in CaltechData

TCCON data from Ascension Island (SH), Release GGG2014.R0

Dataset
2017-03-30
CaltechDATA

Download Edit

Details

Authors	Feist, D. G. Max Planck Institute for Biogeochemistry, Jena (DE) 0000-0002-5890-6687 ORCID
	Arnold, S. G. Max Planck Institute for Biogeochemistry, Jena (DE)
	John, N. Ariane Tracking Station, Ascension Island (SH)
	Geibel, M. C. Stockholm University, Stockholm (SE) 0000-0002-7369-0781 ORCID
Contributors	Roehl, C. M. DataCurator California Institute of Technology, Pasadena, CA (US) 0000-0001-5383-8462 OR
Description	Abstract:
	The Total Carbon Column Observing Network (TCCON) is a network of ground-based Fourier Transform Spectrometers that record direct solar absorption spectra of the atmosphere in the near-infrared. From t spectra, accurate and precise column-averaged abundances of atmospheric constituents including CO_2, C N_2O, HF, CO, H_2O, and HDO, are retrieved. This data set contains observations from the TCCON station on Ascension Island.
Publication Date	2017-03-30
Subject(s)	"atmospheric trace gases", "CO_2", "CH_4", "CO", "N_2O", "column-averaged dry-air mole fractions", "remote sensing", "FTIR spectroscopy", "TCCON"
DOI	10.14291/tccon.ggg2014.ascension01.R0/1149285
Version	GGG2014.R0
Format	application/x-netcdf
Related Identifier(s)	IsPartOf (URL): http://tccon.ornl.gov/
	IsDocumentedBy (DOI): 10.14291/tccon.ggg2014.documentation.R0/1221662
	IsCitedBy (DOI): 10.5194/amt-9-683-2016
	IsCitedBy (DOI): 10.1002/2015JD023389
	IsCitedBy (DOI): 10.5194/acp-16-1653-2016
	IsCitedBy (DOI): 10.5194/amt-9-3491-2016
	IsCitedBy (DOI): 10.5194/amt-9-2381-2016
	IsCitedBy (DOI): 10.1002/2015JD024157
Relevant Dates	Created: 2014-10-01
	Updated: 2017-03-01
	Collected: 2012-05-22
License	other
Funding	Max Planck Society
	Max Planck Institute for Biogeochemistry
Language	eng
Geographic coverage	Ariane Tracking Station (AC)
	-7.9165 -14.3325

TCCON data from Ascension Island (SH), Release GGG2014R0

Dataset

Download

Filename	Size	
README.txt	12.1 kB	Download
ae20120522_20161221.public.nc	10.6 MB	Download

Annotations:

- Creators from outside Caltech/US
- Contributor from Caltech
- References Documentation
- Data cited by other outputs
- German funders
- Data documentation refers to external site for usage policy

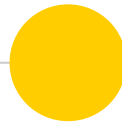TCCON > TCCON Data and Network Policies > Data Use Policy

Data Use Policy

TCCON data can be downloaded from the TCCON archive, hosted by CDIAC as per our Data Protocol. Please acknowledge the use of the data as described below. Citation files in bibtex (or the concise version), RIS, and XML format has been created for your convenience that include all TCCON data citations. More detailed information on the data can be found on the Data Description page. Auxiliary data, including column averaging kernels and a priori profiles are contained in the netCDF files provided for GGG2014. Instructions on how to properly use the averaging kernels and a priori profiles for the TCCON data are on the Auxiliary Data wiki page.

For updates on our data, sign up to our data users list at https://lists.gps.caltech.edu/mailma...fo/tccon_users. Note that the website is self-signed. You can safely add an exception!

1. TCCON Acknowledgement
2. Data Protocol
   2.1. Use of Data
   2.2. Reciprocity
   2.3. Warnings
3. References and Contact Information
   3.1. Park Falls, Lamont, JPL, and Caltech
   3.2. Darwin and Wollongong
   3.3. Bremen and Ny Alesund
   3.4. Bialystok
   3.5. Orleans
   3.6. Lauder
   3.7. Tsukuba
   3.8. Garmisch
   3.9. Eureka

TCCON
TCCON Data and Net...
TCCON Requirements
Data Use Policy
Auxiliary Data
Data Description
Data Description GGG...
Data Description GG...
FAQs
Proposed change to ...
TCCON Charter

log in

$\sum$ **Caltech Data**

= Institutional Repository using Invenio
= World of Legal Interop challenges!

