Brigham Young University
BYU ScholarsArchive

International Congress on Environmental Modelling and Software

7th International Congress on Environmental Modelling and Software - San Diego, California, USA - June 2014

Jun 18th, 9:00 AM - 10:20 AM

Wireless Remote Animal Monitoring (WRAM) - A new international database e-infrastructure for management and sharing of telemetry sensor data from fish and wildlife

Holger Dettki
Swedish University of Agricultural Sciences, holger.dettki@slu.se

Michel Brode
Swedish University of Agricultural Sciences

Timothy Giles
Swedish University of Agricultural Sciences

Jerry Hallgren
Swedish University of Agricultural Sciences

Follow this and additional works at: https://scholarsarchive.byu.edu/iemssconference

Dettki, Holger; Brode, Michel; Giles, Timothy; and Hallgren, Jerry, "Wireless Remote Animal Monitoring (WRAM) - A new international database e-infrastructure for management and sharing of telemetry sensor data from fish and wildlife" (2014). *International Congress on Environmental Modelling and Software*. 19.
https://scholarsarchive.byu.edu/iemssconference/2014/Stream-A/19

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

International Environmental Modelling and Software Society (iEMSs)
7th Int. Congress on Env. Modelling and Software, San Diego, CA, USA,
Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

Wireless Remote Animal Monitoring (WRAM) - A new international database e-infrastructure for management and sharing of telemetry sensor data from fish and wildlife

Holger Dettki[1], Michel Brode[2], Ivan Clegg[2], Timothy Giles[3], Jerry Hallgren[2]

[1]*Swedish University of Agricultural Sciences (SLU), Umeå Center for Wireless Remote Animal Monitoring (UC-WRAM), Dept. of Wildlife, Fish, and Environmental Studies, Skogsmarksgränd, SE – 90183 Umeå, Sweden, holger.dettki@slu.se.*
[2]*Swedish University of Agricultural Sciences (SLU), IT-Dept., Vindbrovägen 1, SE – 75007 Uppsala, Sweden.*
[3]*Swedish University of Agricultural Sciences (SLU), IT-Dept., Skogsmarksgränd, SE – 90183 Umeå, Sweden*

Abstract: New tracking technologies have become available to ecologists, allowing remote real-time data capture from an increasing number of taxa, species and animals. To realize the potential of the data, researchers must be able to share data and collaborate with the global research community. The Wireless Remote Animal Monitoring (WRAM) database system, started in 2003, contains 96.5 million positions and other sensor data and is used to date by 32 user groups from 8 countries, tracking 16 species and 2 179 individual animals. The infrastructure represents the Swedish national data node and data sharing portal for real-time telemetry sensor data from fish and wildlife. WRAM will be part of Swedish LifeWatch and is cooperating with several other international database initiatives. The infrastructure consists of 2 main parts: 1) The WRAM Data Warehouse (WDW) is a high performance data warehouse for real-time 'big data' as position, acceleration, or heartbeat data from fish and wildlife, hosted by the High Performance Computing Center North (HPC2N) in Umeå, Sweden. 2) The WRAM Data Broker (WDB) is the single-sign-on web interface federating the WDW with other similar database systems around the world as Movebank, EuroDeer or CAnMove to enable seamless querying across systems and easy data sharing between data owners, while honoring local authentication and authorization settings. Query results are also accessible through ODBC in a temporary database, enabling users to use local analysis tools for further analyses. Here, we give an overview over the different parts of the e-infrastructure, including automated data capture tools, database models, data sharing approaches, and web-based visualization tools.

***Keywords**:* bio-telemetry; wildlife; database network; tracking; animal monitoring**;** GPS

1 INTRODUCTION

During recent years new tracking and telemetry technologies have become available, allowing remote data capture from a steadily increasing number of taxa, species and individual animals (Urbano et al., 2010). This has resulted in a substantial increase in data volume gathered by research, environmental monitoring and public agencies. In the future, one can expect a rapid increase in collected data as new sensors monitoring e.g., health status, interactions among individuals, or other environmental variables are integrated into units deployed on animals. A crucial limitation for efficient use of the data in the wider society is that in many cases basic research still lacks the infrastructure to collect, store and efficiently share these data, in order to use the full potential of these data. Restricting this to animal research, it is often very costly to deploy tracking devices and bio-sensors to wildlife, resulting in only few sensors being used in a single project. Hence, to realize the full potential of locally collected data, researchers must be able to share and distribute their data with the global research community. While data management, storage and visualization is addressed recently by several authors (e.g., Cagnacci and

Urbano, 2008; Hartog et al., 2009; Urbano et al., 2010; Kranstauber et al., 2011; Dettki et al., 2014), automatic data capture and data sharing remains a major challenge for many researchers (Coyne and Godley, 2005).

