

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



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



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 share

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. Fryxell⁵, Bram Van Moorter⁶, Susan C. Alberts⁷, Abdullahi H. Ali⁸, Andrew M. Allen^{9,10}, Nina Attias¹¹, Tal Avgar¹², Hattie Bartlam-Brooks¹³, Buuveibaatar Bayarbaatar¹⁴, Jerrold L. Belant¹⁵, Alessandra Bertassoni¹⁶, Dean Beyer¹⁷, Laura Bidner¹⁸, Floris M. van Beest¹⁹, Stephen Blake^{20,21}, Niels Blaum²², Chloe Bracis²³, Danielle Brown²⁴, P. J. Nico de Bruyn²⁵, Francesca Cagnacci^{26,27}, Justin M. Calabrese^{3,28}, Constança Camilo-Alves^{23,29}, Simon Chamaillé-Jammes³⁰, Andre Chiaradia^{31,32}, Sarah C. Davidson^{32,30}, Todd Dennis³⁴, Stephen DeStefano³⁸, Duane Diefenbach³⁵, Iain Douglas-Hamilton^{37,33}, Julian Fennessy³⁶, Claudia Fichtel⁴⁰, Wolfgang Fiedler³⁹, Christina Fischer⁴¹, Ilya Fischhoff⁴², Christen H. Fleming^{3,27}, Adam T. Ford⁴³, Susanne A. Fritz^{1,2}, Benedikt Cehr⁴⁴, Jacob R. Coheen⁴⁵, Eliezer Curarie^{3,46}, Mark Hebblewhite⁴⁷, Marco Heinrich^{48,49}, A. J. Mark Hewison⁵⁰, Christian Hof⁵, Edward Hurme³, Lynne A. Isbell^{18,51}, René Janssen⁵², Florian Jeltsch⁵³, Petra Kaczensky^{6,53}, Adam Kane⁵⁴, Peter M. Kappeler⁵⁵, Matthew Kauffman⁵⁶, Roland Kays^{56,57}, Duncan Kimuyu⁵⁸, Flavia Koch^{40,59}, Bart Kranstauber⁴⁶, Scott LaPointe^{60,61}, Peter Leimgruber²⁷, John D. C. Linnell⁶, Pascual López-López⁶¹, A. Catherine Markham⁶², Jenny Mattisson⁶, Emilia Patricia Medici^{63,64}, Ugo Mellone⁶⁵, Evelyn Merrill⁶⁶, Guilherme de Miranda Mourão⁶⁷, Ronaldo G. Morato⁶⁷, Nicolas Morellet⁶⁰, Thomas A. Morrison⁶⁸, Samuel L. Díaz-Muñoz^{69,72}, Atle Mysterud⁷¹, Dejid Nandintsetseg^{1,2}, Ran Nathan⁷², Aidin Niamir¹, John Odden⁷³, Robert B. O'Hara⁷⁴, Luiz Gustavo R. Oliveira-Santos⁷⁵, Kirk A. Olson¹⁴, Bruce D. Patterson⁷⁶, Rogerio Cunha de Paula⁶⁷, Luca Pedrotti⁷⁷, Björn Reineking^{78,79}, Martin Rimmer⁸⁰, Tracey L. Rogers⁸¹, Christer Moe Rolandsen⁶, Christopher S. Rosenberg⁸², Daniel I. Rubenstein⁸³, Kamran Safi^{20,84}, Sonia Said⁸⁵, Nir Sapir⁸⁶, Hall Sawyer⁸⁷, Niels Martin Schmidt^{19,88}, Nuria Selva⁸⁹, Agnieszka Sergiel⁸⁹, Enkhtuvshin Shilegdamba¹⁴, João Paulo Silva^{30,91,92}, Navinder Singh⁹, Erling J. Solberg⁶, Orr Spiegel⁹³, Olav Strand⁶, Siva Sundaresan³⁴, Wiebke Ullmann²², Ulrich Voigt⁹⁵, Jake Wall³⁷, David Wattles⁹⁵, Martin Wikelski^{22,84}, Christopher C. Wilmers⁹⁶, John W. Wilson⁹⁷, George Wittemyer^{37,98}, Filip Zieliński⁹⁹, Tomasz Zwiłłacz-Kozica³⁹, Thomas Mueller^{7,2,27,*}



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.

