

Chapter 1

Wildlife Tracking Data Management: Chances Come from Difficulties

Holger Dettki, Ferdinando Urbano, Mathieu Basille
and Francesca Cagnacci

Abstract In recent years, new wildlife tracking and telemetry technologies have become available, leading to substantial growth in the volume of wildlife tracking data. In the future, one can expect an almost exponential increase in collected data as new sensors are integrated into current tracking systems. A crucial limitation for efficient use of telemetry data is a lack of infrastructure to collect, store and efficiently share the information. Large data sets generated by wildlife tracking equipment pose a number of challenges: to cope with this amount of data, a specific data management approach is needed, one designed to deal with data scalability, automatic data acquisition, long-term storage, efficient data retrieval, management of spatial and temporal information, multi-user support and data sharing and dissemination. The state-of-the-art technology to meet these challenges are relational database management systems (DBMSs), with their dedicated spatial extension. DBMSs are efficient, industry-standard tools for storage, fast retrieval and manipulation of large data sets, as well as data dissemination to client

H. Dettki (✉)
Umeå Center for Wireless Remote Animal Monitoring (UC-WRAM),
Department of Wildlife, Fish, and Environmental Studies,
Swedish University of Agricultural Sciences (SLU), Skogsmarksgränd,
SE-901 83 Umeå, Sweden
e-mail: holger.dettki@slu.se

F. Urbano
Università Iuav di Venezia, Santa Croce 191 Tolentini, 30135 Venice, Italy
e-mail: ferdi.urbano@gmail.com

M. Basille
Fort Lauderdale Research and Education Center, University of Florida,
3205 College Avenue, Fort Lauderdale, FL 33314, USA
e-mail: basille@ase-research.org

F. Cagnacci
Biodiversity and Molecular Ecology Department, Research and Innovation Centre,
Fondazione Edmund Mach, via E. Mach 1, 38010 S.Michele all'Adige, TN, Italy
e-mail: francesca.cagnacci@fmach.it

programs or Web interfaces. In the future, we expect the development of tools able to deal at the same time with both spatial and temporal dimensions of animal movement data, such as spatiotemporal databases.

Keywords GPS tracking · Large data set · Database management system · Spatial database

Introduction

In recent years, new wildlife tracking and telemetry technologies have become available, allowing remote data capture from a steadily increasing number of taxa, species and individual animals. This has resulted in a substantial increase in the volume of data gathered by researchers, environmental monitoring programs and public agencies. In the future, one can expect an almost exponential increase in collected data as new sensors, e.g. to monitor health status, interactions among individuals, or other animal-centred variables, are integrated into current bio-logging systems on animals. Data can be remotely transferred to operators (e.g. using Global System for Mobile Communications (GSM) networks or satellite systems such as Argos, Globalstar and Iridium), making near real-time monitoring of animals possible. Furthermore, positional information can now be complemented with a wide range of other information about the animals' environment made available by satellite remote sensing, meteorological models and other environmental observation systems.

The information embedded in animal-borne data sets is enormous and could be made available in a wider societal context than wildlife research or management. However, there is still a lack of suitable infrastructures to collect, store and efficiently share these data. In this chapter, and in the rest of the book, we offer a solution for a subset of animal-borne information, i.e. wildlife tracking data. With this term, we mainly refer to Global Positioning System (GPS)-based radiotelemetry (Cagnacci et al. 2010; Tomkiewicz et al. 2010). 'GPS' is here a synonym for all different existing or upcoming global navigation satellite system (GNSS)-based animal tracking systems. Most of the concepts and tools proposed in this book, however are also valid for other tracking systems (e.g. very high frequency (VHF) telemetry, radio frequency identification (RFID) systems, echolocation).

In the past, software tools for wildlife tracking studies were mainly developed on the basis of VHF radiotracking data, which are characterised by small and discontinuous data sets, and were focused on data analysis rather than data management (Urbano et al. 2010). Spatial data, such as animal locations and home ranges, were traditionally stored locally in flat files, accessible to a single user at a time and analysed by a number of independent applications without any common standards for interoperability. Now that GPS-based tracking data series have become the main reference, data management practices face new challenges.

As we discuss below, GPS telemetry usually provides locations separated by constant and short time intervals (varying from a few minutes to several hours) that accumulate in large data series. Thus, data should be securely, consistently and efficiently managed in order to minimise errors, increase the reliability and reproducibility of inferences and ensure data persistence (e.g. access to data on multiple occasions and by several persons). Further, there is an increasing call for sharing and distributing data to the global research community, for principle and opportunity, and wildlife tracking data are no exception. Indeed, deploying tracking devices and bio-logging sensors on wildlife is costly, and most projects are local or rely on limited sample sizes. Hence, to realise the full potential of locally collected data, researchers must be able to share with and distribute their data to the global research community (see [Chap. 13](#)).

