DATA SHARING AND STANDARDS IN MOVEMENT ECOLOGY

WHY AND HOW TO ARCHIVE YOUR DATA

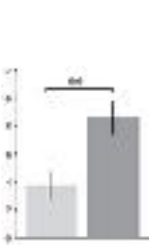
20 June 2018

Sarah C. Davidson, Movebank Data Curator

What happens to your data?

What is data sharing?

archiving for future re-use



compilation with collaborators

sharing “by request” or on university or journal website

informal exchange between colleagues

within an organization or research group

Why share data?

- Give access to results of publicly-funded research
- Allow data to be re-used for other purposes
- Receive additional citations for your research
- Allow analyses to be verified
- Fulfill data sharing and management requirements

Studies for which the related data were made publicly available showed a 69% increase in citations (cancer microarray clinical trial data).

Piwowar et al. 2007

Author reluctance to share data was correlated with weaker statistical evidence against null hypothesis and more apparent errors in reporting statistical results (psychology).

Wicherts et al. 2011

Studies for which the related code were made publicly available were more highly cited (image processing research).

Vandewalle 2012

telecoms.com

Why share data?

Studies for which the related data were made publicly available showed a 69% increase in citations (cancer microarray clinical trial data).

Piwowar et al. 2007

Author reluctance to share data was correlated with weaker statistical evidence against null hypothesis and more apparent errors in reporting statistical results (psychology).

Wicherts et al. 2011

Studies for which the related code were made publicly available were more highly cited (image processing research).

Vandewalle 2012

telecoms.com

Why share data?

- Journals with data-sharing requirements

The American Naturalist
Behavioral Ecology
Biology Letters
Ecological Applications
Ecological Monographs
Ecology and Evolution
Functional Ecology
Journal of Animal Ecology
Journal of Applied Ecology
Journal of Avian Biology
Journal of Avian Biology
Journal of Ecology
Journal of Fish and Wildlife Management
Journal of Zoology
Methods in Ecology and Evolution
Philosophical Transactions of the Royal Society B
PLOS journals
PNAS
Proceedings of the Royal Society B
Science