Flack A, Fiedler W, Blas J, Pokrovski I, Kaatz M, Mitropolsky M, Aghababayan K, Fakriadis Y, Makrigianni E, Jerzak L, Shamina Flack A, Fiedler W, Blas J, Pokrovski I, Kaatz M, Mitropolsky M, Aghababayan K, Fakriadis Y, Makrigianni E, Jerzak L, Azafzal H, Feltrup-Azafzal G, Rotlos 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, Aghababayan K, Khachatryan A, Fakriadis I, Makrigianni E, Jerzak L, Shamir M, Shamir G, Azafzal H, Feltrup-Azafzal G, 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

Cite | Share

Package Identifier **doi:10.5441/001/1.78152p3q**

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.

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)



Citations

Public domain license

Persistent identifier

Metadata

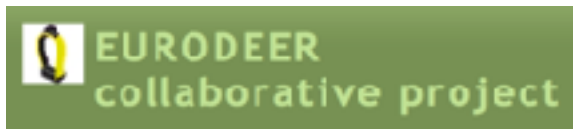
Publish as a data paper

- 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



Seabird Tracking Database
Tracking Ocean Wanderers



OBIS-SEAMAP




seaturtle.org



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?

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

Date format?

Time zone?

Describing data

		<i>What kind of tag? What duty cycle?</i>
tag	id of tag deployed on the animal	
animal	id of animal tracked	<i>What species? manipulated?</i>
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	

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

Who did this study and why?