News   Resources ▾   Projects ▾   Activities ▾   Cooperation   API & Tools ▾   About ▾   Search

Today's new taxa

New material citations

Articles: 19762
Treatments: 191347
Occurrences: 18474
Observation Records: 87839
Observation Records Geo-referenced: 33201

Search TreatmentBank

Search	🔍

Hit enter to submit fulltext search

Extended search

News

07. March 2017

"Copyright and the Use of Images as Biodiversity Data" published

There is considerable uncertainty within the taxonomic community as to how to re-use images that were included in taxonomic publications, especially in regard to whether copyright applies. This article deals with the principles...[more]

Events

17. November 2016

April 7, 2017. 09:00-10.30 RDA 9th Plenary, Barcelona, Spain

On April 7, CalTech and Plazi will demonstrate their applications of the RDA/CODATA Legal Interoperability Principles in their repositories at CalTech and Biodiversity Literature Repository at Zenodo/CERN respectively. Both...[more]

Plazi is a small organization with broad expertise and track record

Life sciences

Legal affairs

Library sciences

Computer sciences

Semantic enhancement

Publishing

Organizing scientific meetings

Biodiversity Literature Repository

Organizers

Community: 22 contributors
DOI recipients: 4 journals
Number of journals: 960

Goal: enriching publications and included data by making it findable, accessible, interoperable, and re-usable

http://biolitrepo.org

BLR: Long tail biodiversity data at Zenodo

	Zenodo	BLR	BLR contribution
Images total	108,643	107,772	99%
Images open	108,572	107,741	99%
Images % open	100%	100%	
Publications all	90,195	21,621	24%
Publications open	43,479	5,938	14%
Publications % open	48%	27%	

(Received: March 31, 2017)

