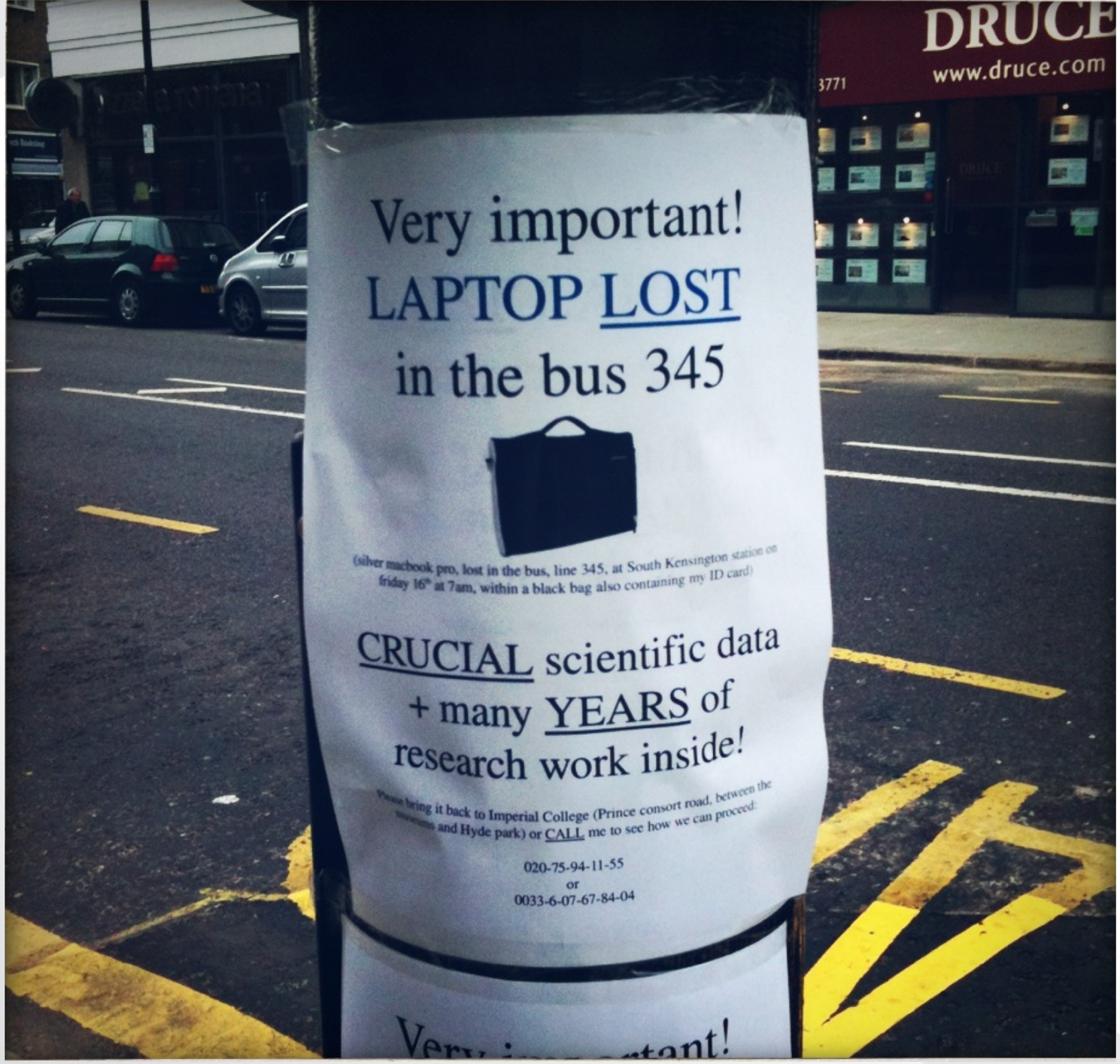
Lars Holm Nielsen
CERN, IT Department


# FAIR data in a generic data repository




Linking Open Science in Austria, Vienna, April 24, 2019

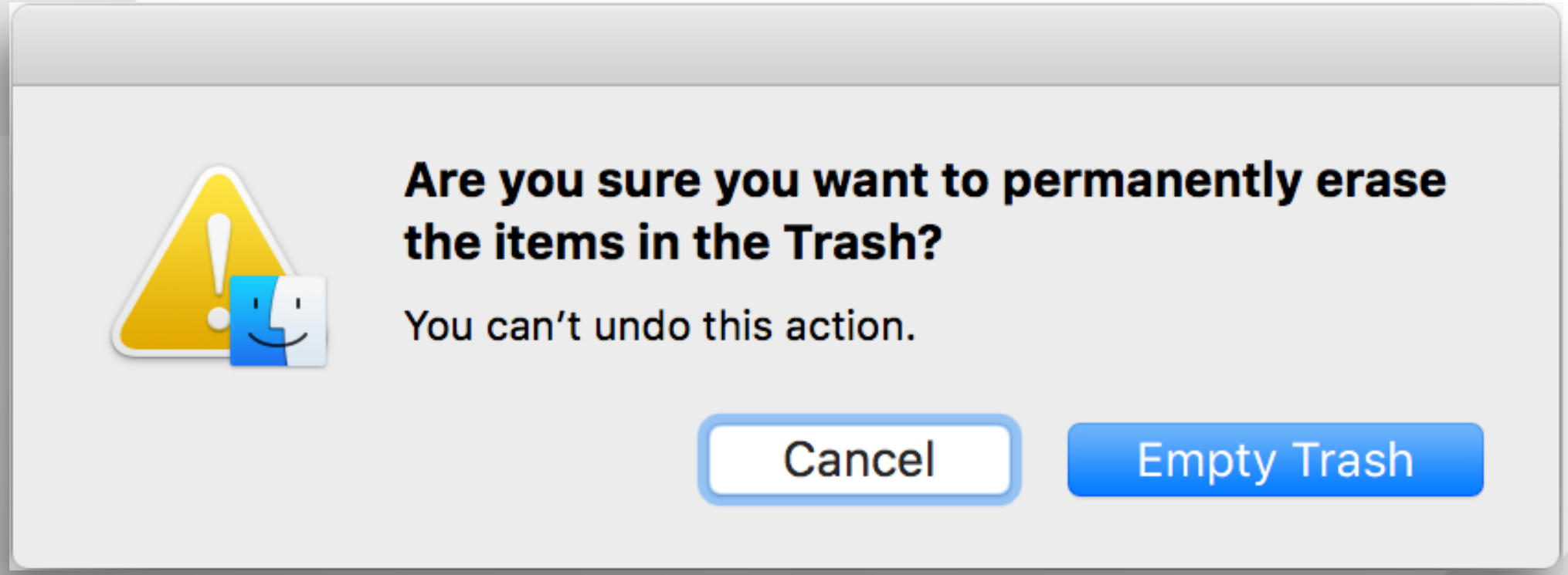



**Credit:** Dave Hill, CC-BY-NC-SA 2.0 Generic. https://www.flickr.com/photos/dmh650/4031607067/in/gallery-wlef70-72157633022909105/

