ASSOCIATION OF CANADIAN MAP LIBRARIES AND ARCHIVES
BULLETIN

Geospatial Data and Software Reviews

Tomasz Mrozewski

Movebank

As GIS librarian at a university with a strong Biology program, species range and habitat data is probably the single largest topic for data requests. As a result, I'm constantly on the lookout for a reliable, high-quality source of detailed digital data on a variety of species. Recently this search led me to Movebank (www.movebank.org/), "a free online infrastructure created to help researchers manage, share, analyze and archive animal movement data"[1].

Movebank is an international project led by the Max Planck Institute for Ornithology in Switzerland and coordinates with the North Carolina Museum of Natural Sciences, Ohio State University, and the University of Konstanz with funding from the National Science Foundation, the German Aerospace Center, the German Science Foundation, and NASA. Notably and promisingly, Movebank boasts "long-term (>20 years) funding through the Max Planck Society and the University of Konstanz" with the goal of serving as "a global data archive for animal movement data."[2]

Initiated in 2007, Movebank has been collecting data since 2012. As of January 2018, Movebank data comprises "over 600 million animal locations and over 1.1 billion measurements from non-location sensors such as acceleration bursts and light levels from geolocators [from] nearly 4,200 studies and [representing] movements of 756 taxa all over the world"[3]. Several published papers describe the Movebank infrastructure and many more describe datasets hosted on the platform, dating back to 1975 - a comprehensive bibliography may be found on Movebank's "Literature" page.[4]

The base unit of Movebank's data model is the study - a research project or observation program[5]. A study dataset includes tag (a tracking device or sensor), animal (the individual to which a tag is attached), deployment (the time a tag is deployed), and event (each individual observation of a tag, featuring a time and location) and follows a custom metadata schema described in the Movebank Attribute Dictionary (www.movebank.org/node/2381). As a result, all datasets on the site have a consistent structure that allows for spatial representation in the Tracking Data Map and linkage through the Env-DATA System, both described below. Figure 1 illustrates the Movebank Study data model.

---

[1] "About Movebank," https://www.movebank.org/node/2 - all URLs last retrieved March 20, 2018.

[2] Ibid.

[3] Ibid.

[4] "Literature," https://www.movebank.org/node/63062.

[5] "The Movebank data model," https://www.movebank.org/node/9.
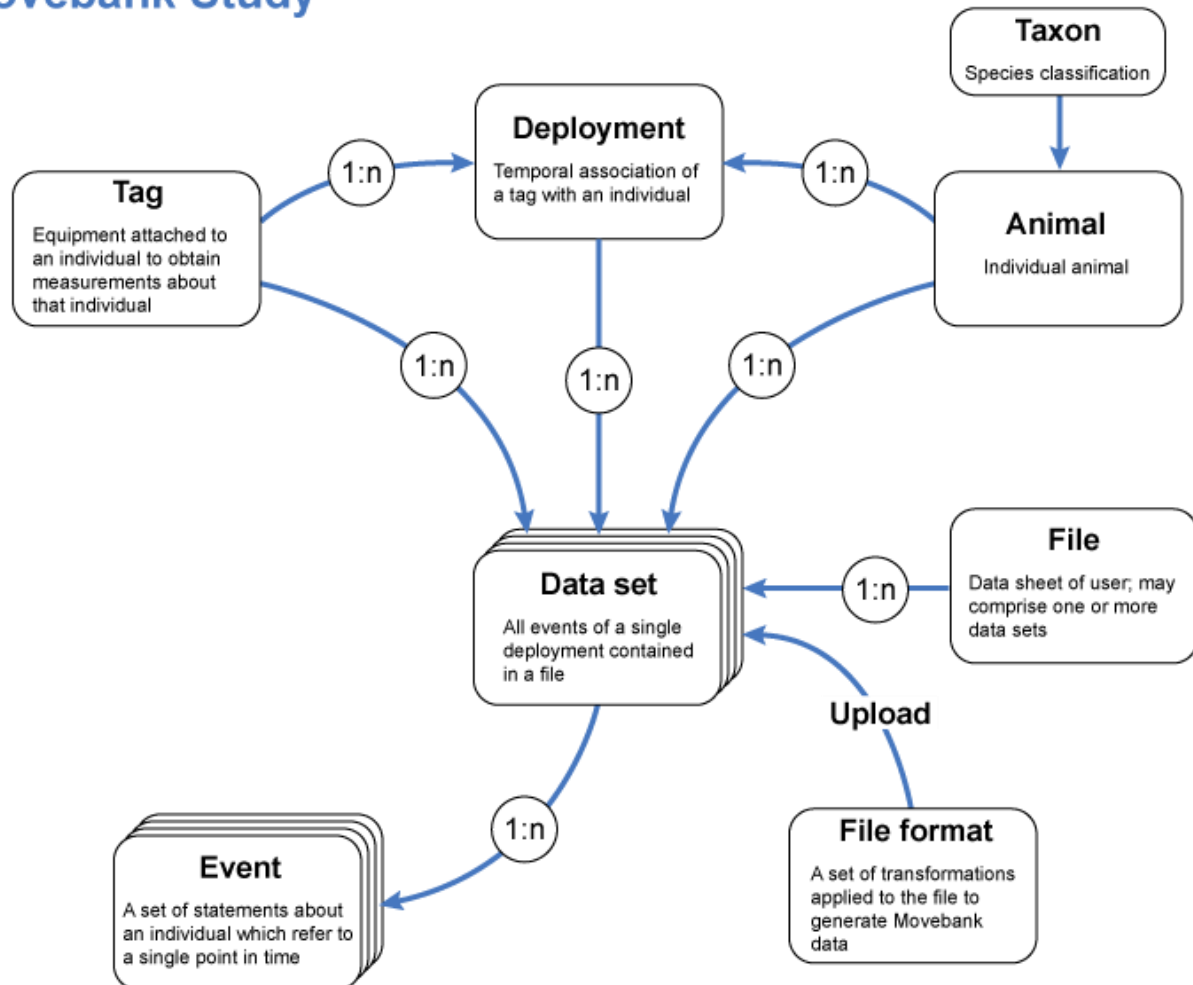
Figure 1. The Movebank Study data model