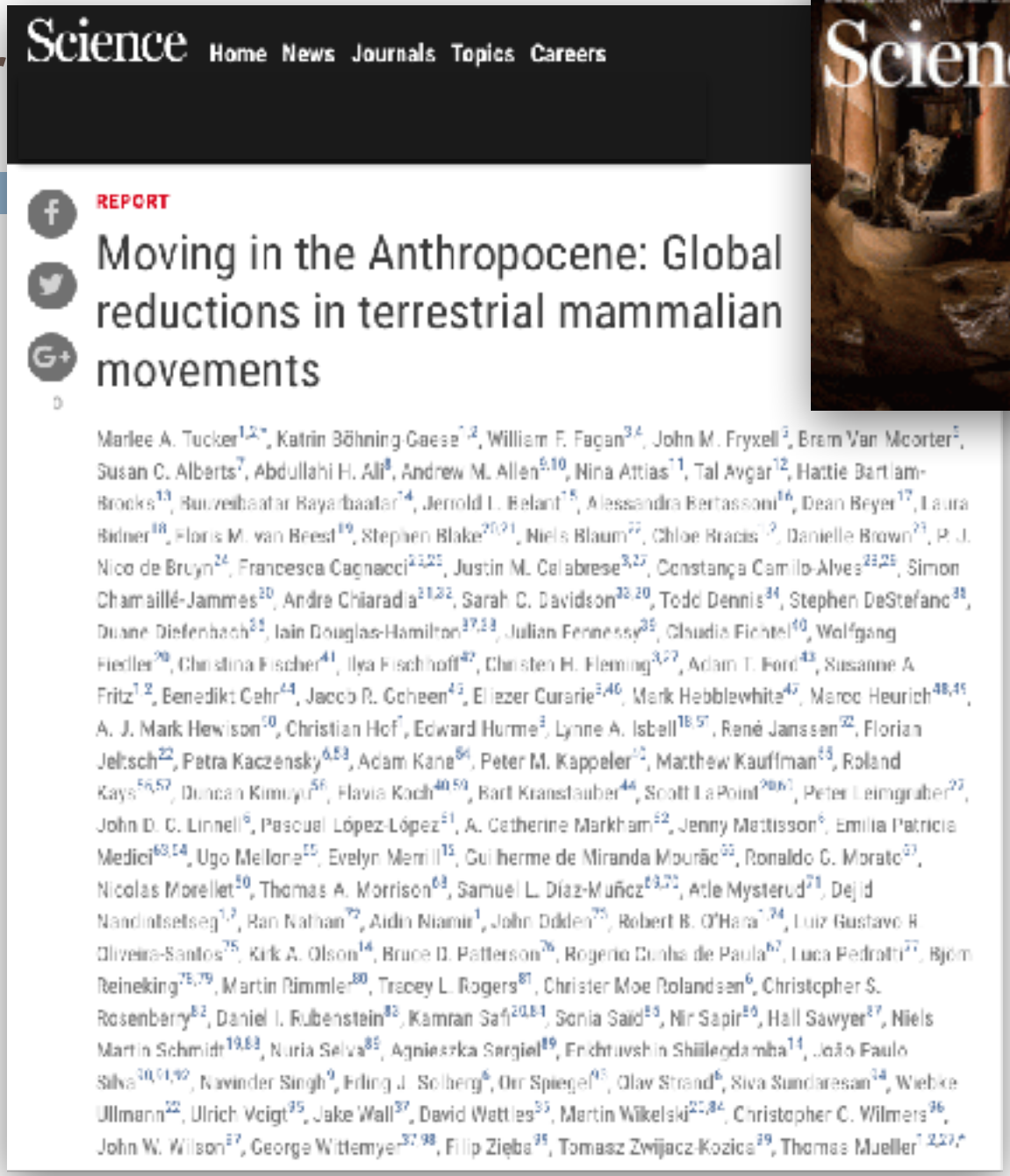
Why shar

species: 57

individuals: 803

projects: dozens

January 2018



Science Home News Journals Topics Careers

REPORT

Moving in the Anthropocene: Global reductions in terrestrial mammalian movements

Marlee A. Tucker[1,2,*], Katrin Böhning-Gaese[1,2], William F. Fagan[3,4], John M. Fryxell[5], Bram Van Moorter[6], Susan C. Alberts[7], Abdullahi H. Ali[8], Andrew M. Allen[9,10], Nina Attias[11], Tal Avgar[12], Hattie Bartlam-Brooks[13], Buuveibaatar Bayarbaatar[14], Jerrold L. Belant[15], Alessandra Bertassoni[16], Dean Beyer[17], Laura Bidner[18], Floris M. van Beest[19], Stephen Blake[20,21], Niels Blaum[22], Chloe Bracis[23], Danielle Brown[21], P. J. Nico de Bruyn[24], Francesca Cagnacci[25,26], Justin M. Calabrese[9,27], Constança Camilo-Alves[28,29], Simon Chamaillé-Jammes[30], Andre Chiaradia[31,32], Sarah C. Davidson[33,20], Todd Dennis[34], Stephen DeStefano[35], Duane Diefenbach[36], Iain Douglas-Hamilton[37,38], Julian Fennessy[39], Claudia Fichtel[40], Wolfgang Fiedler[20], Christina Fischer[41], Ilya Fischhoff[42], Christen H. Fleming[9,27], Adam T. Ford[43], Susanne A. Fritz[1,2], Benedikt Gehr[44], Jacob R. Goheen[45], Eliezer Gurarie[9,46], Mark Hebblewhite[47], Marco Heurich[48,49], A. J. Mark Hewison[50], Christian Hof[1], Edward Hurme[3], Lynne A. Isbell[18,51], René Janssen[52], Florian Jeltsch[22], Petra Kaczensky[6,53], Adam Kane[54], Peter M. Kappeler[55], Matthew Kauffman[56], Roland Kays[56,57], Duncan Kimuyu[58], Flavia Koch[40,59], Bart Kranstauber[44], Scott LaPoint[20,60], Peter Leimgruber[27], John D. C. Linnell[6], Pascual López-López[61], A. Catherine Markham[62], Jenny Mattisson[6], Emilia Patricia Medici[63,64], Ugo Mellone[65], Evelyn Merrill[15], Guilherme de Miranda Mourão[66], Ronaldo G. Morato[67], Nicolas Morellet[50], Thomas A. Morrison[68], Samuel L. Díaz-Muñoz[69,70], Atle Mysterud[71], Dejid Nandintsetseg[1,7], Ran Nathan[72], Aidin Niamir[1], John Odden[73], Robert B. O'Hara[74], Luiz Gustavo R Oliveira-Santos[75], Kirk A. Olson[14], Bruce D. Patterson[76], Rogerio Cunha de Paula[67], Luca Pedrotti[77], Bjorn Reineking[78,79], Martin Rimmler[80], Tracey L. Rogers[81], Christer Moe Rolandsen[6], Christopher S. Rosenberry[82], Daniel I. Rubenstein[83], Kamran Safi[20,84], Sonia Said[85], Nir Sapir[86], Hall Sawyer[87], Niels Martin Schmidt[19,88], Nuria Selva[89], Agnieszka Sergiel[89], Enkhtuvshin Shiilegdamba[14], João Paulo Silva[30,91,92], Navinder Singh[9], Erling J. Solberg[6], Orr Spiegel[93], Olav Strand[6], Siva Sundaresan[94], Wiebke Ullmann[22], Ulrich Voigt[95], Jake Wall[87], David Wattles[35], Martin Wikelski[27,96], Christopher C. Wilmers[36], John W. Wilson[97], George Wittemyer[37,98], Filip Zięba[99], Tomasz Zwijacz-Kozica[29], Thomas Mueller[1,27,*]