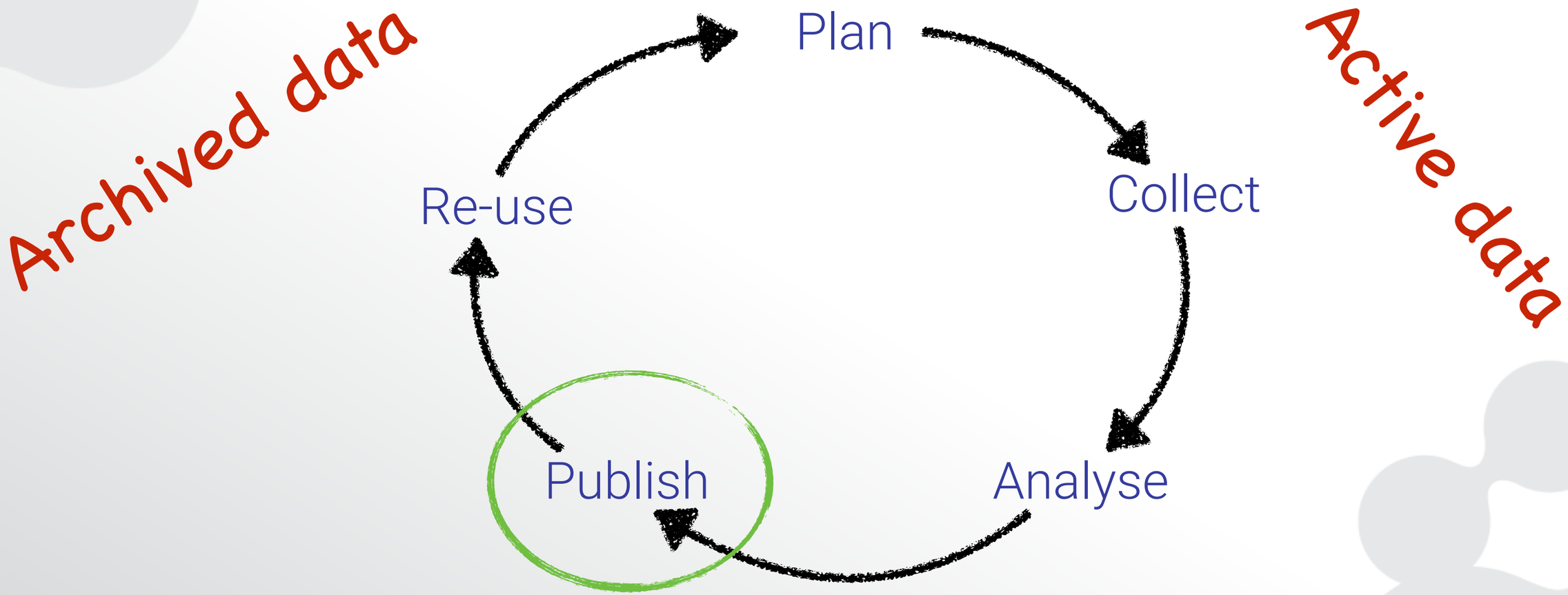

**Credit:** By Brian Herzog / licensed under Creative Commons Attribution-NonCommercial-ShareAlike 2.0 — https://www.flickr.com/photos/herzogbr/6756173595/in/gallery-wlef70-72157633022909105/

**Credit:** By Bryan Tong Minh / CC-BY-2.5 (http://commons.wikimedia.org/wiki/File:Brand_bouwkunde_-_TU_Delft_-_13_Mei_2008.jpg)

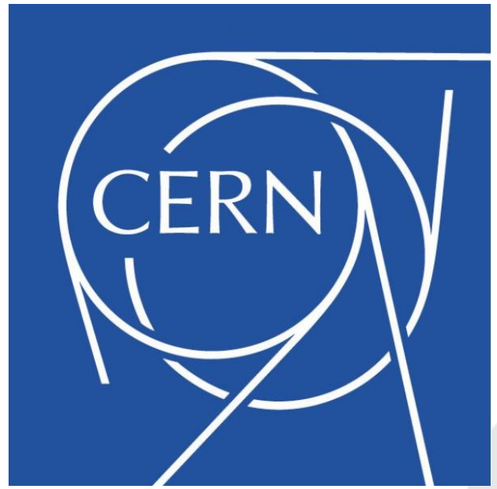

Are you sure you want to permanently erase the items in the Trash?
You can’t undo this action.
Cancel
Empty Trash

