

Data integration and provenance in the Encyclopedia of Life

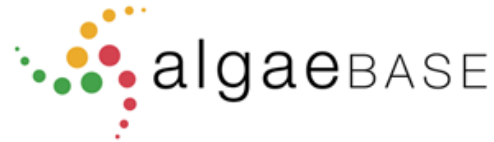


Katja Schulz
Encyclopedia of Life
Smithsonian Institution
National Museum of Natural History

EOL aggregates data from content partners



CONABIO



Smithsonian



HARVARD UNIVERSITY



FishBase



WIKIMEDIA COMMONS



Biodiversity Heritage Library



The Paleobiology Database
revealing the history of life



► Taxonomy

Taxon recognized by [AntWeb](#):

[Arthropoda](#) ±

[Insecta](#) ±

[Hymenoptera](#) ±

[Formicidae](#) ±

[Adetomyrma](#) ±

[Adetomyrma venatrix](#)

[Adetomyrma aureocuprea](#)

[Adetomyrma bressleri](#)

[Adetomyrma caputleae](#)

[Adetomyrma cassis](#)

- Taxonomy
- Media



Description

FossilSpecimen UT-NHM - TUG - 668-242 IMG#4

Image of *Natica globosa*

© Museum of Geology, University of Tartu

Source information

(cc) BY-SA

Some rights reserved

© Museum of Geology, University of Tartu

[View source](#)

Supplier: [Open Up!](#)

Creator: Museum of Geology, University of Tartu

- ▶ Taxonomy
- ▶ Media
- ▶ Text

General Description

All species of *Pelegrina* are easily distinguished from relatives by the male embolus, which have two rami placed retrolaterally to the opening. The embolus is further differentiated in *P. montana* by small denticles on its anterior surface. On average, *P. montana* is larger and of a darker colouration than related species. On males, the side band and forehead band common to all *Pelegrina* are a faded white and the cheek band is either weakly present or completely absent. The clypeus is brown, with dark setae overhanging the chelicerae. The white marks on the dorsal portion of the abdomen are absent except on the latter half, where they are reduced, forming thin, pale bars. Females possess the characteristic thickened epygynal flaps present in all *Pelegrina*. A prominent diagnostic characteristic is these flaps' short length and convex shape. The first curve of the duct is narrow, and the second curve follows obliquely. Immediately behind the flaps is a prominent ridge shaped like a rounded rectangle. The carapace of the spider is dark, but covered in sparse, white hairs. These white scales are found in higher density on the clypeus. The abdomen is dark brown with small pairs of white spots (Maddison, 1996). Males: avg. body length 4.6 mm, carapace length 2.1 mm; females: avg. body length 5.9-6.9 mm, carapace length 2.5 mm (Maddison, 1996).

Source information



Copyright © 2015 University of Alberta Museums

© University of Alberta Museums

[View source](#)

Supplier: [University of Alberta Museums](#)

Source: [E.H. Strickland Entomological Museum](#)

Author: Trevoy, S.

EOL content

► Taxonomy

► Media

► Text

► Data

Colobus satanas

Black Colobus

body length 1,378 mm (female)

Data about this record

catalog number	598560
measurement unit	mm
sex	female
source	http://collections.mnh.si.edu/search/mammals/?nb=598560
measurement remarks	Verbatim measurements: Total Length: 1378mm: ; Tail Length: 860mm: ; Hind Tarsus Length: 174mm: ; Ear Notch Length: 35mm: ; Specimen Weight: 8600.0g:
institution code	USNM
field notes	Stomach Content: digested leaves = 483.9g
citation	Smithsonian Institution, National Museum of Natural History, Department of Vertebrate Zoology, Division of Mammals. Data for specimen 598560. http://collections.mnh.si.edu/search/mammals/
preparations	Skin: ; Skull: ; Skeleton:
occurrence remarks	Progeny Details: 1: embryo: 42: mm: ; Mammals Field Number: 4055; Mammals Field Number: TU 4256
locality	Precise Locality: Evuenam;Province/State: Rio Muni;Country: Equatorial Guinea;
recorded by	Jones, Clyde
scientific name	Colobus satanas

EOL content

At black light among
mixed forest, farmed
fields, tidal creek
with salt marsh

BRED
PARASITE?
PHYLLOTOMA
NEMORATA



Limestone cave in Sierra
Cachiri NW of Maracaibo,
Distrito Mara, Edo Zulia,
Venezuela.