Plazi workflow: Open Access through text and data mining

Search

Upload   Communities
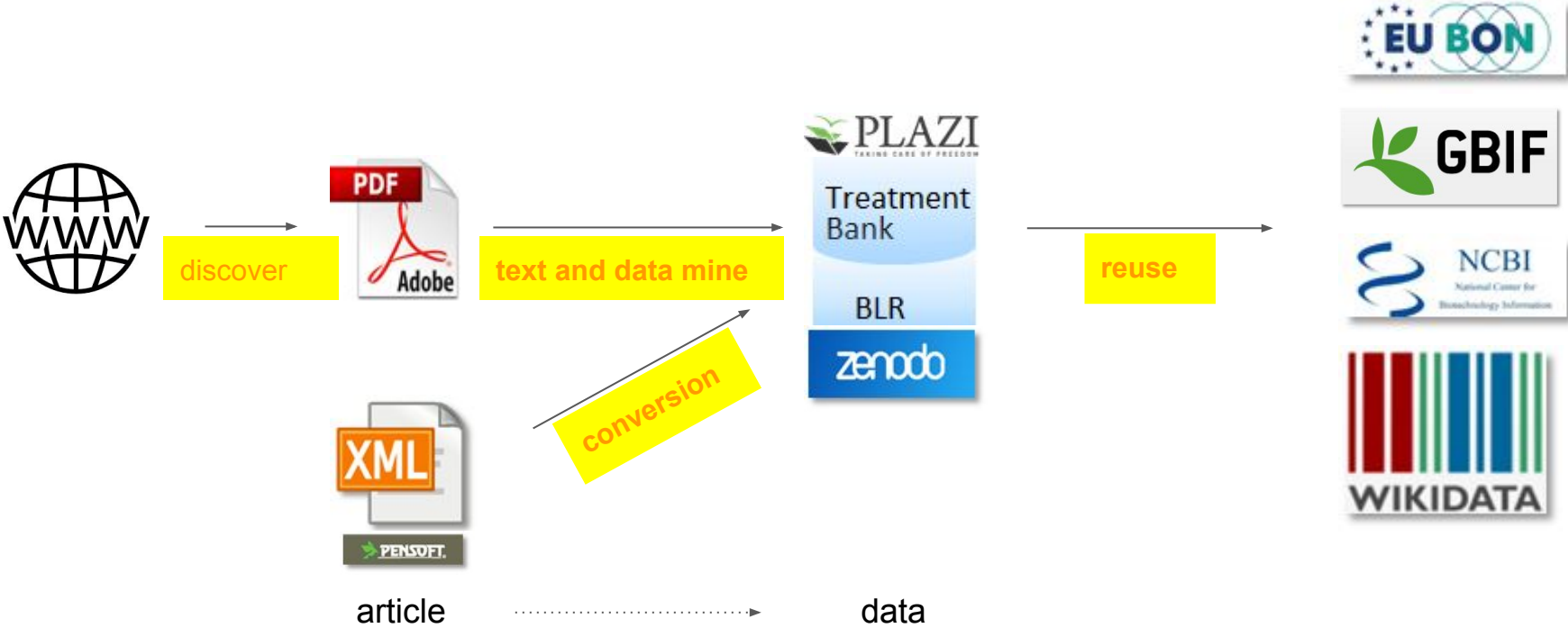
Access

December 31, 2017

Journal article   Closed Access

Three new species of Ecuadorian Rhinotragini (Coleoptera, Cerambycidae)

Bezark, Larry G.; Santos-Silva, Antonio

Bezark, Larry G., Santos-Silva, Antonio (2017): Three new species of Ecuadorian Rhinotragini (Coleoptera, Cerambycidae). Zootaxa 4247 (5): 585-592, DOI: https://doi.org/10.11646/zootaxa.4247.5.5

Files

🔒 Closed Access

Files are not publicly accessible.

Publication date:
December 31, 2017

DOI:
DOI 10.11646/zootaxa.4247.5.5

Original article DOI

Keyword(s):
Biodiversity Taxonomy Animalia Arthropoda
Insecta Coleoptera Cerambycidae

Keywords

Published in:
Zootaxa: 4247 pp. 585-592.

Related identifiers:
Has part:
http://treatment.plazi.org/id/03CE9B411F78FD13
FF1D5198FEFD595B,
http://treatment.plazi.org/id/03CE9B411F7AFD15
FF1D558AFB7D5E32,
http://treatment.plazi.org/id/03CE9B411F7CFD17
FF1D5222FB395CF7, 10.5281/zenodo.439186,
urn:lsid:plazi.org:fig:2-187-1392-196-1924 (LSID),
10.5281/zenodo.439187, *urn:lsid:plazi.org:fig:6-
167-1419-561-1775* (LSID)

Taxonomic treatment

Image

Alternate identifiers:
*urn:lsid:zoobank.org:pub:427F743A-5286-4023-
80A6-66659E5B7B49* (LSID)

Alternative identifier

Communities:
Biodiversity Literature Repository

Community

Share

Cite as

Bezark, Larry G., & Santos-Silva, Antonio. (2017).
Three new species of Ecuadorian Rhinotragini
(Coleoptera, Cerambycidae). Zootaxa, 4247(5), 585–
592. http://doi.org/10.11646/zootaxa.4247.5.5

Start typing a citation style...

Export

BibTeX CSL DataCite Dublin Core JSON
MARCXML Mendeley

zenodo

Search

Upload   Communities   Sign up

Access

December 31, 2017   Figure   Open Access

FIGURES 9 – 12 in Three new species of Ecuadorian Rhinotragini (Coleoptera, Cerambycidae)

Bezark, Larry G.; Santos-Silva, Antonio

FIGURES 9 – 12. Lygrocharis setosus sp. nov., holotype female: 9. Dorsal habitus; 10. Ventral habitus; 11. Lateral hatibus; 12. Head, frontal view.
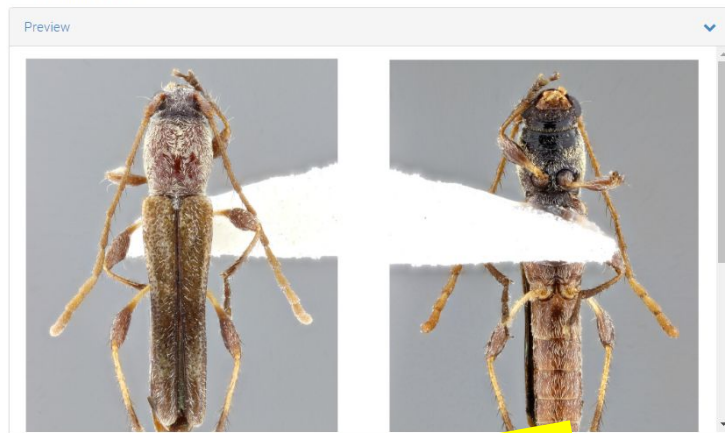
Preview