# **Don’t** do it **yourself**!

# Research data lifecycle



# Funded by



Alfred P. Sloan
FOUNDATION

# Behind Zenodo

## CERN Data Centre

- ~300PB disk
- ~400PB tape
- ~110k CPUs

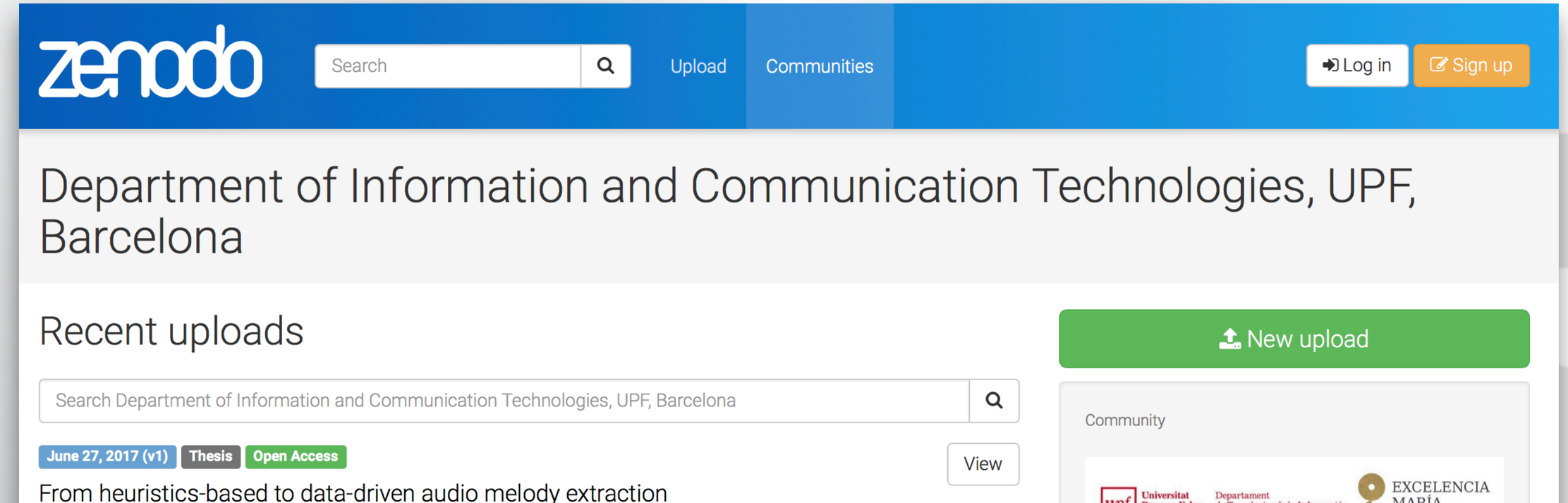
## Services

- Digital repositories
- Data preservation for High-Energy Physics.

INVENIO

# Zenodo

# Upload Describe Publish



# Department of Information and Communication Technologies, UPF, Barcelona

## Recent uploads

Search Department of Information and Communication Technologies, UPF, Barcelona

June 27, 2017 (v1) Thesis Open Access

View

### From heuristics-based to data-driven audio melody extraction

Bosch, Juan J.;

The identification of the melody from a music recording is a relatively easy task for humans, but very challenging for computational systems. This task is known as "audio melody extraction", more formally defined as the automatic estimation of the pitch sequence of the melody directly from

Uploaded on December 20, 2017

November 14, 2017 (v2) Thesis Open Access

View

### Knowledge Extraction and Representation Learning for Music Recommendation and Classification

Oramas, Sergio;

In this thesis, we address the problems of classifying and recommending music present in large collections. We focus on

New upload

Community

upf. Universitat Pompeu Fabra Barcelona — Departament de Tecnologies de la Informació i les Comunicacions — EXCELENCIA MARÍA DE MAEZTU

### Department of Information and Communication Technologies, UPF, Barcelona

Department of Information and Communication Technologies. Maria de Maeztu (MdM) Unit of Excellence. UPF, Barcelona

Maria de Maeztu Unit of Excellence -

MdM Strategic Research Program on data-driven knowledge extraction

# How can a generic data repository provide services for FAIR data?

# Case study: Taxonomic treatments

## Describe the discovery of new biological species

**Example:**

Journal article describing 22 new millipedes, published in European Journal of Taxonomy

European Journal of Taxonomy 445: 1–90
https://doi.org/10.5852/ejt.2018.445
ISSN 2118-9773
www.europeanjournaloftaxonomy.eu
2018 · Enghoff H.



**Monograph**

urn:lsid:zoobank.org:pub:852A3F68-B728-413A-B12E-56F306D56C35

**A mountain of millipedes VII: The genus *Eviulisoma* Silvestri, 1910, in the Udzungwa Mountains, Tanzania, and related species from other Eastern Arc Mountains. With notes on *Eoseviulisoma* Brolemann, 1920, and *Suohelisoma* Hoffman, 1963 (Diplopoda, Polydesmida, Paradoxosomatidae)**

Henrik ENGHOFF

Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, DK-2100 København Ø, Denmark.

Email: henghoff@snm.ku.dk

urn:lsid:zoobank.org:author:FB09A817-000D-43C3-BCC4-2BC1E5373635

**Abstract.** Twenty-two new species of the genus *Eviulisoma* Silvestri, 1910, from the Eastern Arc Mountains, Tanzania, are described: *E. acaciae* sp. nov., *E. aequilobatum* sp. nov., *E. akkariae* sp. nov., *E. angulatum* sp. nov., *E. articulatum* sp. nov., *E. biquintum* sp. nov., *E. breviscutum* sp. nov., *E. cetafi* sp. nov., *E. chitense* sp. nov., *E. commelina* sp. nov., *E. coxale* sp. nov., *E. ejti* sp. nov., *E. grumslingslak* sp. nov., *E. kalimbasiense* sp. nov., *E. navuncus* sp. nov., *E. nessiteras* sp. nov., *E. ottokrausi* sp. nov., *E. paradisiacum* sp. nov., *E. sternale* sp. nov. and *E. zebra* sp. nov. from the Udzungwa Mts, *E. culter* sp. nov. from the Rubeho Mts and *E. kangense* sp. nov. from the Kanga Mts. *Eviulisoma kwabuniense* Kraus, 1958, and *E. dabagaense* Kraus, 1958, both from the Udzungwa Mts, are redesribed based on new material. Notes are provided on *E. iuloideum* (Verhoeff, 1941) based on type material. *Eoseviulisoma* Brolemann, 1920, is synonymized under *Eviulisoma*, based on newly collected material of *E. julinum* (Attems, 1909), type species of *Eoseviulisoma*. New material of *Suohelisoma ulugurense* Hoffman, 1964, type species of *Suohelisoma* Hoffman, 1964, has revealed that the gonopod structure is more similar to that of *Eviulisoma* than originally thought, but *Suohelisoma* is retained as a valid genus. Four species groups are recognized among *Eviulisoma* species from the Udzungwa Mts, but the need for a revision of the entire genus is emphasized. Two types of epizootic fungi are recorded from *Eviulisoma* spp., and an enigmatic amorphous mass, which may be a kind of plugging substance, is recorded from the gonopod tips and excavated sixth sternum of several species.