In conclusion, it has become impractical, if not impossible, for single research groups to build and maintain local database infrastructures to facilitate all needs and demands from the international research community (Hartog et al., 2009; Urbano et al., 2010; Dettki et al., 2014). A centralized database e-infrastructure for sensor data from fish, wildlife, and domesticated animals as provided by WRAM, enables researcher to focus on world-class data analysis instead of local data handling and management, providing the pre-conditions for cutting edge research.

The 'Wireless Remote Animal Monitoring' (WRAM) is a national Swedish e-infrastructure for research and management in animal ecology, behavior, and physiology. It is part of the Swedish LifeWatch (2014) program and is complementing this 'biodiversity' focused program with the ability to handle spatial and non-spatial bio-sensor data. WRAM provides software tools and a database infrastructure to automatically capture, store, share, and analyze sensor data from animals, currently primarily from fish and wildlife. It has provided these services since 2003 for an ever-growing number of research and management projects in Scandinavia. Currently the system stores location, acceleration, proximity, and temperature data for 2 179 animals from 16 species in 8 countries, resulting in 96.5 million database records (May 2014), making it one of the largest database for sensor data from fish and wildlife in the world. The system handles at the moment 13 data formats from 9 different sensor providers. The system is of wide national and international interest, as it is used by 32 different projects, representing both highly qualified research projects as the Scandinavian Brown Bear Research Project (SBBRP, 2014) or the Fennoscandian Moose Research initiative (FennoScandMoose, 2014), as well as wildlife management projects on a European level as the EU Life-funded Raccoondog-management project (MIRDINEC, 2013).

2 WRAM INFRASTRUCTURE OVERVIEW

The WRAM 2 infrastructure represents an open platform- and location-independent database network e-infrastructure. In general, it is build up in a modular way, to enable upgrades to different modules when needed, without affecting the overall structure or functionality of the system.

It consists of two main structures (Fig. 1), the WRAM Data Warehouse (WDW) and the WRAM Data Broker (WDB). The WDW main data storage uses backend spatio-relational databases building on Microsoft SQL Server 2012, while the frontend database, accessible through the WDB, features both a spatial enabled PostgreSQL engine and a Mircosoft SQL Server 2012 engine. WRAM 2 is an open access infrastructure project, and all source code developed within WRAM 2 will be documented and published in appropriate Open Source repositories and the WRAM web site after the end of the development phase in July 2014. Currently (March 2014) the following system components of the WRAM 2 system are under development or tested as prototypes.

2.1 WRAM Data Warehouse (WDW)

The WDW is a Data Warehouse for the reception, data cleaning, data aggregation and data publication for sensor data from fish and wildlife. It adapts established standards for spatial data (e.g. Darwin Core, 2009; INSPIRE, 2013; OGC, 2013) and uses these standards to ensure the highest possible probability to share or combine datasets with other systems across networks. To enable the WDW to cope with 'big-data' in real-time, it is hosted by the High Performance Computing Center North (HPC2N) in Umeå, Sweden. HPC2N is one of six national High Performance Computing Centers tasked by the Swedish Research Council to provide computing and grid services to the Swedish research community. The WDW features a staging area to process operational data before making it available in the Warehouse. In the WDW meta-, summary-, and raw data will be available as well as several Data Marts, i.e., views customized to the different requirements of the users (Fig. 2).