Files

Name

figure.png   5.2 MB   Preview   Download
md5:a1b763f6de4ab303b7b96674e0f35c2b

Source file

Publication date:
December 31, 2017
DOI:
DOI 10.5281/zenodo.439187
Keyword(s):
Biodiversity   Taxonomy   Animalia   Arthropoda
Insecta   Coleoptera   Cerambycidae   Lygrocharis
Related identifiers:
Cited by:
http://treatment.plazi.org/id/03CE9B411F7CFD1
FF1D5222FB395CF7
Part of:
10.11646/zootaxa.4247.5.5,
urn:lsid:plazi.org:pub:FFF7E3391F79FD10FF8A5
5FFF25D27 (LSID)
Communities:
Biodiversity Literature Repository
License (for files):
License Not Specified

DataCite DOI from Zenodo

Keywords

Taxonomic treatment

Journal article DOI

Community

Licence

Share

Cite as

Bezark, Larry G., & Santos-Silva, Antonio. (2017, December 31). FIGURES 9 – 12 in Three new species of Ecuadorian Rhinotragini (Coleoptera, Cerambycidae). Zenodo. http://doi.org/10.5281/zenodo.439187

Start typing a citation style...

Export

BibTeX   CSL   DataCite   Dublin Core   JSON
MARCXML   Mendeley

Export

$$\sum$$ **BLR Data**

= Community Repository using Invenio
= Opening up a huge corpus of imprisoned data from
    scientific publications
= Enriching publications and its data by linking the data and
    publication, making it findable, accessible, inter-operable
    and re-usable

3 Comparisons between Cases

Principle One:
Facilitate the lawful access to and reuse of research data

- Open access and unrestricted access and reuse

- Place data in public domain

- Apply rights waivers e.g., CC0 or PDDL

- Rights assertions and restrictions should make data available equitably, espec to disadvantaged

P1: Facilitate the lawful access to and reuse of research data

Open and unrestricted access to the over 105,000 scientific illustrations extracted from publications, as well as taxonomic treatments of species

Items with closed access remain closed

Items published on or before Dec 31 1999 are read-only access

Select types of data provided are considered to be ineligible for copyright protection (Patterson et al, 2015: The Blue List http://plazi.org/activities/blue-list/   ).

Conditions of reuse depend on the applicable national copyright law

Access to self-archived data, software, and other research outputs from Caltech that may (or may not be) associated with publications and theses

Items with closed access not accepted by repository

Ownership of provided data governed by Caltech IP Policy

All releases must be accompanied by required disclaimers of responsibility

Waivers or Licenses for reuse accompany each data set.
CC0 waiver is the default.
CC-By, CC-BY-NC Licenses also available

Conditions of reuse depend on Caltech IP Policy and US national copyright law, including user exceptions e.g. Fair Use

Principle Two:
Determine the rights to and responsibilities for the data

- Establish rightsholder + rights before dissemination

- Abide by applicable rights

- Participate in law/policy that opens access to data

- Design and deliver training on rights and responsibilities with research data

P2: Determine the rights to and responsibilities for the data

The data source are exclusively publications. Publications *per se* are made accessible depending on their licences.

Metadata is exposed and enhanced with links to the exposed data.

BLR has its domicile in Switzerland and is ruled by Swiss law. Note that E.U. database protection is not applicable in Switzerland.

BLR is publishing their legal arguments and makes them as widely accessible as possible.

The BLR community is involved in lecturing on providing open access to scientific research results

Content in CaltechData generally falls under the Institute Copyright and Software Policy

For content that may be non-Caltech in origin, the depositor verifies that they have the rights to license the content and identifies/acknowledges owner(s).