**Keywords.** Taxonomy, new species, epizootic fungi, copulatory plug.

Enghoff H. 2018. A mountain of millipedes VII: The genus *Eviulisoma* Silvestri, 1910, in the Udzungwa Mountains, Tanzania, and related species from other Eastern Arc Mountains. With notes on *Eoseviulisoma* Brolemann, 1920, and *Suohelisoma* Hoffman, 1963 (Diplopoda, Polydesmida, Paradoxosomatidae). *European Journal of Taxonomy* 445: 1–90. https://doi.org/10.5852/ejt.2018.445



**Enghoff H. 2018.** A mountain of millipedes VII: The genus *Eviulisoma* Silvestri, 1910, in the Udzungwa Mountains, Tanzania, and related species from other Eastern Arc Mountains. With notes on *Eoseviulisoma* Brolemann, 1920, and *Suohelisoma* Hoffman, 1963 (Diplopoda, Polydesmida, Paradoxosomatidae). *European Journal of Taxonomy* 445: 1–90. https://doi.org/10.5852/ejt.2018.445

# Treatments: Data in disguise

***Eviulisoma breviscutum*** sp. nov.
urn:lsid:zoobank.org:act:D7C4195B-37DF-4B02-BD3B-4447DBCBB23C
Fig. 36

**Diagnosis**

Differs from other Udzungwan species of *Eviulisoma* by the combination of unmodified sterna 5 and 6 and a very short *map* (ca half as long as solenophore).

**Etymology**

The name is a noun in apposition meaning 'short shield' and refers to the short, shield-like mesal acropodital process.

**Material** (total: 3 ♂♂)

**Holotype**

TANZANIA: ♂, Mwanihana Forest, above Sanje, 1650 m a.s.l., pitfall trap, 18 Aug. 1982, M. Stoltze and N. Scharff leg. (ZMUC).

**Paratypes**

TANZANIA: 1 ♂, Morogoro Region, Kilombero District, Udzungwa Mts National Park, forest below Mwanihana Peak, 7°49′ S, 36°50′ E, 1800 m a.s.l., sifted from leaf litter, 20 Aug. 2017, T. Pape leg. (ZMUC); 1 ♂, Morogoro Region, Udzungwa Mts National Park, Mito Mitatu, above Mang'ula,

# Treatments: Material

Geographic coordinates

Date collected

Collector

**Material** (total: 3 ♂♂)

**Holotype**

TANZANIA: ♂, Mwanihana Forest, above Sanje, 1650 m a.s.l., pitfall trap, 18 Aug. 1982, M. Stoltze and N. Scharff leg. (ZMUC).

**Paratypes**

TANZANIA: 1 ♂, Morogoro Region, Kilombero District, Udzun
Mwanihana Peak, 7°49′ S, 36°50′ E, 1800 m a.s.l., sifted from l
(ZMUC); 1 ♂, Morogoro Region, Udzungwa Mts National P
07°49′3″ S, 36°52′58″ E, 1487 m a.s.l., 16 Dec. 2016, T. Pape and

Collection code

**Enghoff H. 2018.** A mountain of millipedes VII: The genus *Eviulisoma* Silvestri, 1910, in the Udzungwa Mountains, Tanzania
*Eoseviulisoma* Brolemann, 1920, and *Suohelisoma* Hoffman, 1963 (Diplopoda, Polydesmida, Paradoxosomatidae). *Europea*

*EUROPEAN JOURNAL OF TAXONOMY*

MATERIAL CITATIONS FORMATTING GUIDE

In accordance with the European Journal of Taxonomy's FAIR & Open Science policy, the formatting gui
entomology, zoology and palaeozoology material citations is provided below (guidelines for botany ava
soon). Authors are encouraged to prepare their manuscripts according to this model prior to submissio
they will also be given the opportunity to comply upon acceptance of the article.

While EJT strongly recommends that authors adhere to the guidelines given below, the fine-grain formatt
the material citations is not compulsory: if an author decides not to comply or the material is not approp
EJT will perform reduced formatting during production. In this case, the majority of the material citatio
not be tagged, extracted or disseminated to scientific databases.