Movebank's Tracking Data Map (www.movebank.org/panel_embedded_movebank_webapp) will be the site's most interesting feature for most data users. The Tracking Data Map is the primary interface for accessing tracking data. There are two ways to discover datasets: a search/browse panel and a map window. The map displays Study Markers - points coloured green or grey, depending on whether they are openly accessible - representing each study with a single point. Studies can be selected by clicking Study Markers in the map or by clicking them in the search/browse results pane. When a study is selected, the map will display all events associated with that study. It is also possible to toggle the display of lines that connect events from trackers in order to represent the tracks taken by individual animals. Once selected, a study's entry will also expand in the search/browse panel to allow users to highlight individual animals tagged in a study on the map. Selecting a study through the search/browse panel or through the map allows the user to then click through to the link animal, tracking device, and study metadata associated with it, as well as to download the datasets in CSV, SHP, or other formats. The map may also generate custom URLs for specific map views.

The Tracking Data Map is a fascinating tool but does have some drawbacks. It appears to be a custom web app embedded in the Drupal CMS on which the site is built. When viewing study data through the Tracking Data Map, it appears in an HTML window within the map pane of the app with poor navigation features. It is difficult to return to same map view from a data window and I had to use the right mouse button menu to access the Back and Forward buttons to navigate between animal, tracker, and study data pages. Data discovery may also be problematic for some users since, as a single point, the Study Marker is inadequate for representing the range of a study. The search function also lacks the ability to search locations. To illustrate the problem: a study of adult Golden Eagles included events ranging from the far Northwest extremity of Alaska to the state of Durango in Mexico to the South to Arizona in the West; the Study Marker for this study was placed over central Montana, which may be the centroid of the events according to some map projection but not for the one used as a base map. Although this study contains events in British Columbia, for example, neither the search function nor Study Markers allow for the identification of events in that province - the user may only see that there are events in BC for the study by selecting the study to visualize its contents. Finally, in testing, I found that the web app was prone to stalling when switching between or zooming to selected datasets. Although it never actually crashed, it sometimes took a minute or longer to view a study on the map.