Risks and concerns

Someone will steal my data!

- Others must acknowledge data authorship
- Share after publication and embargo

Others won't understand my data!

- Thoroughly describe your data

The data are sensitive!

- Controlled and limited sharing possible

I don't have time!

- Practice good data management
- Use available resources
- Publishing provides academic credit

Options for data sharing

Publish in a data repository

Add to a shared database

Publish as a data paper

Provide upon request

Publish as supplementary files with a paper

When asked to provide data from papers recently published in high-impact psychology journals, 73% of authors failed to comply, despite data-sharing policies and six months of repeated requests (n=141).

Wicherts et al. 2006

26% of supplemental data links for biomedical research papers published 1998-2005 were no longer accessible (n=655).

Anderson et al. 2006

Data “Publishing”

Review process

- Quality control

Persistent identifiers

- Permanence

Licenses

- Explicit re-use conditions

Persistent identifiers

Unique and persistent reference to a digital object

- find the object if it is moved or renamed

Examples

- **DOI (digital object identifier)**
- LSID (life science identifier)
- non-digital: ISBN (international standard book number)

Licensing data

Data licensing options for re-use

- Creative Commons **CC-Zero**

  creativecommons.org
- Open Data Commons **PDDL, OCD-By**

  opendatacommons.org

License restrictions

- Require attribution
- Restrict to non-commercial uses
- Prohibit modification

Public Domain

Credit is still required based on professional norms!

Avoids unintended constraints

Publish in a data repository

Publish data underlying published journal articles

- Movebank Data Repository: animal movement data

  datarepository.movebank.org

- ZoaTrack: animal movement data

  datadryad.org

- Dryad: scientific and medical data

  datadryad.org

Data from: Costs of migratory decisions: a comparison across eight white stork populations

When using this dataset, please cite the original article.

Citations

Flack A, Fiedler W, Blas J, Pokrovski I, Kaatz M, Mitropolsky M, Aghababyan K, Fakriadis Y, Makrigianni E, Jerzak L, Shamina Flack A, Fiedler W, Blas J, Pokrovski I, Kaatz M, Mitropolsky M, Aghababyan K, Fakriadis Y, Makrigianni E, Jerzak L, Azafzaf H, Feltrup-Azafzaf C, Rotics S, Mokotjomela TM, Nathan R, Wikelski M, 2016, Costs of migratory decisions: a comparison across eight white stork populations. Science Advances 2(1): e1500931. doi:10.1126/sciadv.1500931

Additionally, please cite the Movebank data package:

Flack A, Fiedler W, Blas J, Pokrovski I, Mitropolsky B, Kaatz M, Aghababyan K, Khachatryan A, Fakriadis I, Makrigianni E, Jerzak L, Shamin M, Shamina C, Azafzaf H, Feltrup-Azafzaf C, Mokotjomela TM, Wikelski M (2015) Data from: Costs of migratory decisions: a comparison across eight white stork populations. Movebank Data Repository. doi:10.5441/001/1.78152p3q