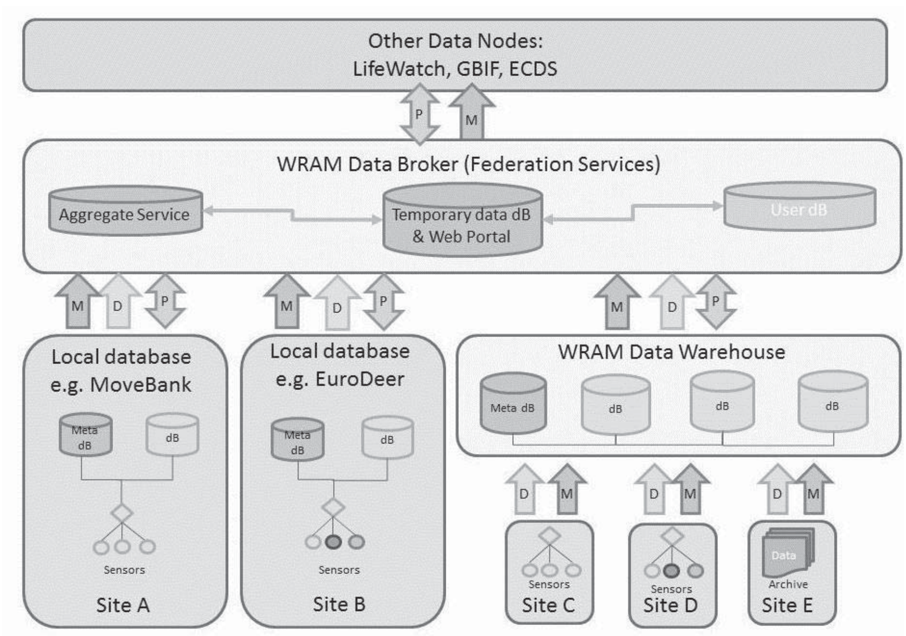

Fig. 1. WRAM 2 e-infrastructure (WDW & WDB): The WRAM Data Warehouse accepts data and meta data from different sources and links up with other local databases in the WRAM Data Broker. Here, a web portal and a temporary database are hosting results from user queries, aggregated data is submitted to other, high level data nodes, and user access is managed.

WRAM automates the process of retrieving and parsing data from different sensors and providers into the database as far as possible, enabling users to concentrate on analysis and interpretation of the data instead of data management. Where manual data upload is necessary, e.g., when data is downloaded manually from the sensor, the user simply emails the data file in the original proprietary format to a functional email address. This also enables users to set up their own automatic receiving stations and send off the received data files to the central WRAM database. This is particularly important when using the GSM system or equivalent cell phone network for data transfer, as sending data-SMS between different cell phone network providers or countries can increase the data transfer costs many times due to network roaming.

The way data can be collected by the system depends on the proprietary infrastructure of the sensor provider. While some providers allow the users to set up their own receiving stations to receive data-SMS through a local GSM-modem using their proprietary software to decode and export the data to a text file, other providers use a company-owned receiving system. They are then simply using a 'Push' system sending emails with the data in the email-body or as attachment to the user. Additionally, 'Pull' systems, as e.g., web services can make the data available on the provider's side. The WRAM system can handle sensor data sent by Argos, Iridium, or GSM in a uniform way: The data is received as SMS or email by proprietary software of the sensor provider or as email by the WRAM program 'ReadMail', if necessary parsed into a readable format, and exported into a standardized text file. These text files are then read by the 'WRAM Data Turbine' and inserted record-wise into the database. Additionally, web services can be read on a regular basis. The user can also send an email with any manually retrieved data from the sensors to the program 'ReadMail'. The system is working with both Vectronic Aerospace GmbH (VAS) and Lotek, Inc. hardware (GPS/GSM-units & GPS/Iridium-units) and software (GPS Plus, GPS Plus X), Televilt AB (TGV) GPSDirect and Tellus GPS/GSM-units, FollowIT AB (FIT) GPS/GSM-Tellus and Tellus Domestic GPS/GSM-collars, EcoTone GPS/GSM-units, Cellular Tracking Technologies (CTT) GPS/GSM-units, and Microwave Telemetry, Inc. (MTI) GPS/Argos-units and software (GPS Parser). Data from the Norwegian provider Telespor AS is pulled as XML-data from their web service. Data retrieved manually from the units through a VHF or UHF-link or by cable-download can be sent into the database by email for VAS, Lotek, and Telemetry Solutions, Inc. GPS-units.