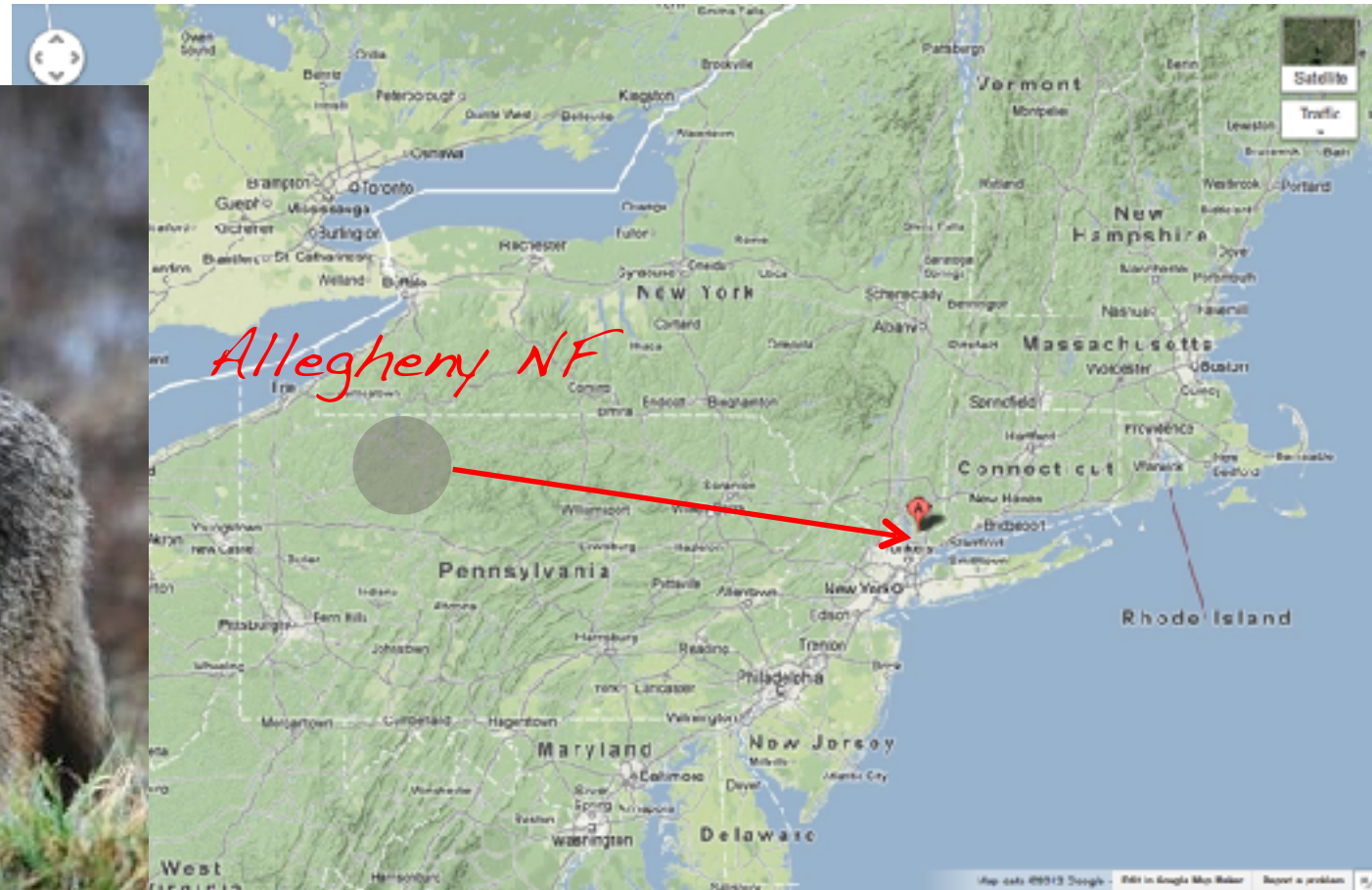


Photo by Rebecca Richardson

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.
		Effects of coyote reintroduction on native population and territory distribution near White Plains, New York
title	title	
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



WEB OF SCIENCE™

Search | Return to Search Results | My Tools | Search History

MPG | Save to EndNote online | Add to Marked List

Data from: Migratory connectivity and population specific migration routes in a long-distance migratory bird

From Repository: Movebank Data Repository

By: Trierweiler, Christiane; Klaassen, Raymond HG; Drent, Rudi H; Exo, Klaus-Michael; Komdeur, Jan; Bairlein, Franz; Koks, Ben J

Movebank Data Repository
DOI: <http://dx.doi.org/10.5441/001/1.tb272j.c1>
Viewed Date: 12 Mar 2014
Published: 2014

Abstract

Knowledge about migratory connectivity, the degree to which individuals from the same breeding site migrate to the same wintering site, is essential to understand processes affecting populations of migrants throughout the annual cycle. Here, we study the migration system of a long-distance migratory bird, the Montagu's harrier *Circus pygargus*, by tracking individuals from different breeding populations throughout northern Europe. We identified three main migration routes towards wintering areas in

Citation Network

1 Times Cited
0 Cited References
Create Citation Alert
(data from Web of Science)

All Times Cited In

- 1 In All Databases
- 1 In Web of Science
- 1 In BIOSIS Previews
- 0 In Chinese Science Database
- 0 In Data Cite
- 0 In Sciendo



Data standards: Discoverability



Biodiversity
Information
Standards
TDWG

ARGOS Satellite Tracking of animals

Occurrence dataset published by Australian Antarctic Data Centre

346,511
Occurrences

View occurrences

InformationStatsActivity

Summary

FULL TITLE
ARGOS Satellite Tracking of animals

DESCRIPTION
Various species have been tracked using ARGOS PTT trackers since the early 1990's. These include Emperor, King and Adelia penguins, Light-mantled Sooty, Grey-headed and Black-browed albatrosses, Antarctic and Australian fur seals, Southern Elephant Seal and Blue and Humpback whales. Note that not all data for any species or locations is or will be exposed to OBIS. Geographic coverage is from Heard Island to the west and Macquarie Island to the east and several islands near the southern end of Chile. The data has been filtered to remove most but not all erroneous positions.

LANGUAGE OF METADATA
ENGLISH

DOI [doi:10.15488/1ancjy](https://doi.org/10.15488/1ancjy)

PUBLISHED BY
[Australian Antarctic Data Centre](#)

REGISTRATION DATE
Apr 3, 2007

SERVED BY
[DiGIR Installation](#)

LINKS
• [Dataset homepage](#)

ALTERNATIVE IDENTIFIERS
• [GBIF Portal ID](#)
[http://data.gbif.org/datasets/...](http://data.gbif.org/datasets/)

EXTERNAL DATA
• [DiGIR](#)

METADATA DOCUMENTS
• [GBIF annotated version \(EML\)](#)