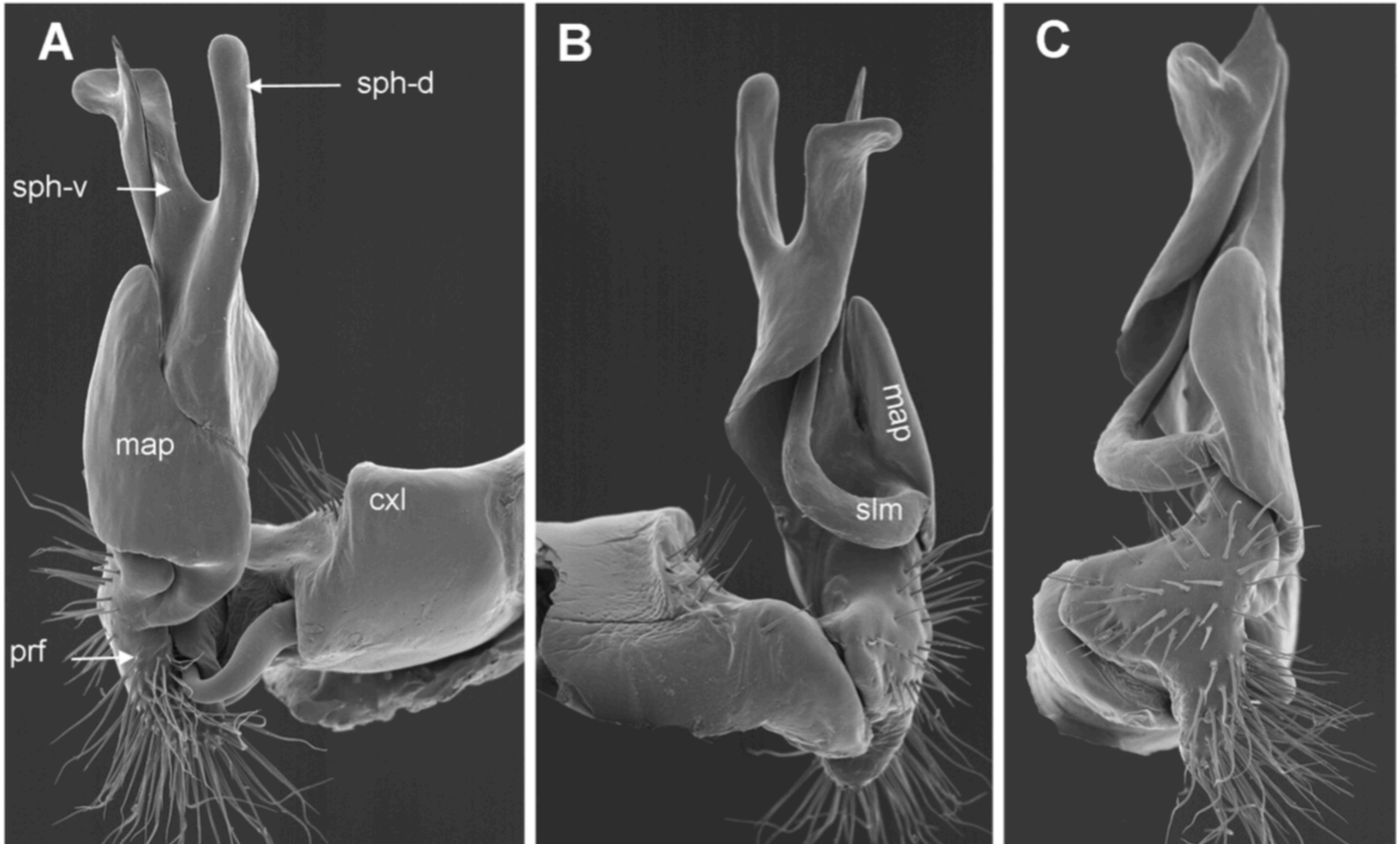
## GENERAL PRESENTATION

### Order

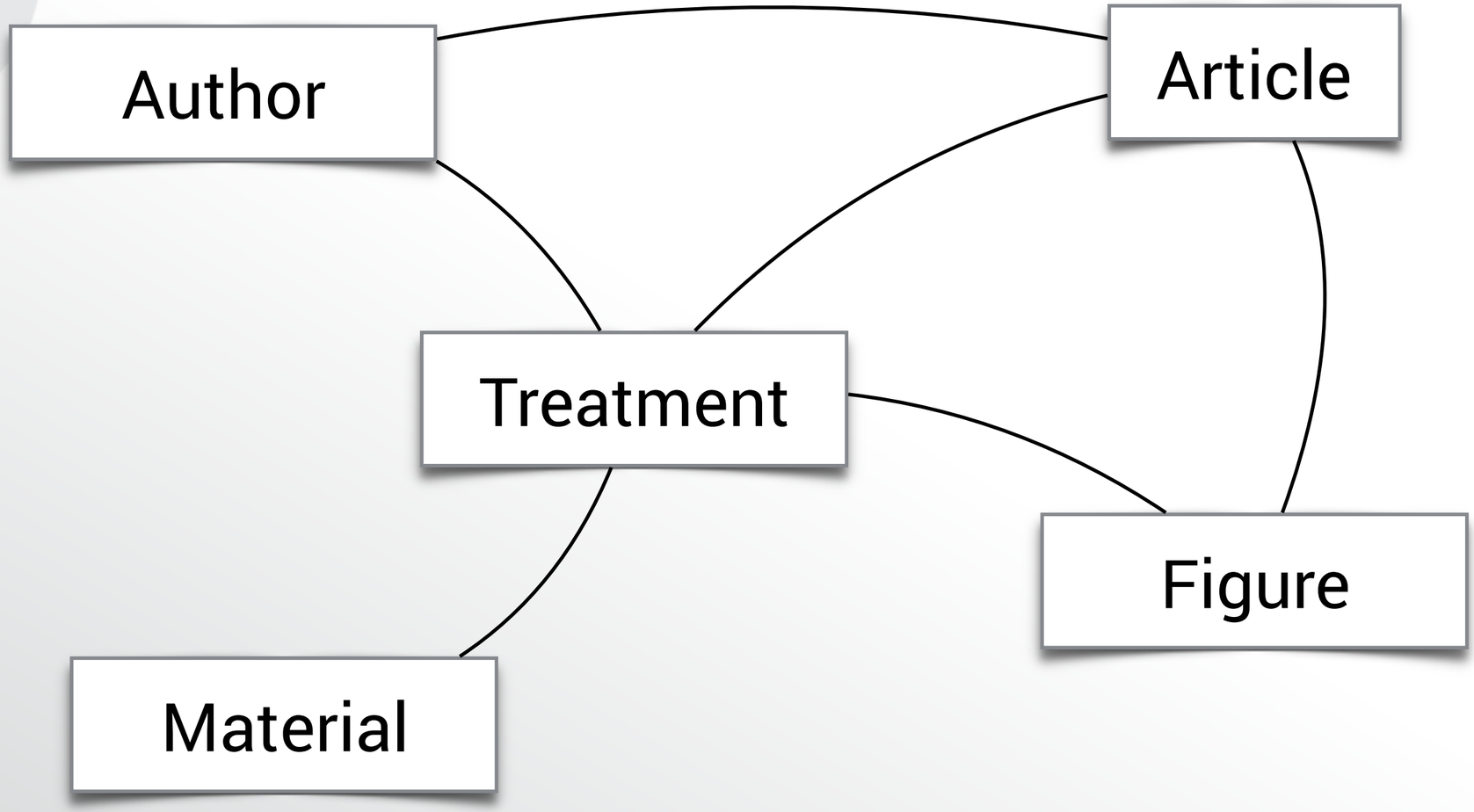
Each material citation is composed of diverse data fields (number of specimens, locality, date collected,
that EJT identifies using Darwin Core (DWC) terms. To efficiently perform this, it is important to ensure tha
different fields of a material citation are consistently presented in the same order throughout the article
the very least, within a taxon treatment.