Movebank also hosts the Movebank Data Repository (www.datarepository.movebank.org/) hosted by the Library of the University of Konstanz. The Repository allows for data publication in compliance with journal and/or funder data sharing and publication requirements. It publishes datasets under a CC0 license with DOIs that are registered with Datacite. As of March 2018 it contains 102 datasets and 272 data files. Publication is free but all datasets must first be entered as studies in Movebank using the standard data model and metadata schema and must be curated by staff from the Max Planck Institute for Ornithology and University of Konstanz Library - although this may be seen as a barrier by some researchers, it does guarantee data cleanliness.
Unfortunately, the Data Repository suffers from some glaring omission of discovery functions. The only way to discover data in the Repository is through an Atom feed of the 10 most recent datasets on the front page or through the Author Browse feature - there is no search function or title or subject browse. It seems that the Data Repository is meant to be accessed via direct DOI link to the datasets rather than as a tool for data discovery itself, which is perhaps unsurprising if its main goal is to host replication data for published papers rather than to support discovery for novel, secondary reuse.

Movebank also provides a suite of tools that will be of more use to the researcher than to the data user since they support the collection, management, and analysis of data. Movebank supports the creation and management of live feeds from common tracking hardware which enable direct download and near-live updating of studies (Movebank boasts 800 live studies as of January 2018[6]). It also features a deployment tracker and a collection of community-developed software libraries for common data analysis platforms such as R and Java. However, the perhaps the most interesting tool is the Env-DATA System, which stands for Environmental Data Automated Track Annotation System[7]. The Env-DATA System harvests environmental data from 19 different data sources from providers including NASA, ESA, USGS, and others and allows the data to be linked

---

[6] "January 2018 Newsletter," https://www.movebank.org/node/90724.
[7] "The Env-DATA System," https://www.movebank.org/node/6607.

to events in a study. This greatly facilitates the correlation of environmental parameters to animal behaviours.

Although Movebank provides many useful and interesting features for researchers and data consumers alike, it remains to be seen whether it will succeed in being taken up as the premier repository for global animal tracking data by researchers. As a repository, its main competition will be institutional repositories or discipline-specific repositories such as Dryad (after which the Movebank Data Repository is modeled[8]). To try to gauge uptake of Movebank vis-à-vis Dryad, I conducted a brief comparison by searching Dryad for keywords commonly appearing in Movebank datasets. Although there did appear to be some location-based datasets in Dryad, searches seemed to result in a number of non-tracking datasets and datasets without geographic references. Unfortunately, a systematic and quantitative comparison is beyond the scope of this review so I was unable to ascertain how many datasets eligible for Movebank were published in Dryad instead. From a data user's perspective, Movebank is the more attractive of the two repositories due to the narrow focus and standardized data format. However, it may not have the same reputation as Dryad and some researchers may find the data format onerous.

During the comparison of repositories, I did make an illustrative discovery when I searched Dryad for the title of one recent dataset in the Movebank Data Repository ("Data from: Correcting for missing and irregular data in home-range estimation").
The study and article[9] compares tracking data for three individual animals of different species. The dataset in the Movebank Data Repository[10] was deposited by two of the eleven co-authors, covering one of the three species and conforming to Movebank metadata standards. However, a different dataset appears under the same title[11] in Dryad, attributed to all eleven co-authors, covering a different one of the three species and containing sparse, undocumented data. While this may seem an individual case of inconsistency within a research team, I see it as exemplary of the tension between data managers who create repositories and promote adherence to standards that will lead to robust, usable data collections, and busy researchers with competing priorities who want their papers and data published as smoothly as possible. In other words: data managers propose, researchers dispose.

---

[8] "Movebank Data Repository FAQ: Are there other repositories like this?"
https://www.movebank.org/node/2220#other_repositories.

[9] Fleming, C. H., Sheldon, D. , Fagan, W. F., Leimgruber, P. , Mueller, T. , Nandintsetseg, D. , Noonan, M. J., Olson, K. A., Setyawan, E. , Sianipar, A. and Calabrese, J. M. (2018), Correcting for missing and irregular data in home-range estimation. *Ecol Appl.* Accepted Author Manuscript. doi:10.1002/eap.1704.

[10] Setyawan E, Sianipar A (2018) Data from: Correcting for missing and irregular data in home-range estimation. Movebank Data Repository. doi:10.5441/001/1.3gj67c2k.

[11] Fleming CH, Sheldon D, Fagan WF, Leimgruber P, Mueller T, Nandintsetseg D, Noonan MJ, Olson KA, Setyawan E, Sianipar A, Calabrese JM (2018) Data from: Correcting for missing and irregular data in home-range estimation. Dryad Digital Repository. https://doi.org/10.5061/dryad.n42h0.