To enhance performance and allow for both excellent end user query performance and real-time data reception and loading (Extract-Transform-Load, ETL) at the same time, the following design is utilized (Fig. 2). The WDW is split into two physical data warehouses with an identical structure to facilitate different needs from different user groups. While the 'Real Time WRAM Data Warehouse' will contain only data current, 'real-time' data not older than e.g., 4 weeks, the 'Historic WRAM Data Warehouse' will hold all data older than 4 weeks. By this, the relatively small, but 'most-used' data on a daily basis is accessible to many users with excellent performance, while less often performed queries against the complete dataset of e.g., a project for analysis or publication will run against the bulk of the data storage, with slightly lower performance. The time frame for the Real Time and Historic WDW is fully configurable and can be changed from 4 weeks to any other time frame if need arises. Two ETL processes are in place after the data is received and cleared from the staging area. The first process (Real Time ETL) runs every five minutes and loads new data into either the Real Time or Historic WDW. A second 'Daily ETL' process moves data from the Real Time WDW and into the Historic WDW when it is 4 weeks old. In addition, aggregation data is calculated to be forwarded to the Swedish LifeWatch (2014) Analysis Portal. To guarantee data quality, two Quality Control and Correction (QC+C) processes are developed (Fig. 2). The Real Time QC+C will run only rules that do not require excessive processing time. The Daily QC+C process will run the rest of the rules on recently received records in both data warehouses. The individual rules can be tailored towards different sensors, species or specific project needs. All QC+C rules and results will be recorded as metadata so that users can view the data treatment and confidence level of the individual data records. Additionally, metadata about the data accuracy is recorded, based on the details of the sensors, to provide information on how accurate the given measurement is.

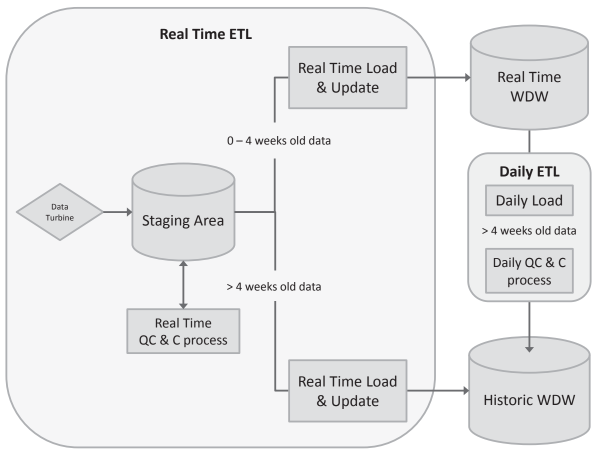

Fig. 2. WRAM Data Warehouse data flow: The Real Time ETL handles data collection and quality control in a staging area and distributes the data, depending on the age of the measurement, to the Real Time or Historic WDW. A Daily ETL moves older data from the Real Time to the Historic WDW and checks data quality.

Within the WDW, a main core Normalized Data Store (NDS) is used to store all the data received from the data providers. All measurement data is stored in a single generic table by serializing the individual sensor measurements (Fig. 3). This enables the system to easily add new, future sensors to the system without the need to apply structural changes to the underlying database. This system also enables the storage of other types of data, not directly connected to sensors, as e.g., animal measurements as weight or size, and results from genetic analyses of bio-samples. The end user, however, is presented with more traditional data views and new possibilities for multidimensional cube-analyses.

Information on projects, animals, tags, sensors, and measurements are organized in hierarchical information layers, adhering to the normalized data structure (Fig. 4). By this, it is easy for the end user to administer e.g., deployment of tags with one or more sensors onto animals, which can belong to one or more projects, or the re-use of sensors or tags on different animals. This structure also aids the security model of the WDW, which can restrict data access at any level and combination for any user. The data owner is responsible to set the appropriate permissions for the different users or user groups. This can e.g., include data access to several animals during a specific time period, or a more restricted access to an individual animal's sensor reading over a longer time period.