The preferred order of data fields is as follows:

# Treatments: Figures



# Treatments: Relations



# Locked up data

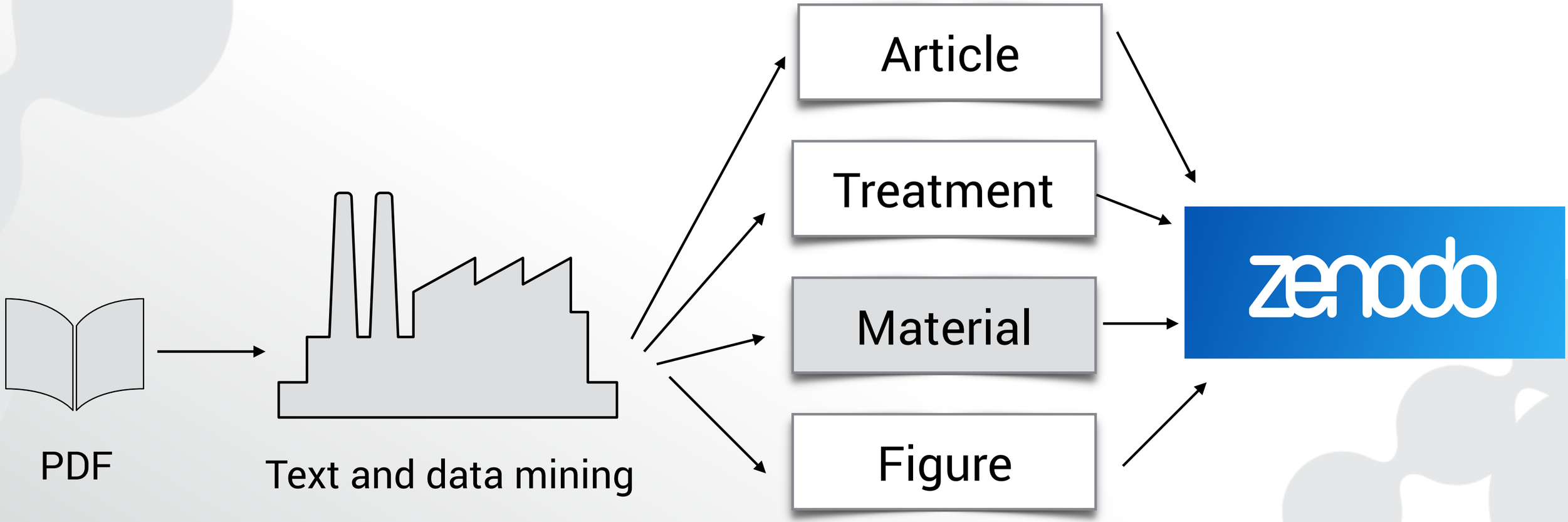
**Unanswerable questions:**

- How many species have been described by my collection?
- Give me a list of all new species?
- Retrieve all images for a given taxon?
- What's know about a geographic region?

- Treatments:
  - Past 260 years: **~10+ millions** published
  - Every year: **~17k** new **/ ~130k** augmented

# Unlocking FAIR treatments?

# Overview of process




PLAZI
TAKING CARE OF FREEDOM

# Example




December 31, 2018

Figure
Open Access

# Fig. 7 in A mountain of millipedes VI. New records, new species, a new genus and a general discussion of Odontopygidae from the Udzungwa Mts, Tanzania (Diplopoda, Spirostreptida, Odontopygidae)

Enghoff, Henrik

Fig. 7. Hoffmanides dissutus (Hoffman, 1963), ♂, from Udzungwa Mts National Park. Photograph by A. Illum. Scale bar = 5 mm.

Preview

Part of

Biodiversity Literature Repository

Indexed in

OpenAIRE

**Publication date:**
December 31, 2018

**DOI:**
DOI 10.5281/zenodo.1146170

**Keyword(s):**
Biodiversity
Taxonomy
Animalia
Arthropoda
Diplopoda
Spirostreptida
Odontopygidae
Hoffmanides

**Published in:**
European Journal of Taxonomy: 394 pp. 1-29.

**Related identifiers:**
Cited by:
http://treatment.plazi.org/id/038D2864FFA3FFA2FDA4FEBA88B9FCD9

Part of:
10.5852/ejt.2018.394,
*urn:lsid:plazi.org:pub:FFB4501CFFB2FFB1FF94FFB0892EFF8E* (LSID), https://zenodo.org/record/1146158, 10.5852/ejt.2018.394, https://zenodo.org/record/1146158

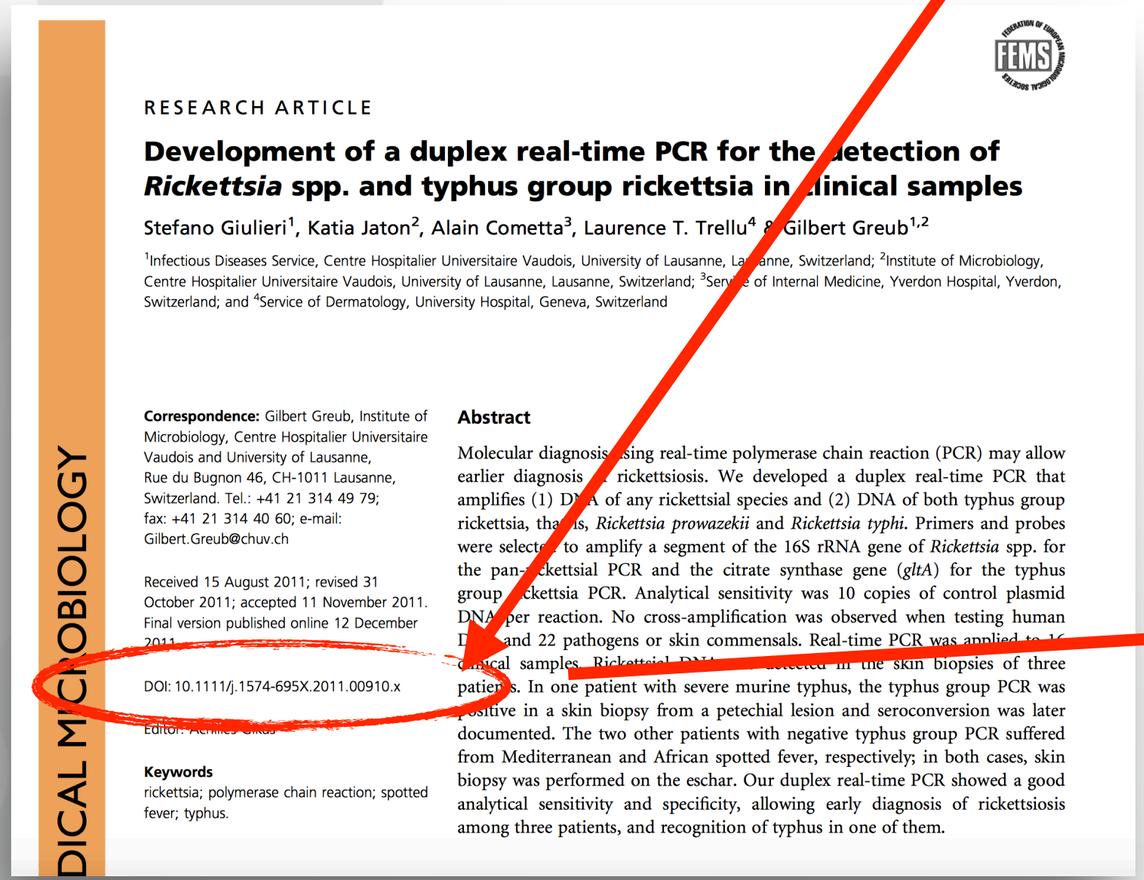
**Communities:**
Biodiversity Literature Repository