Dacus psidii
in guava
Noumea New
Caledonia VI-50
N. KRAUSS-IX
50-11833

Dendroica petechia chrysodeta NKK
ovary: 4.5 x 4 mm, ovum minute, oviduct: 1 mm, no brood patch present
skull: 100% oss.; iris dark brown, maxilla dark gray, mandible medium
13.5 g., trace fat; pearl gray, tarsi + toes yellow/brown, pale yellow pads
stomach saved; insect parts; no remige, rectrix molt, light malar molt, trace head molt
keel: 14.3 mm, interorbital distance: 2.8 mm
left wing chord: 64.2 mm, arcs 66 mm, tail: 52 mm; probable mate of NKK 366
shot in beach area 200 m from mangrove, foraging in isolated *Avicennia*

Data integration: names reconciliation

- ▶ Name strings
- ▶ Synonym relationships
- ▶ Authority strings
- ▶ Ranks
- ▶ Parent taxa
- ▶ Child taxa

Catalogue of Life



WoRMS
World Register of Marine Species



The Paleobiology Database
revealing the history of life

algaeBASE



Interim
Register of
Marine and
Nonmarine
Genera

NCBI

FishBase



Data integration: community standards

Darwin Core

taxa
occurrences
measurements or facts

Dublin Core

rights & attribution
language, format
general resource description

Audubon Core

images
videos
sounds

Species Profile Model

subjects

Data integration: semantic annotations



The OBO Foundry

Open Biological and Biomedical Ontology

envo	environments
oba	biological attributes
pato	phenotypic quality
to	plant traits
uberon	multi-species anatomy
vt	vertebrate traits
ro	relations
uo	units of measurements
stato	statistics
chebi	chemical entities

Marineregions.org



ECOREGIONS



WIKIDATA



GeoNames



Experimental Factor Ontology

Semanticscience

Integrated Ontology (SIO)

GRBio

GLOBAL
REGISTRY OF
BIODIVERSITY
REPOSITORIES



Data integration: semantic annotations

body length

[OBA:VT0001256](#)

The length of a multicellular organism.

6

cm

[UO:0000015](#)

A length unit which is equal to one hundredth of a meter or 10^{-2} m.

male

[PATO:0000384](#)

A biological sex quality inhering in an individual or a population whose sex organs contain only male gametes.

mature

[PATO:0001701](#)