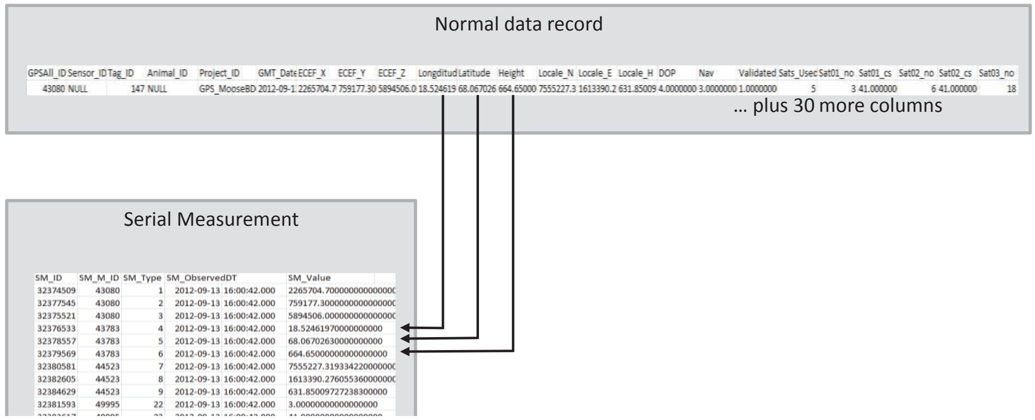

Fig. 3. An example showing the serialization of sensor measurements compared to older 'parallel table solutions'

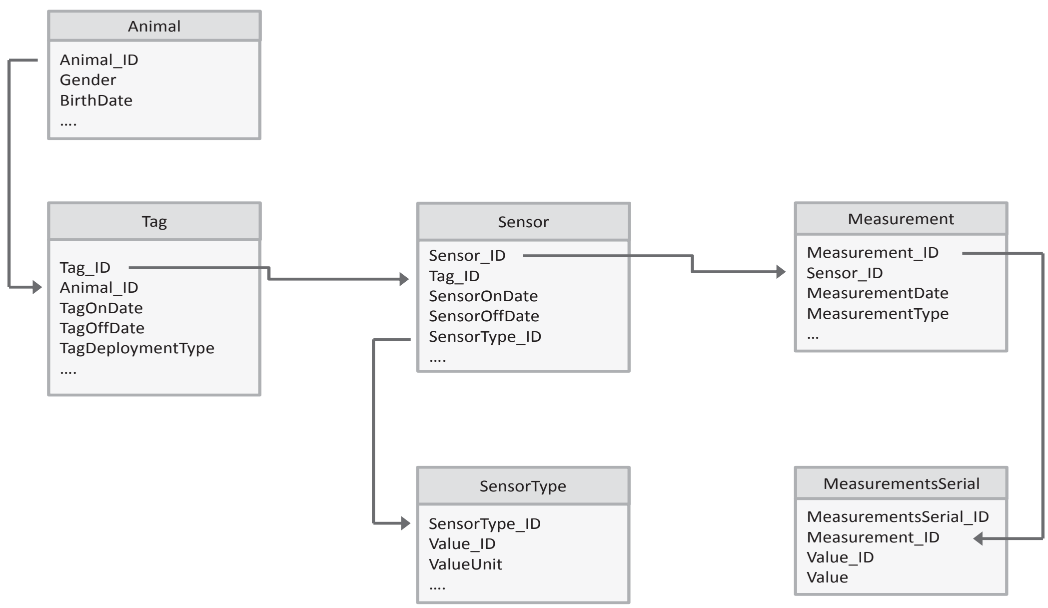

Fig. 4. Normalized structure of the information flow from animals to sensor readings.