Public domain license

Cite | Share

Package Identifier	doi:10.5441/001/1.78152p3q Persistent identifier
Abstract	Annual migratory movements can range from a few tens to thousands of kilometers, creating unique energetic requirements for each specific species and journey. Even within the same species, migration costs can vary largely because of flexible, opportunistic life history strategies. We uncover the large extent of variation in the lifetime migratory decisions of young white storks originating from eight populations. Not only did juvenile storks differ in their geographically distinct wintering locations, their diverse migration patterns also affected the amount of energy individuals invested for locomotion during the first months of their life. Overwintering in areas with higher human population reduced the stork's overall energy expenditure because of shorter daily foraging trips, closer wintering grounds, or a complete suppression of migration. Because migrants can change ecological processes in several distinct communities simultaneously, understanding their life history decisions helps not only to protect migratory species but also to conserve stable ecosystems.
Keywords	animal tracking, avian migration, Ciconia ciconia, Env-DATA, Movebank, movement ecology, white storks

Metadata

MPIO white stork lifetime tracking data (2013-2014)-gps.csv View File Details

Download: README.txt (14.21Kb)
Download: MPIO white stork lifetime tracking data (2013-2014)-gps.csv.zip (24.20Mb)

MOVEBANK

Publish as a data paper

- Biodiversity Data Journal
- BioRisk
- Dataset Papers in Science
- Ecological Archives
- Ecological Research
- Scientific Data (Nature)
- GBIF Integrated Publishing Toolkit

Share in an online tracking database

OBIS-SEAMAP

and many others!

Describing data

Animal or researcher?

24556		41.1247	-73.7935	08.12.2008	2:00
24556		41.1247	-73.8125	08.12.2008	8:00
24556	Anne	41.1248	-73.8287	08.12.2008	14:00
24556	Anne	41.1248	-73.7991	08.12.2008	20:00
24556	Anne	41.1247	-73.7975	09.12.2008	8:00
24556	Anne	41.1158	-73.7883	09.12.2008	14:00
24556	Anne	41.1158	-73.7883	09.12.2008	20:00
24556	Anne	41.1243	-73.7883	10.12.2008	2:00

Describing data

Why are these blank?

Measured or modelled?

Date format?

tag	animal	latitude	longitude	date	time
24556		41.1247	-73.7935	08.12.2008	2:00
24556		41.1247	-73.8125	08.12.2008	8:00
24556	Anne	41.1248	-73.8287	08.12.2008	14:00
24556	Anne	41.1248	-73.7991	08.12.2008	20:00
24556	Anne	41.1247	-73.7975	09.12.2008	8:00
24556	Anne	41.1158	-73.7883	09.12.2008	14:00
24556	Anne	41.1158	-73.7883	09.12.2008	20:00
24556	Anne	41.1243	-73.7883	10.12.2008	2:00

Time zone?

Describing data

What kind of tag? What duty cycle?

tag	id of tag deployed on the animal
animal	id of animal tracked
latitude	measured latitude, decimal degrees, WGS84 datum
longitude	measured longitude, decimal degrees, WGS84 datum
date	date of observation, dd.mm.yyyy
time	time of observation, UTM, hh:mm

What species? manipulated?

tag	animal	latitude	longitude	date	time
24556		41.1247	-73.7935	08.12.2008	2:00
24556		41.1247	-73.8125	08.12.2008	8:00
24556	Anne	41.1248	-73.8287	08.12.2008	14:00
24556	Anne	41.1248	-73.7991	08.12.2008	20:00
24556	Anne	41.1247	-73.7975	09.12.2008	8:00
24556	Anne	41.1158	-73.7883	09.12.2008	14:00
24556	Anne	41.1158	-73.7883	09.12.2008	20:00
24556	Anne	41.1243	-73.7883	10.12.2008	2:00