The Library coordinates w/ Office of General Counsel; Research Compliance Office; Office of Technology Transfer to align repository terms and conditions.

Caltech Library has a non-exclusive, worldwide, perpetual license to use, duplicate and distribute the deposited content, and to transfer the content to any format or medium now known or later developed for archiving, preservation and access purposes.

Depositor retains right to publish and/or share deposited content as provided by the Institute Copyright and Software Policy.

Principle Three: Balance the Legal Interests

- Waive rights in publicly funded data where possible

- Limit time embargoes

P3: Balance the Legal Interests

Zenodo is publicly funded and supports the approach by the BLR community.

BLR is supported by a not for profit organization, a commercial publisher and various domain specific publishers such as scientific associations. All BLR data is openly accessible.

CaltechDATA supports embargos in order to encourage data deposit and support privacy while a publication is in preparation. However, it not an indefinite embargo and the researcher is responsible for complying with funder, publisher, or institutional requirements.

The repository will restrict access to the data until the end of the embargo period; at which time, the content will become publicly available automatically.

Users assume responsibility for complying with funder, publisher, or organizational embargo requirements (such as the Graduate School policy on thesis embargo).

Principle Four: State the Rights Clearly and Transparently

- Use standard electronic rights statements

- Consult qualified legal counsel to clarify rights

- Rightsholders inform users about terms and conditions of use

P4: State the Rights Clearly and Transparently

BLR is making use of Zenodo's rights statements.

BLR is supported by a not for profit organization, a commercial publisher and various domain specific publishers such as scientific associations. All BLR data is openly accessible.

BLR includes legal counsels to accompany the data extraction workflow, and the data policy of BLR and presents the findings in public (Agosti and Egloff, 2009; Patterson et al., 2014, Egloff et al., 2017).

Users are informed about their rights on the front page of BLR.

The license or waiver name is reported as part of the data record and metadata

A future goal is to include license information in the metadata deposited with the DOI registration agency (generally DataCite) for a Caltech-created dataset.

Caltech community members can access the Office of General Counsel for legal advice about data rights.

CaltechDATA does not yet have a standard mechanism for supporting special terms and conditions of use. We want to encourage researchers to select standard licenses.

Principle Five: Promote harmonization of rights in research data

- Leverage user rights in applicable statutes

- Devise workflows and best practices as useful tools to promote harmonization

P5: Promote harmonization of rights in research data

Plazi is using the BLR to demonstrate the power of data sharing to funding agencies and law makers.

The data extraction workflow has been described by Agosti & Egloff, 2009; a taxononomy of data and their legal status has been compiled in the Blue List (Patterson et al., 2014) and the augmented list (Egloff et al., 2017). BLR promotes harmonization by encouraging data providers to use copyright waivers.

By limiting custom rights statements, the CaltechDATA repository encourages users to adopt standard licenses for their data.

Caltech Library applies community norms, e.g., the Association of Research Libraries' Code of Best Practices in Fair Use -- where warranted -- to make accessible via campus network any publications associated with the deposited dataset that are necessary to understand or interpret that data.

Principle Six: Provide proper attribution and credit for research data

◉ Establish and follow normative practices for giving credit where due

P6: Provide proper attribution and credit for research data

All data in BLR are linked to their source. Attribution is therefore guaranteed.

As follow-up of discussions in presentations on BLR, a special chapter on attribution has been included in the publication on copyright on scientific illustration (Egloff et al. 2017), accompanying BLR.

By introducing DOIs for images extracted from articles, we explicitly want to highlight the value of citing subarticle elements.

Caltech Library encourages and provides training in data attrbiution through Data and AuthorCarpentry workshops as well as a program on Giving Credit Where Due offered for advisors, instructors, and students

 We are pursuing offering of example citations as part of the data record landing page, based on the best practices promoted by Fenner et al (2016, http://biorxiv.org/content/early/2016/12/28/097196) "**A Data Citation Roadmap for Scholarly Data Repositories"** but this is not currently implemented.

Thanks!

Any **questions** ?

You can find us at

gclement@library.caltech.edu

agosti@plazi.org

egloff_bader@bluewin.ch

tmorrell@library.caltech.edu