2.2 WRAM Data Broker (WDB)

The WDB enables end users and data owners to query data from a number of data providers on sensor data from animals with a single sign-on through one single access point, the WDB Client Web Portal (CWP). Currently, several data providers such as Movebank (Kranstauber et al., 2011) and CAnMove (2014) are in the process of connecting to the WDB. The data is retrieved from the federated database systems in real-time, if necessary translated into a unified format and presented back to the user. The WDB has been designed and implemented in a modular and extendible fashion. This leads to a more evolutionary development cycle in which parts of the full system can be easily updated, extended or even completely replaced if desired or needed without affecting the rest of the system. The WDB consists of several sub-systems which are described below.
The WDB Client Web Portal (CWP) provides end users with the ability to make queries against data stored in one or more database systems connected in the WDB and visualize the data from the various data sources they have access to. The resulting data is available both in a tabular format and – for

spatial data – as maps. It is built on prior developments (Dettki et al., 2004) and will also offer some analysis tools to the participating projects which can be used either directly in the web application or locally through Web Processing Services (OGC, 2013). However, it must be stressed that it is our firm belief that most researchers never will rely only on static built-in web analysis tools but will always try to further develop and perfect local analytical methods. Hence, it is more important to supply the users with an easy way to access the data with the user's own statistical packages or tools of choice. This is provided by the frontend-database of the WDB. Users can also administer the data sources, check and upload data in to the WDW and administer their tags and animals in the CWP.
Further, the WDB aggregates suitable data into an open 'biodiversity' format and forwards it to Swedish LifeWatch (2014) and GBIF (2014). Hence, the data becomes discoverable through e.g., the Swedish LifeWatch (2014) Analysis Portal.
There are two levels of authentication and authorization. The first level is within the WDB via the Client Web Portal, while the second occurs with each data provider. First a user must register with the WDB system in order to make queries. This requires the user's email, name, organization and an OpenID. An OpenID account is free and available via OpenID providers as e.g., Google, Wordpress, Blogger, or Yahoo, and is used as a means to authenticate users. After registration the user can make queries in the portal. Which data providers that are listed for a particular user to choose from is based on whether the user has permissions at each data provider to access data from. The second level of authentication and authorization is used every time a user requests data from a particular data provider or across multiple data providers. At this level, the user is authenticated with each data provider and the data returned is based on what the user is authorized to access. Authorization at this level will be administered by the data providers or data owners. By this, the internal security model of each external data providers remains unaffected by the cooperation with WRAM.

2.3 Frontend Database (FEDB)

One of the novel approaches in the WDB is the ability of a user to access query results returned by the CWP with his local analysis tools in the same way as one would access a local database. An option in the CWP allows the user to have the result of a query be stored temporarily in a spatial database, which the user can then access using a tool of their choice (e.g., ArcGIS, R Statistics, MS Access.) for a specific period of time using the standard ODBC protocol for data access. The user can choose between SQL Server 2012 or PostgreSQL 9.2/PostGIS 2.0.1 spatial database engines. The data is stored in a temporary table and only available for a limited amount of time to avoid long-term caching of data from different data providers, as this would make the WRAM infrastructure effectively a central database for all data instead of a database federation system.

3 CONCLUSIONS AND RECOMMENDATIONS

The WRAM e-infrastructure represents an internationally much needed environmental software tool. Opposed to other similar infrastructures as Movebank (Kranstauber et al., 2011) or EuroDeer (2014), it not only offers long-term storage of sensor data from animals, but features a true data federation layer. This enables researchers around the world to easily cooperate and share data with each other, while keeping full local control of their data and sharing permissions. The novel front-end database access bypasses earlier data access limitations and enables simple data access from any ODBC-enabled local analysis-program. The infrastructure is open and free to use for all researchers and managers tasked with the handling and analysis of sensor data from all animals; primarily at this stage fish, wildlife, and domesticated species. WRAM strives to give researchers and managers free access to the collected data in WRAM or in the federated national and international databases by enabling them to share and merge data with each other across project or national borders. However, all data in the WDW is property of the research groups who originally delivered the data or made the data accessible. Hence, it is the responsible of the data owners to decide on data sharing and publishing policies for their data sets according to their formalized data publishing plans which are part of the individual project plans.
The WRAM e-infrastructure offers research groups an indefinite data storage and sharing opportunity, which can be used even after the end of the individual research project. By this, the system over time will become an international data repository and archive for animal sensor data, which can be used to utilize long term environmental studies on animals across national borders.
All developed source code comprising the infrastructure is open-source and can be freely used by researchers and managers to set up clones of the WRAM Data Warehouse (WDW) to be utilized in

other disciplines or countries. All WRAM software tools, open-source code, or database structures will be easily accessible for registered users through the WRAM website web-page (www.slu.se/wram/) after the development phase is finished in July 2014. User registration is using industry-standard OpenID-technology and hence is independent of organization-based user authentication methods, which greatly simplifies international collaboration between research or management organizations.
The scientific results of the infrastructure in terms of publications (currently 165 entries) or published datasets are continuously summarized and described at the WRAM website (www.slu.se/wram) and at national and international conferences. For example, to date, at least 43 peer-reviewed conference contributions using data in the WRAM system were reported to WRAM.