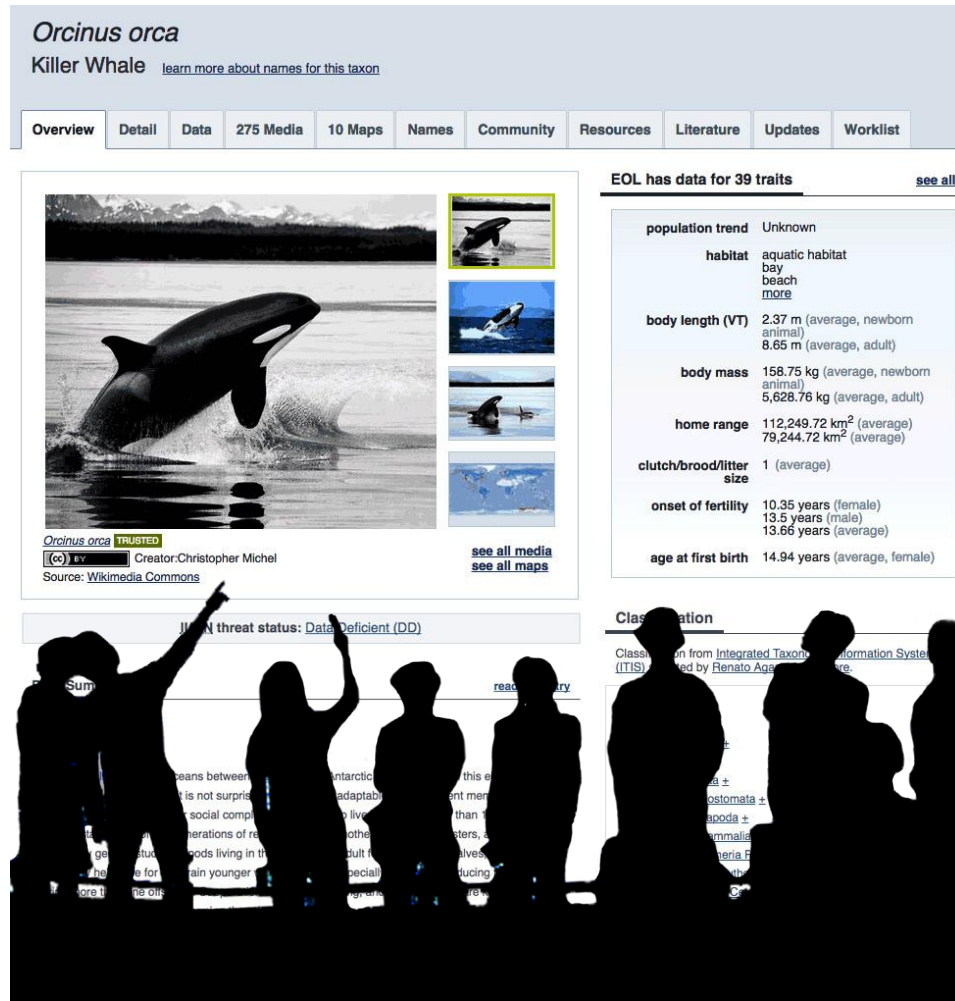
A quality inhering in a bearer by virtue of the bearer's exhibiting complete growth, differentiation, or development.

Dissemination to diverse audiences

humans

Orcinus orca
Killer Whale [learn more about names for this taxon](#)

Overview Detail Data 275 Media 10 Maps Names Community Resources Literature Updates Worklist



EOL has data for 39 traits [see all](#)

population trend	Unknown
habitat	aquatic habitat bay beach more
body length (VT)	2.37 m (average, newborn animal) 8.65 m (average, adult)
body mass	158.75 kg (average, newborn animal) 5,628.76 kg (average, adult)
home range	112,249.72 km ² (average) 79,244.72 km ² (average)
clutch/brood/litter size	1 (average)
onset of fertility	10.35 years (female) 13.5 years (male) 13.66 years (average)
age at first birth	14.94 years (average, female)

[see all media](#)
[see all maps](#)

[IUCN threat status: Data Deficient \(DD\)](#)

Classification

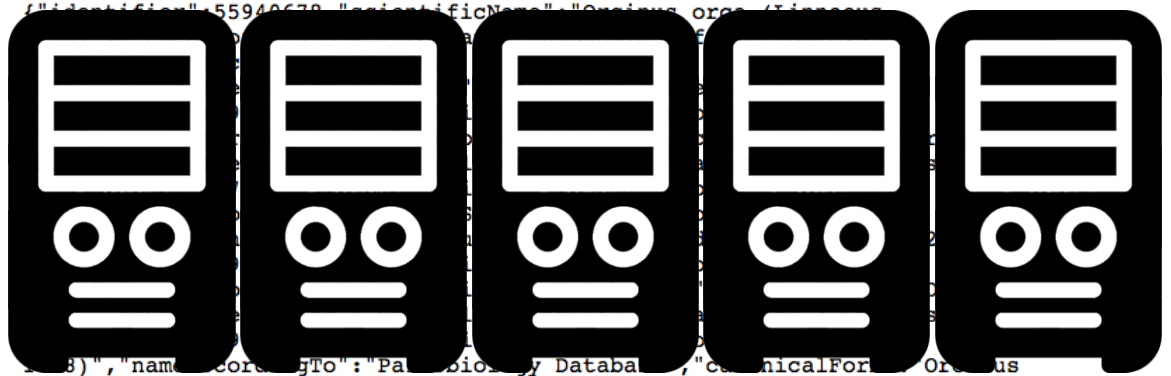
Classified from Integrated Taxonomic Information System (ITIS) created by Renato Aguiar.