# Key features provided

- **Persistent identifier** management and registration
- Widely adopted **metadata schemas** (DataCite, JSON-LD, …)
- **Indexing and registration** in searchable resources

**F**indable **A**ccessible

# Digital Object Identifier



DICAL MICROBIOLOGY

RESEARCH ARTICLE

**Development of a duplex real-time PCR for the detection of *Rickettsia* spp. and typhus group rickettsia in clinical samples**

Stefano Giulieri[1], Katia Jaton[2], Alain Cometta[3], Laurence T. Trellu[4] & Gilbert Greub[1,2]

[1]Infectious Diseases Service, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland; [2]Institute of Microbiology, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland; [3]Service of Internal Medicine, Yverdon Hospital, Yverdon, Switzerland; and [4]Service of Dermatology, University Hospital, Geneva, Switzerland

**Correspondence:** Gilbert Greub, Institute of Microbiology, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Rue du Bugnon 46, CH-1011 Lausanne, Switzerland. Tel.: +41 21 314 49 79; fax: +41 21 314 40 60; e-mail: Gilbert.Greub@chuv.ch

Received 15 August 2011; revised 31 October 2011; accepted 11 November 2011. Final version published online 12 December 2011.

DOI: 10.1111/j.1574-695X.2011.00910.x

Editor: Alexander Sikkes

**Keywords**
rickettsia; polymerase chain reaction; spotted fever; typhus.

**Abstract**

Molecular diagnosis using real-time polymerase chain reaction (PCR) may allow earlier diagnosis of rickettsiosis. We developed a duplex real-time PCR that amplifies (1) DNA of any rickettsial species and (2) DNA of both typhus group rickettsia, thanks, *Rickettsia prowazekii* and *Rickettsia typhi*. Primers and probes were selected to amplify a segment of the 16S rRNA gene of *Rickettsia* spp. for the pan-rickettsial PCR and the citrate synthase gene (*gltA*) for the typhus group rickettsia PCR. Analytical sensitivity was 10 copies of control plasmid DNA per reaction. No cross-amplification was observed when testing human DNA and 22 pathogens or skin commensals. Real-time PCR was applied to 16 clinical samples. Rickettsial DNA was detected in the skin biopsies of three patients. In one patient with severe murine typhus, the typhus group PCR was positive in a skin biopsy from a petechial lesion and seroconversion was later documented. The two other patients with negative typhus group PCR suffered from Mediterranean and African spotted fever, respectively; in both cases, skin biopsy was performed on the eschar. Our duplex real-time PCR showed a good analytical sensitivity and specificity, allowing early diagnosis of rickettsiosis among three patients, and recognition of typhus in one of them.

**Credit:** Screenshot of article (free to read) Copyright 2011 Federation of European Microbiological Societies.

## References

1. Walker DH (2007) Rickettsiae and Rickettsial Infections: The Current State of Knowledge. Clin Infec Dis 45: S39–S44.
2. Bavaro MF, Kelly DJ, Dasch DA, Hale BR, Olson P (2005) History of U. S. Military Contributions to the study of Rickettsial Diseases. Military Medicine 170:49–60. PMID: 15916283
3. Kelly DJ, Richards AL, Temenak J, Strickman D, Dasch GA (2002) The past and present threat of rickettsial diseases to military medicine and international public health, Clinical Infect. Dis. 34: S145–S169.
4. Rolain JM, Jensenius M, Raoult D (2004) Rickettsial infections—a threat to travellers? Curr Opin Infect Dis 17:433–7. PMID: 15353963
5. Kovacova E, Kazar J (2000) Rickettsial diseases and their serological diagnosis. Clin Lab 46:239–45. PMID: 10853230
6. La Scola B, Raoult D (1997) Laboratory diagnosis of rickettsioses: current approaches to diagnosis of old and new rickettsial diseases. J Clin Microbiol 35:2715–27. PMID: 9350721
7. Renvoisé A, Rolain JM, Socolovschi C, Raoult D (2012) Widespread use of real-time PCR for rickettsial diagnosis. FEMS Immunol Med Microbiol 64:126–9. doi: 10.1111/j.1574-695X.2011.00899.x PMID: 22092999
8. Giulieri S, Jaton K, Cometta A, Trellu LT, Greub G (2012) Development of a duplex real-time PCR for the detection of Rickettsia spp. and typhus group rickettsia in clinical samples. FEMS Immunol Med Microbiol 64:92–7. doi: 10.1111/j.1574-695X.2011.00910.x PMID: 22098502
9. Prakash JAJ, Reller ME, Barat N, Bunler JS (2009) Assessment of a quantitative multiplex 5' nuclease real-time PCR for spotted fever and typhus group rickettsioses and *Orientia tsutsugamushi*. Clin Microbiol Infect Suppl. 2:292–3.
10. Jiang J, Chan TC, Temenak JJ, Dasch GA, Ching WM, et al. (2004) Development of a quantitative real-time polymerase chain reaction assay specific for *Orientia tsutsugamushi*. Am J Trop Med Hyg 70: 351–6. PMID: 15100446
11. Paris DH, Aukkanit N, Jenjaroen K, Blacksell SD, Day NP (2009) A highly sensitive quantitative real-time PCR assay based on the *groEL* gene of contemporary 339 Thai strains of *Orientia tsutsugamushi*. Clin Microbiol Infect 15:488–495. doi: 10.1111/j.1469-0691.2008.02671.x PMID: 19416296
12. Roux V, Rydkina E, Eremeeva M, Raoult D (1997) Citrate synthase gene comparison, a new tool for phylogenetic analysis, and its application for the *rickettsiae*. Int J Syst Bacteriol 47: 252–261. PMID: 9103608