Describing data

tag	animal	deployOnDate	taxon	tagType	manipulation
24556	Anne	08.12.2008 12:00	Canis latrans	GPS	none
24558	Hans	08.12.2008 12:15	Canis latrans	GPS	translocated from Allegheny National Forest
24559	Fred	08.12.2008 13:42	Canis latrans	GPS	translocated from Allegheny National Forest

tag	animal	latitude	longitude	date	time
24556		41.1247	-73.7935	08.12.2008	2:00
24556		41.1247	-73.8125	08.12.2008	8:00
24556	Anne	41.1248	-73.8287	08.12.2008	14:00
24556	Anne	41.1248	-73.7991	08.12.2008	20:00
24556	Anne	41.1247	-73.7975	09.12.2008	8:00
24556	Anne	41.1158	-73.7883	09.12.2008	14:00
24556	Anne	41.1158	-73.7883	09.12.2008	20:00
24556	Anne	41.1243	-73.7883	10.12.2008	2:00

Describing data

Who did this study and why?

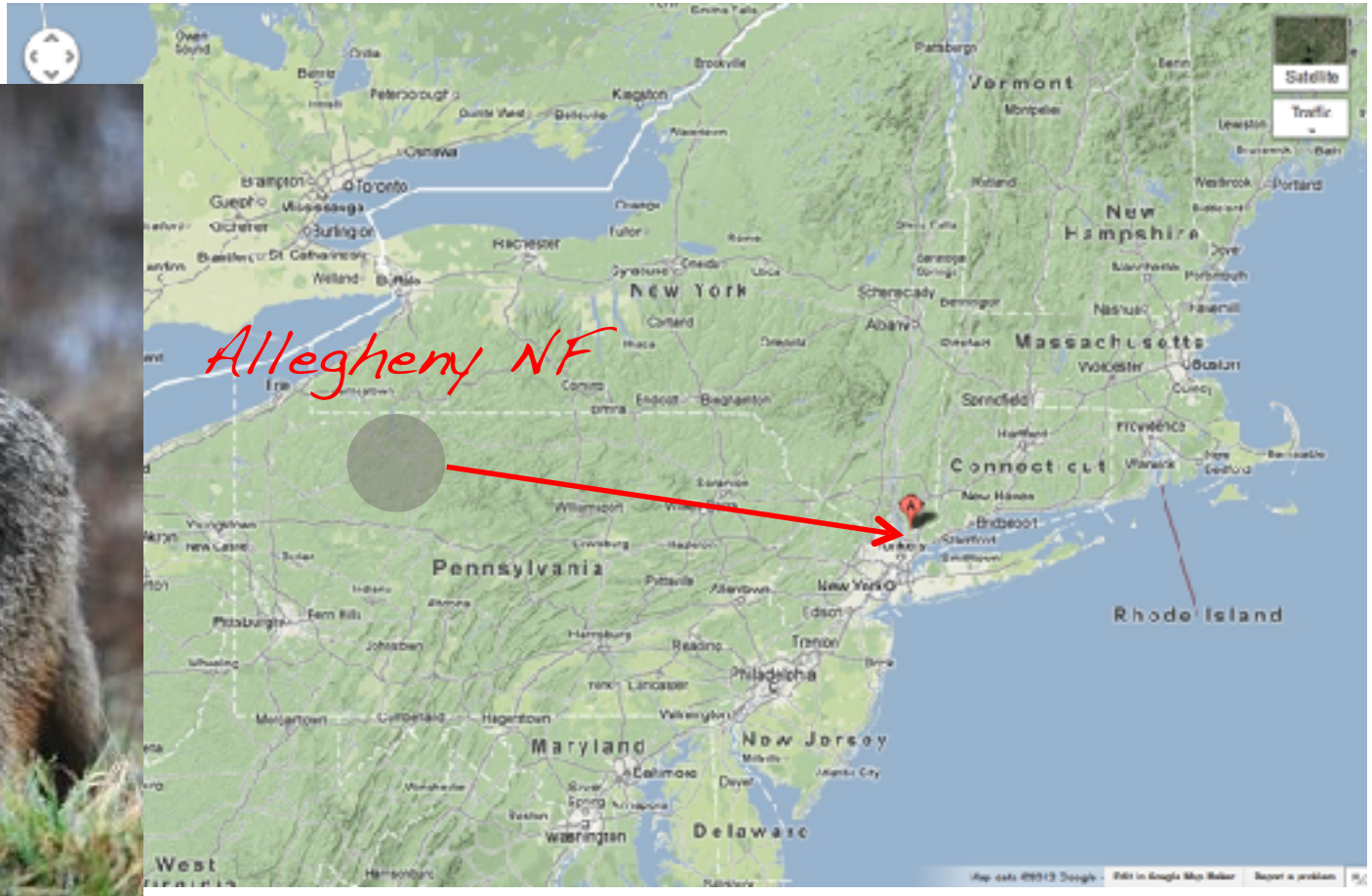


Photo by Rebecca Richardson

Describing data

How can others search for and find this dataset?