Summary

means between... antarctic... this e...
it is not surpris... adaptabl... ent men...
social comp... on live... than m...
generations of re... other... nters, a...
ge stud... ods living in th... adult fi...
he me for... rain younger... ducing...
ore to the dis... ngrants... are

machines

```
{
  "identifier": 328534,
  "scientificName": "Orcinus orca (Linnaeus, 1758)",
  "richness_score": 900.0,
  "taxonConcepts": [
    {
      "identifier": 52576326,
      "scientificName": "Orcinus orca (Linnaeus, 1758)",
      "nameAccordingTo": "Species 2000 & ITIS Catalogue of Life: April 2013",
      "canonicalForm": "Orcinus orca",
      "sourceIdentifier": "6850349",
      "taxonRank": "Species"
    },
    {
      "identifier": 49338079,
      "scientificName": "Orcinus orca (Linnaeus, 1758)",
      "nameAccordingTo": "Paleobiology Database",
      "canonicalForm": "Orcinus orca",
      "sourceIdentifier": "urn:paleodb:tn64541",
      "taxonRank": "Species"
    },
    {
      "identifier": 56582740,
      "scientificName": "Orcinus orca (Linnaeus, 1758)",
      "nameAccordingTo": "WORMS Species Information (Marine Species)",
      "canonicalForm": "Orcinus orca",
      "sourceIdentifier": "137102",
      "taxonRank": "Species"
    },
    {
      "identifier": 61335391,
      "scientificName": "Orcinus orca (Linnaeus, 1758)",
      "nameAccordingTo": "WORMS Species Information (Marine Species)",
      "canonicalForm": "Orcinus orca",
      "sourceIdentifier": "137102",
      "taxonRank": "Species"
    },
    {
      "identifier": 24950241,
      "scientificName": "Orcinus orca",
      "nameAccordingTo": "IUCN Red List (Species Assessed for Global Conservation)",
      "canonicalForm": "Orcinus orca",
      "sourceIdentifier": "IUCN-15421"
    },
    {
      "identifier": 46596031,
      "scientificName": "Orcinus orca (Linnaeus, 1758)",
      "nameAccordingTo": "Integrated Taxonomic Information System (ITIS)",
      "canonicalForm": "Orcinus orca",
      "sourceIdentifier": "180469",
      "taxonRank": "Species"
    },
    {
      "identifier": 55940678,
      "scientificName": "Orcinus orca (Linnaeus, 1758)",
      "nameAccordingTo": "Paleobiology Database",
      "canonicalForm": "Orcinus orca"
    }
  ]
}
```



Attribution standards

Darwin Core

institutionCode
collectionCode
catalogNumber
recordedBy
identifiedBy
measurementDeterminedBy

Audubon Core

furtherInformationURL
derivedFrom
rights:owner

Dublin Core

bibliographicCitation
publisher
contributor
creator
source

agent – role:

author	funder
compiler	illustrator
copyright holder	metadata contact
curator	photographer
data contributor	provider
data manager	recordist
editor	reviewer

Attribution can be complex



creator: Michele Esposito

recorded by: Epifânio, S.; Chaul, J.

identified by: Chaul, J.

rights holder: California Academy of Sciences

dataset: AntWeb

publisher: California Academy of Sciences

supplier: Global Biodiversity Information Facility (GBIF)

collection code: ANTWEB

institution code: CAS

owner institution code: UFV-LABECOL

catalogue number: ufv-labecol-007525

occurrence id: CAS:ANTWEB:ufv-labecol-007525

identifier: CAS:ANTWEB:ufv-labecol-007525

citation: Fisher B, Fong J (2018). AntWeb. California Academy of Sciences.