Below, we summarise the requirements that wildlife tracking data represent in terms of data management, and opportunities offered by potential solutions. This analysis is largely drawn by Urbano et al. (2010), with updated considerations.

Requirements

The methodological approach and software architecture for managing wildlife tracking data have to meet the specific requirements of spatiotemporal data series which are the result of individual animals' behaviour. Thus, the first step is the definition of both data and marked animals' characteristics, as well as the users' needs.

- **Scalability:** GPS-based devices can currently record thousands of locations per animal over short time intervals (hours, days, months). The number of monitored individuals and species has steadily increased in recent years, due in part to decreases in costs, decreases in device size and availability of a growing range of device models. Data collected by additional bio-logging sensors can vastly increase the total amount of data collected. Data management methods must be able to accommodate this growing volume of data.
- **Periodic and automatic data acquisition:** Automated procedures to receive, process and store data from GPS telemetry devices are required when a near real-time data transfer system is provided by the tracking units.
- **Long-term storage for data reuse:** Data must be consistently stored and properly documented beyond the period of data collection and analysis to permit data archiving, reuse and sharing.
- **Efficient data retrieval:** As the data sets increase in size, effective data analysis depends on efficient data retrieval tools.
- **Management of spatial information:** GPS data are by definition spatiotemporal data (i.e. usually they represent moving objects). Retrieval, manipulation and management tools should then be specific to the spatial data domain.

- **Global spatial and time references:** The use of global time and spatial reference systems enables comparison with data sets from different regions and at different scales.
- **Heterogeneity of applications:** The complex nature of movement ecology requires that sensor data are visualised, explored and analysed by a wide range of task-oriented applications; therefore, the software architecture should support the integration of different software tools.
- **Integration of additional data sources:** Animal locations can be enhanced by other spatial (e.g. remote sensing, socioeconomic) and non-spatial information (e.g. capture details or life-history traits), as well as data from other bio-logging sensors; multiple spatial and non-spatial data sets should be correctly managed and efficiently integrated into a comprehensive data structure.
- **Multi-user support:** Wildlife tracking data sets are of interest to researchers, but also to a range of stakeholders, including for example public institutions (wildlife management offices, national parks), and private organisations (environmental groups, hunters). These users might need to access data simultaneously, both locally and remotely, with different access privileges.
- **Data sharing:** There is an increasing call for sharing data publicly or among research groups. This is discussed in more detail in [Chap. 13](#). Technically, data sharing requires adherence to standard data formats, definition of metadata and methods for data storage and management that, in turn, guarantee interoperability.
- **Data dissemination/outreach:** Dissemination of data to the scientific community or outreach activities targeting the general public is important to supporting management decisions, fundraising and promoting a larger awareness of issues related to ecosystem changes and resilience to changes. This requires the integration of specific tools to visualise and make data accessible (e.g. Web-based data interfaces, mapping tools, or search engines).
- **Cost-effectiveness:** By choosing cost-effective software tools that can meet the above requirements, funding can be focused on the collection and analysis of data, rather than on data management.

Chances

All of these requirements must be satisfied to take full advantage of the information that wildlife tracking devices can provide. As the volume and complexity of these data sets increase, the software tools used in the past by most animal tracking researchers are not sustainable, and thus there is an urgent need to adopt new software architectures.

Fortunately, software solutions exist and have a large user base. The reference solutions for data management are relational or object-relational database management systems (DBMSs), with their dedicated spatial extensions. DBMSs are efficient tools for storage, fast retrieval and manipulation of data (Urbano et al. 2010).

From a strictly technical point of view, advantages of DBMSs for tracking and movement ecology studies include the following:

- **Storage capacity:** Virtually any potential volume of data from wildlife GPS tracking or other sensor data can be stored in a DBMS.
- **Data integrity:** Data entry, changes and deletions can be checked to comply with specific rules.
- **Data consistency:** DBMSs fully support reversible transactions and transaction logging to ensure traceability of data operations and proper data management.
- **Automation of processes:** DBMSs can be empowered by defining internal functions and triggers; thus, a wide range of routinely complex work procedures can be automatically and efficiently performed inside the database itself.
- **Data retrieval performance:** The use of indexes effectively decreases querying time.
- **Management of temporal data types:** Time zones or daylight saving settings linked to temporal data types are supported and allow time consistency across study areas and times of year.
- **Reduced data redundancy:** The use of primary keys avoids data replication and the adoption of a normalized relational data model reduces data redundancy.
- **Client/server architecture:** Advanced DBMSs provide data through a central service, to which many applications can be connected and used as database front-end clients.
- **Advanced exploratory data analysis:** Data mining techniques for automatic knowledge discovery of information embedded in large spatial data sets must be applied in consistent and structured environments such as DBMSs.
- **Data models:** Data models are the logical core of DBMSs and allow linking and integration of data sources by means of complex relationships; this is not only necessary for consistently structuring the database, but is also an extremely useful way to force users to clarify the ecological/biological relational links between groups of data. This will be discussed extensively in [Chaps. 2, 3 and 4](#).
- **Multi-user environment:** Data can be accessed by multiple users at the same time, keeping control on the coherence between operations performed by them, and maintaining a structured data access policy (see below).
- **Data security:** A wide range of data access controls can be implemented, where each user is constrained to the use of specific sets of operations on defined subsets of data.
- **Standards:** Consolidated industry standards for databases, data structure and metadata facilitate interoperability with client applications and data sharing among different research groups (see [Chap. 13](#)).
- **Backup and recovery:** Regular backup and potential disaster recovery processes can be efficiently managed.
- **Cost-effectiveness:** Multiple open source DBMSs software solutions are available that have large user and development communities, as well as extensive free and low-cost resources for training and support.

Spatial and Spatiotemporal Extensions

In addition to the important features listed above, spatial tools are increasingly integrated within databases that now accommodate native spatial data types (e.g. points, lines, polygons, rasters). These spatial DBMSs are designed to store, query and manipulate spatial data, including spatial reference systems. In a spatial database, spatial data types can be manipulated by a spatial extension of the Structured Query Language (SQL), where complex spatial queries can be generated and optimised with specific spatial indexes. Today, all major DBMS providers offer native spatial capabilities and functions in their products.

Spatial databases can easily be integrated with Geographical Information System (GIS) software, which can be used as client applications. Further, few desktop GIS are optimised for managing large vector data sets and complex data structures. Spatial databases, instead, are the tool of choice for performing simple spatial operations on a large set of elements. Thus, simple but massive operations on raw data can be preprocessed within the spatial database itself, while more advanced spatial analysis on subsequent data sets can rely on GIS and the spatial statistics packages connected to it.

A further promising extension to spatial data models is the adoption of spatiotemporal data models (e.g. Kim et al. 2000; Pelekis et al. 2004; Güting and Schneider 2005). In these models, locations are characterised by both a spatial and a temporal dimension that are combined into one unique, double-faced attribute of movement. Spatiotemporal databases will extend the spatial data model for animals by integrating data types and functions specifically related to the spatiotemporal nature of animal movements (e.g. considering ‘movement’ as an attribute of the animal instead of relying on clusters of location objects with timestamps). This approach would help to decipher the relationships between animal movement, habitat use and environmental conditions. Although commonly used DBMSs do not yet support an integrated spatiotemporal extension, spatiotemporal databases (e.g. SECONDO¹, Güting et al. 2004), which are undergoing intense development, will be the natural evolution for wildlife tracking data management tools in the future.

References

- Cagnacci F, Boitani L, Powell RA, Boyce MS (2010) Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. *Philos Trans R Soc B* 365:2157–2163. doi:10.1098/rstb.2010.0107
- Güting RH, Behr T, de Almeida VT, Ding Z, Hoffmann F, Spiekermann M (2004) SECONDO: an extensible DBMS architecture and prototype. *Fernuniversität Hagen, Informatik-Report* 313

¹ <http://dna.fernuni-hagen.de/secondo/>.

- Güting RH, Schneider M (2005) Moving objects databases. Academic Press, USA. ISBN 978-0-12-088799-6
- Kim DH, Ryo KH, Kim HS (2000) A spatiotemporal database model and query language. *J Syst Softw* 55:129–149
- Pelekis N, Theodoulidis B, Kopanakis I, Theodoridis Y (2004) Literature review of spatio-temporal database models. *Knowl Eng Rev* 19:235–274. doi:[10.1017/S026988890400013X](https://doi.org/10.1017/S026988890400013X)
- Tomkiewicz SM, Fuller MR, Kie JG, Bates KK (2010) Global positioning system and associated technologies in animal behaviour and ecological research. *Philos Trans R Soc B* 365:2163–2176. doi:[10.1098/rstb.2010.0090](https://doi.org/10.1098/rstb.2010.0090)
- Urbano F, Cagnacci F, Calenge C, Dettki H, Cameron A, Neteler M (2010) Wildlife tracking data management: a new vision. *Philos Trans R Soc B* 365:2177–2185. doi:[10.1098/rstb%202010.0081](https://doi.org/10.1098/rstb%202010.0081)