**Credit:** Reference list from https://doi.org/10.1371/journal.pntd.0003884

# Digital Object Identifier



**Credit:** Screenshot of https://www.ncbi.nlm.nih.gov/pubmed/22098502 obtained on 2017-05-21

# Digital Object Identifier

## Resolvable

**http://doi.org/**10.5281/zenodo.1100973

↓

https://zenodo.org/record/1100973

# Digital Object Identifier

Persistent

**http://doi.org/**10.5281/zenodo.1100973

↓

https://zenodo.org/record/1100973

# Digital Object Identifier

Globally unique

**http://doi.org/**10.5281/zenodo.1100973

# Digital Object Identifier

## Metadata

```
<?xml version="1.0" encoding="utf-8"?>
<resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://datacite.org/schema/kernel-3" xsi:schemaLoc
  <identifier identifierType="DOI">10.5281/zenodo.1100973</identifier>
  <creators>
    <creator>
      <creatorName>Oramas, Sergio</creatorName>
      <nameIdentifier nameIdentifierScheme="ORCID" schemeURI="http://orcid.org/">0000-0002-8028-2890</nameIdentifier>
      <affiliation>Universitat Pompeu Fabra</affiliation>
    </creator>
  </creators>
  <titles>
    <title>Knowledge Extraction And Representation Learning For Music Recommendation And Classification</title>
  </titles>
  <publisher>Zenodo</publisher>
  <publicationYear>2017</publicationYear>
```

# Indexing and reuse

Get data Share Tools Inside GBIF Login

**Classification**

Select a species

You are browsing: A mountain of millipedes VII: The genus Eviulisoma Silvestri, 1910, in the Udzungwa Mountains, Tanzania, and related species from other Eastern Arc Mountains. With notes on Eoseviulisoma Brolemann, 1920, and Suohelisoma Hoffman, 1963 (Diplopoda, Polydesmida, Paradoxosomatidae)

| | |
|---|---|
| Kingdom | Animalia |
| Phylum | Arthropoda |
| Class | Diplopoda |
| Order | Polydesmida |
| Family | Paradoxosomatidae |
| Genus | *Eviulisoma* |
| Species | *Eviulisoma breviscutum* Enghoff, 2018 |

No children

SPECIES | ACCEPTED

## *Eviulisoma breviscutum* Enghoff, 2018

In: Enghoff, Henrik (2018): A mountain of millipedes VII: The genus Eviulisoma Silvestri, 1910, in the Udzungwa Mountains, Tanzania, and related species from other Eastern Arc Mountains. With notes on Eoseviulisoma Brolemann, 1920, and Suohelisoma Hoffman, 1963 (Diplopoda, Polydesmida, Paradoxosomatidae). European Journal of Taxonomy 445: 1-90, DOI: https://doi.org/10.5852/ejt.2018.445

Mediated through: Plazi.org taxonomic treatments database

TREATMENT VERBATIM SOURCE

Eviulisoma breviscutumsp. nov.

urn:lsid:zoobank.org:act:D7C4195B-37DF-4B02-BD3B-4447 DBCBB 23C

Fig. 36

Diagnosis

Differs from other Udzungwan species of Eviulisoma by the combination of unmodified sterna 5 and 6 and a very short map (ca half as long as solenophore).

Etymology

The name is a noun in apposition meaning ‘short shield’ and refers to the short, shield-like mesal acropodital process.

Material (total: 3 ♂♂)

Holotype

TANZANIA : ♂ , Mwanihana Forest, above Sanje , 1650 m a.s.l. , pitfall trap , 18 Aug. 1982 , M. Stoltze and N. Scharff leg. (ZMUC).

Paratypes

TANZANIA : 1 ♂ , Morogoro Region , Kilombero District , Udzungwa Mts National Park, forest below Mwanihana Peak , 7°49' S , 36°50' E , 1800 m a.s.l. , sifted from leaf litter , 20 Aug. 2017 , T. Pape leg. ( ZMUC ); 1 ♂ , Morogoro Region , Udzungwa Mts National Park, Mito Mitatu, above Mang’ula , 07°49'3" S , 36°52'58" E , 1487 m a.s.l. , 16 Dec. 2016 , T. Pape and N. Scharff leg. (ZMUC).

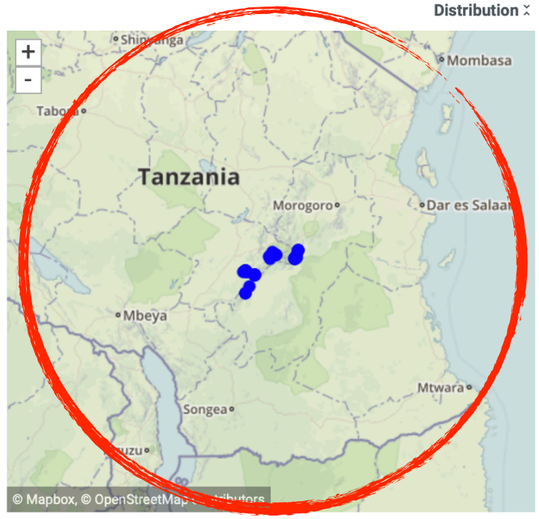
Description (male)

SIZE . Length 14–15 mm , max. width 1.5–1.6 mm .

COLOUR. After 3 months in alcohol dorsally dark brown to blackish brown, vertex and metazonites medium brown, rest of head, antennae and legs pale yellowish.

ANTENNAE. Reaching back to middle of ring 3.

**Distribution**




**Figures**


Global Biodiversity Information Facility

GBIF

# Key challenge

Generic vs domain-specific

**Material**-level vs. **treatment**-level:
Geospatial / Temporal

# Key challenge (solution)

- Map metadata to DataCite with detail loss.

- Simplified custom metadata (searchable):
  - Key-value pairs based on vocabulary (JSON-LD)
  - `http://rs.tdwg.org/dwc/terms/collectionCode`

- Domain-specific metadata files (not searchable)
  - DarwinCore XML

# How can a generic data repository provide services for FAIR data?

# Zenodo for FAIR data

- **Stable home**
  - Key for Findable & Accessible data

- **Metadata (both generic and domain specific)**
  - Key for Interoperable & Reusable data

# It takes an ecosystem





ORCID, ZooBank, Publishers, ….

Thank you!

zenodo.org

# FAIR data in action!

It takes an ecosystem of independent services

Data repositories provide a stable home

OpenAIRE

CERN

Alfred P. Sloan FOUNDATION

ARCADIA