Occurrence Dataset <https://doi.org/10.15468/wqmijt> accessed via GBIF.org on 2018-02-12. <https://www.gbif.org/occurrence/1563377197>

Attribution can be complex

Colobus satanas
Black Colobus

body length 1,378 mm (female)

Data about this record

catalog number	598560
measurement unit	mm
sex	female
source	http://collections.mnh.si.edu/search/mammals/?nb=598560
measurement remarks	Verbatim measurements: Total Length: 1378mm: ; Tail Length: 860mm: ; Hind Tarsus Length: 174mm: ; Ear Notch Length: 35mm: ; Specimen Weight: 8600.0g:
institution code	USNM
field notes	Stomach Content: digested leaves = 483.9g
citation	Smithsonian Institution, National Museum of Natural History, Department of Vertebrate Zoology, Division of Mammals. Data for specimen 598560. http://collections.mnh.si.edu/search/mammals/
preparations	Skin: ; Skull: ; Skeleton:
occurrence remarks	Progeny Details: 1: embryo: 42: mm: ; Mammals Field Number: 4055; Mammals Field Number: TU 4256
locality	Precise Locality: Evuenam;Province/State: Rio Muni;Country: Equatorial Guinea;
recorded by	Jones, Clyde
scientific name	Colobus satanas

Contribution vs. authorship

TAXON — 1 Sep 2017: 3 pp.

Rouhan & al. • Natural History Collections and co-authorship

POINT OF VIEW

The time has come for Natural History Collections to claim co-authorship of research articles

Germinal Rouhan,¹ Laurence J. Dorr,² Laurent Gautier,³ Philippe Clerc,³ Serge Muller¹ & Myriam Gaudeul¹

1 Muséum national d'Histoire naturelle, Institut de Systématique Évolution Biodiversité (ISYEB, UMR 7205 CNRS, MNHN, UPMC, EPHE), Sorbonne Universités, Herbar national, 57 rue Cuvier, CP39 – 75005, Paris, France

2 National Museum of Natural History, Department of Botany, MRC-166, Smithsonian Institution, P.O. Box 37012, Washington, D.C. 20013-7012, U.S.A.

3 Conservatoire et Jardin botaniques de la Ville de Genève, C.P. 71, 1292 Chambésy, Switzerland

Author for correspondence: *Germinal Rouhan, rouhan@mnhn.fr*

DOI <https://doi.org/10.12705/665.2>

Measuring the impact of bioresources

Developing a guideline to standardize the citation of bioresources in journal articles (CoBRA)

Elena Bravo , Alessia Calzolari, Paola De Castro, Laurence Mabile, Federica Napolitani, Anna Maria Rossi and Anne Cambon-Thomsen

BMC Medicine 2015 **13**:33

<https://doi.org/10.1186/s12916-015-0266-y> | © Bravo et al.; licensee BioMed Central. 2015

Received: 25 June 2014 | Accepted: 2 January 2015 | Published: 17 February 2015

Quantifying the use of bioresources for promoting their sharing in scientific research

Laurence Mabile, Raymond Dalglish, Gudmundur A Thorisson, Mylène Deschênes, Robert Hewitt, Jane Carpenter, Elena Bravo, Mirella Filocamo, Pierre Antoine Gourraud, Jennifer R Harris ... [Show more](#)

GigaScience, Volume 2, Issue 1, 1 December 2013, Pages 1–8, <https://doi.org/10.1186/2047-217X-2-7>