ACKNOWLEDGMENTS

We want to thank the numerous users for many suggestions over the years improving the current WRAM system, Fredrik Stenbacka, Åke Nordström and Eric Andersson for identifying gaps and errors in the system during field work, and Ulf Tingelöv and Martin Norrsken-Ericsson for early help with development. We are grateful for the financial support from the Swedish Research Council and the Kempe Foundation.

REFERENCES

Cagnacci, F., Urbano, F., 2008. Managing wildlife: A spatial information system for GPS collars data. Environ. Modell. Softw. 23, 957-959.
CAnMove, 2014. Centre for Animal Movement Research. http://canmove.lu.se (last accessed 05.03.2014.).
Coyne, M.S., Godley, B.J., 2005. Satellite Tracking and Analysis Tool (STAT): an integrated system for archiving, analyzing and mapping animal tracking data. Mar. Ecol. Prog. Ser. 301, 1-7.
Darwin Core, 2009. Project website. http://rs.tdwg.org/dwc (last accessed 05.03.2014.).
Dettki, H., Ericsson, G., Edenius, L., 2004. Real-time moose tracking: an internet based mapping application using GPS/GSM-collars. Alces 40, 13-21.
Dettki, H., Urbano, F., Basille, M., Cagnacci, F., 2014. Wildlife tracking data management: changes come from difficulties. In: Urbano, F., Cagnacci, F. (Eds.), Spatial Database for GPS Wildlife Tracking Data, Springer, Switzerland, pp. 1-7.
EuroDeer, 2011. Project website. http://sites.google.com/site/eurodeerproject (last accessed 04.03.2014.).
FennoScandMoose, 2014. Project website. http://www.moose-research.se (last accessed 04.03.2014.).
GBIF, 2014. Global Biodiversity Information Facility. http://www.gbif.org (last accessed 04.03.2014.).
Hartog, J.R., Patterson, T.A., Hartmann, K., Jumppanen, P., Cooper, S., Bradford, R., 2009. Developing integrated database systems for the management of electronic tagging data. In: Nielsen, J.L., Arrizabalaga,H., Fragoso, N., Hobday, A., Lutcavage, M., Sibert, J. (Eds.), Tagging and Tracking of Marine Animals with Electronic Devices, Springer, The Netherlands, pp. 367–380.
INSPIRE, 2013. Infrastructure for Spatial Information in the European Community. http://inspire.jrc.ec.europa.eu (last accessed 05.03.2014.).
Kranstauber, B., Cameron, A., Weinzerl, R., Fountain, T., Tilak, S., Wikelski, M., Kays, R., 2011. The Movebank data model for animal tracking. Environ. Modell. Softw. 26, 834-835.
MIRDINEC, 2013. Management of the invasive Raccoon Dog (Nyctereutes procyonoides) in the north-European countries. http://ec.europa.eu/environment/life/project/Projects/index.cfm?fuseaction=search.dspPage&n_proj_id=3784&docType=pdf (last accessed 05.03.2014.).
OGC, 2013. Open Geospatial Consortium, Inc. http://www.opengeospatial.org (last accessed 05.03.2014.).
SBBRP, 2014. The Scandinavian Brown Bear Research Project. http://www.bearproject.info (last accessed 04.03.2014.).
Swedish LifeWatch, 2014. Project website. http://www.svenskalifewatch.se/en (last accessed 04.03.2014.).
Urbano, F., Cagnacci, F., Calenge, C., Dettki, H., Cameron, A., Neteler, M., 2010. Wildlife tracking data management: a new vision. Phil. Trans. R. Soc. B 365, 2177-2185.