Darwin Core →

Data standards: Biodiversity & GIS

Biodiversity



Darwin Core
ABCD



Ecological
Metadata
Language
(EML)

GIS & remote sensing



Open Geospatial Consortium



Data standards: Animal tracking



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

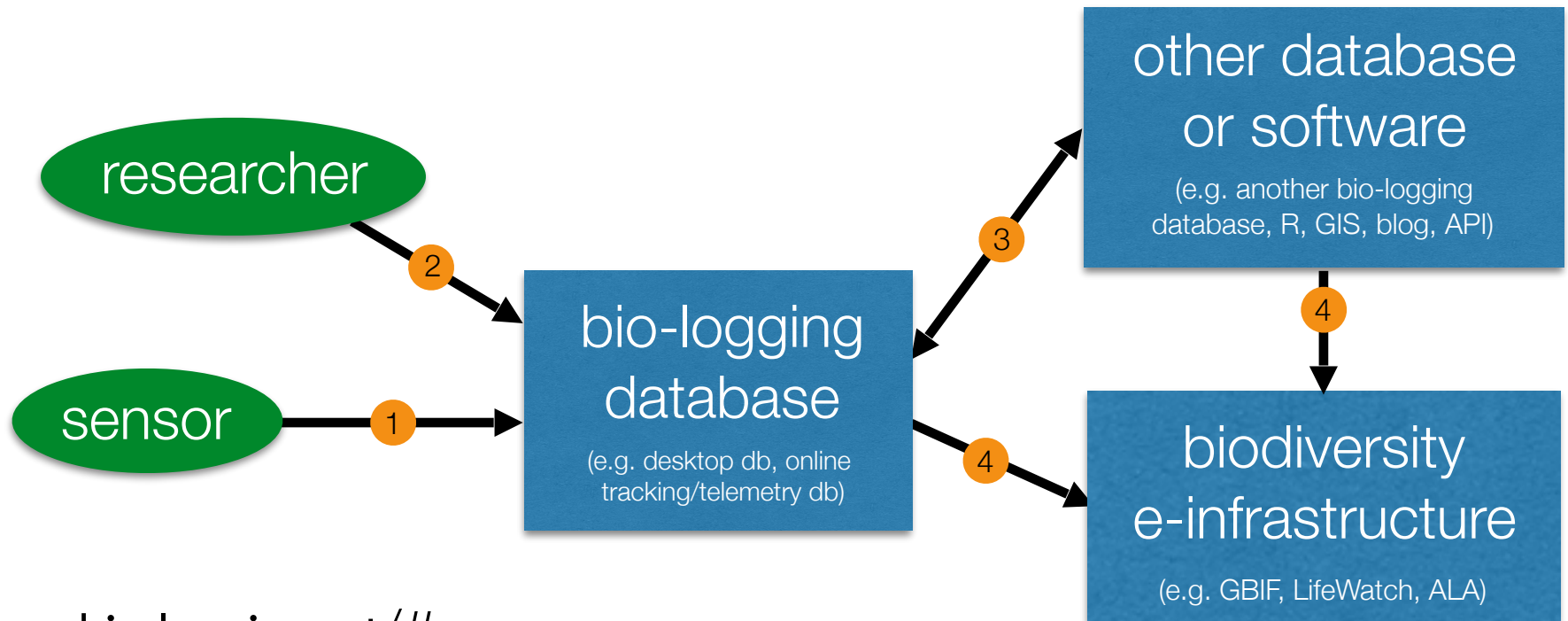
Many datasets remain **poorly documented or undiscoverable**

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



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!

The background of the slide is an abstract digital illustration. It features a prominent DNA double helix structure in shades of blue and grey. Overlaid on this are various elements of binary code, including circles and lines containing the numbers '0' and '1' in different colors like yellow, orange, and purple. In the lower-left quadrant, there is a glowing sphere with a complex, crystalline or molecular structure inside it. The overall aesthetic is high-tech and scientific.

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.

This work is licensed under a Creative Commons Attribution 3.0 Unported License (http://creativecommons.org/licenses/by/3.0/deed.en_US). If you reuse information from this presentation, please cite it as “Davidson SC, Urbano F, Cagnacci F (2018) Data sharing and standards in movement ecology. Trento, Italy, 1–21 June 2018.”