Published: 01 May 2013 **Article history** ▼



Transitive Credit for contributors

Issues in Research Software

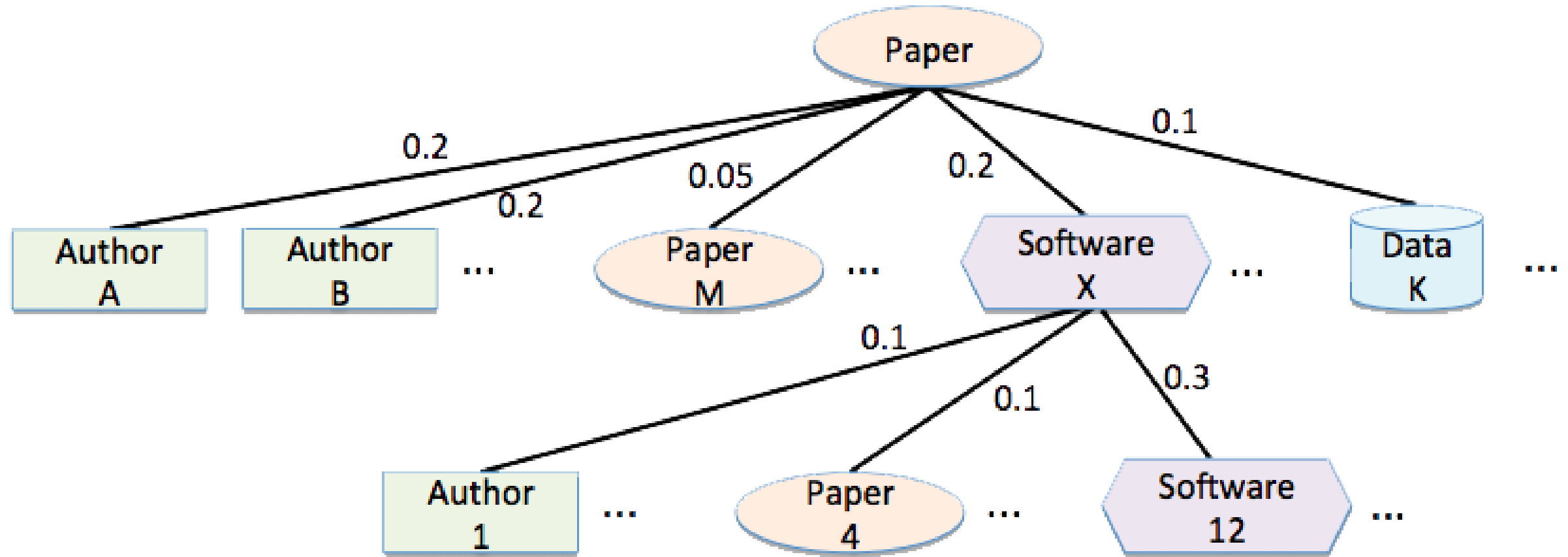
Transitive Credit as a Means to Address Social and Technological Concerns Stemming from Citation and Attribution of Digital Products

Author: Daniel S. Katz 

Keywords: [citation](#), [credit](#), [attribution](#), [software](#)

How to Cite: Katz, D.S., (2014). Transitive Credit as a Means to Address Social and Technological Concerns Stemming from Citation and Attribution of Digital Products. Journal of Open Research Software. 2(1), p.e20. DOI: <http://doi.org/10.5334/jors.be>

Transitive Credit for contributors



Transitive Credit for contributors

- ▶ unique identifiers for products
- ▶ unique identifiers for people
- ▶ unique identifiers for institutions
- ▶ standards to define contributor roles
- ▶ automated provenance systems
 - ▶ harvest contribution data
 - ▶ build contribution metrics



What can aggregators do?

- ▶ support attribution standards
- ▶ expose identifiers
- ▶ provide attribution guidance to data users
- ▶ provide metrics for individuals & institutions