author	Doe, J.I.
title	Effects of coyote reintroduction on native population and territory distribution near White Plains, New York
year published	2012
key words	animal tracking, habitat use
temporal coverage	12/8/2008-2/20/2012
taxon	Canis latrans
type of resource	dataset
file format	csv
language	English

Data standards

Interoperability: ability of systems (databases, search engines, analysis software) to work together

- Basic: Describe my data set so others can find it online.
- Advanced: Describe and format my data so it can be seamlessly analyzed with others' data in R.
- Requires standard formats and descriptions

Why standardize?

- Standardized and well-documented **archiving**
- Increased **discoverability**
- **Less time needed** to collect and integrate datasets
- **Reduced error** and uncertainty
- More **collaborations** and re-use
- Increased accessibility and reduced support needs for **shared tools**
- Enabled **conservation efforts, media coverage** and **outreach**

Data standards

Personal database

author	Doe, John I.
title	Effects of coyote reintroduction on native population and territory distribution near White Plains, New York
year published	2012
key words	animal tracking, habitat use
temporal coverage	12/8/2008-2/20/2012
taxon	Canis latrans
type of resource	dataset
file format	csv
language	English

Data standards: Terms

DataCite term	Personal database	
creatorName	author	Doe, John I.
title	title	Effects of coyote reintroduction on native population and territory distribution near White Plains, New York
publicationYear	year published	2012
subject	key words	animal tracking, habitat use
StartDate, EndDate	temporal coverage	12/8/2008-2/20/2012
taxonID	taxon	Canis latrans
type	type of resource	dataset
Format	file format	csv
Language	language	English

Data standards: Terms and format

```
<resource xmlns="http://datacite.org/schema/kernel-2.2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://datacite.org/schema/kernel-2.2 http://schema.datacite.org/meta/kernel-2.2/metadata.xsd">
  <identifier identifierType="DOI">12.3456/789/xx00xx1</identifier>
  <creators>
    <creator>
      <creatorName>Doe, John. I.</creatorName>
    </creator>
  </creators>
  <titles>
    <title>Effects of coyote reintroduction on native population and territory distribution near White Plains, New
      York</title>
  </titles>
  <publicationYear>2012</publicationYear>
  <subjects>
    <subject>animal tracking</subject>
    <subject>habitat use</subject>
  </subjects>
  <language>eng</language>
  <resourceType resourceTypeGeneral="Dataset"/>
  <formats>
    <format>csv</format>
  </formats>
  <rights>Creative Commons CCZero Waiver and Public Domain Dedication, http://creativecommons.org/
    publicdomain/zero/1.0/</rights>
</resource>
```

Machine readable

Data standards: Discoverability

DataCite

Data standards: Discoverability

GBIF

Biodiversity Information Standards TDWG

Darwin Core

Data standards: Biodiversity & GIS

Biodiversity

Biodiversity Information Standards TDWG

Darwin Core

ABCD

knb

Ecological Metadata Language (EML)

GIS & remote sensing

OGC®
Making location count.

Open Geospatial Consortium

netCDF

ISO

fgdc
Federal Geographic Data Committee

Data standards: Animal tracking

INTERNATIONAL BIO-LOGGING SOCIETY

6th International Bio-logging Symposium
27 September 2017
Konstanz, Germany

A FUTURE FOR A COMMON BIO-LOGGING LANGUAGE?

Discussions about data standards and interoperability in the bio-logging world

Francesca Cagnacci
Sarah Davidson
Lee Belbin
Hamish Campbell
Peter Desmet
Holger Dettki
Bernie McConnell
Peggy Newman
Alessandro Oggioni
Ferdinando Urbano

Data standards: Animal tracking

Many and **diverse sensors** and manufacturers

Many and **diverse databases** for bio-logging data

Many and diverse **analysis tools**

Resources often customized based on taxonomy, administrative unit, geography, etc.

There are lots of good reasons for this! ***But…***

Data standards: Animal tracking

There are lots of good reasons for this! ***But…***

Inconsistent data formats, terms, documentation

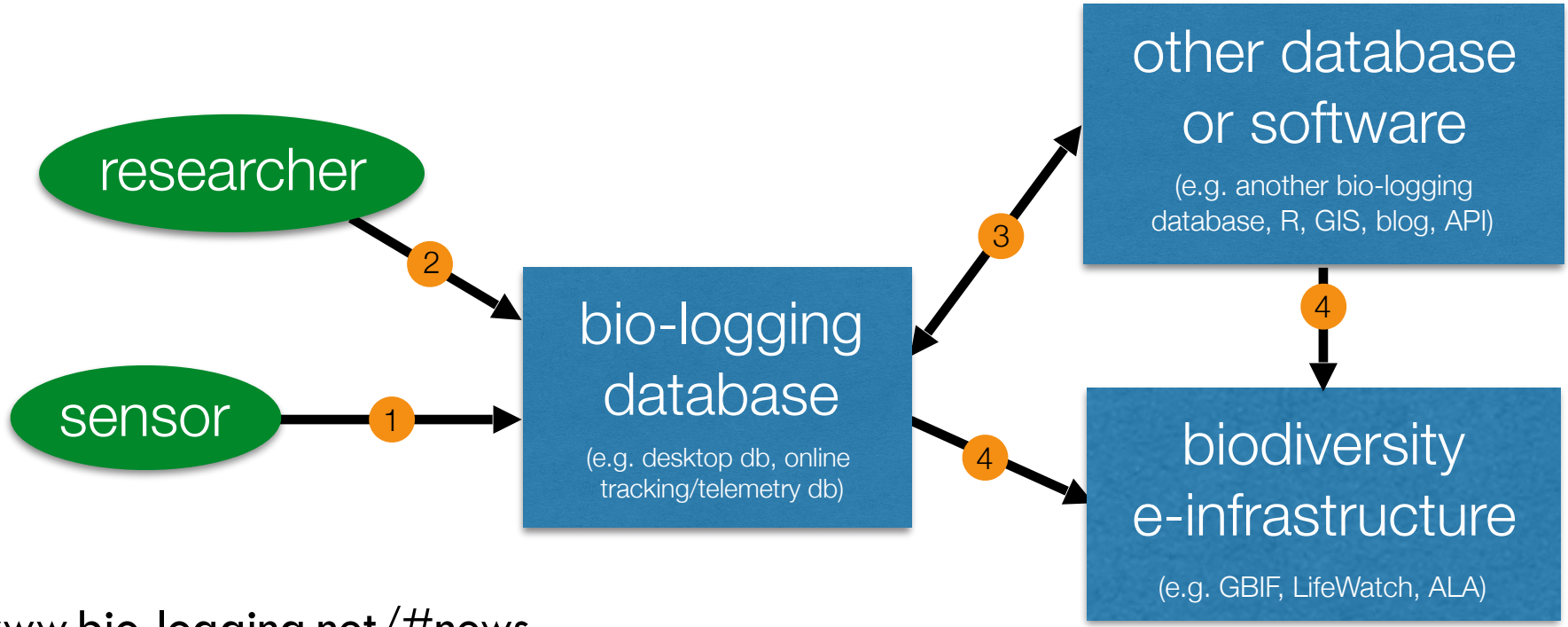
Many datasets remain **poorly documented or undiscoverable**

Little guidance for many data users and providers

Data standards: Animal tracking

IBLS Data Standards Working Group

Goal: Standards to enable the integration and use of data collected by animal-borne sensors originating from different projects, sensor types and manufacturers



www.bio-logging.net/#news

Conclusions

- Data sharing has many potential benefits
- Data sharing requires thorough data description and good management
- Many resources exist to help make your data accessible
- Efforts are underway to increase data standardization for movement ecology

Thank you!

Sarah C. Davidson

Data Curator

Max Planck Institute for Ornithology

The Ohio State University

sdavidson@orn.mpg.de

This presentation was prepared for the course “Dealing with Spatio-temporal Data in Movement and Population Ecology” which took place 17–22 June 2018, at the Fondazione Edmund Mach in San Michele all’Adige, Trento, Italy.