Transactions in GIS, 2009, 13(s1): 63–83

Research Article

A Customization of the Arc Marine Data Model to Support Whale Tracking via Satellite Telemetry

Brett K Lord-Castillo
Office of Emergency Management
St. Louis County Police

Bruce R Mate
Marine Mammal Institute, Oregon State University

Dawn J Wright
Department of Geosciences
Oregon State University

Tomas Follett
Marine Mammal Institute, Oregon State University

Abstract

The Arc Marine data model is a generalized template to guide the implementation of geographic information systems (GIS) projects in the marine environment. It developed out of a collaborative process involving research and industry shareholders in coastal and marine research. This template models and attempts to standardize common ocean and coastal data types to facilitate data sharing and analytical tool development. In this study, Arc Marine is extended from its core model to fit the research goals of the whale satellite-telemetry-tagging program of the Oregon State University Marine Mammal Institute (MMI). The study sought the best customization of the generic Arc Marine data model to enhance the key advantages of satellite telemetry for mapping the distribution and movement of endangered marine mammal species. It was found that three new groups of object classes were needed (Animal, Telemetry, and Operations). Further customization involved the development of a comprehensive framework for animal tracking with Argos satellite telemetry data. A new multidimensional data cube model was also devised, showing how this extension of Arc Marine serves as an appropriate target schema for the application of on-line analytical processing (OLAP) tools and spatial data mining of satellite telemetry tracking datasets.

Address for correspondence: Dawn Wright, Department of Geosciences, 104 Wilkinson Hall, Oregon State University, Corvallis, OR 97331-5506, USA. E-mail: dawn@dusk.geo.orst.edu

doi: 10.1111/j.1467-9671.2009.01159.x

1 Introduction

The management of the nation's marine ecological resources depends on the constant improvement of scientific methods and information resources among the researchers of the marine community. These improvements must come in the form of better information and better access to information. To this end, the marine community must develop standard methods of data management and analysis, which provide rapid dissemination of data, easy comparability of research findings, and simple means to carry out complex analysis. These are among the goals of the Arc Marine data model, which is a schema devised by Wright et al. (2007) to support the collection and management of multidimensional spatial data within a variety of marine applications, including mapping the ocean floor, fisheries management in the water column, marine animal tracking in the water column and on the sea surface, shoreline change, and the integration of physical oceanographic numerical models. The empty geodatabase resulting from the Arc Marine schema when it is translated to ArcGIS via ESRI's Schema Creation Wizard can then be filled with data, automatically organized with appropriate feature classes and relationships for assembling, managing, analyzing, and even publishing data for these application areas. It thus provides a framework for developing associated analytical tools that take advantage of known entities in the resulting geodatabase.

Arc Marine was initially tested and vetted via an initial group of thirteen case studies from the U.S. and Europe and in several of the aforementioned marine application areas (all described in Wright et al. (2007) and consisting mainly of slight modifications of existing core feature classes and subsequent mapping and analysis with the resulting geodatabase). Indeed data models often serve as a starting point that practitioners can extend by adding their own attributes, tables and relationships to suit their specific application needs (Figure 1). This article presents a new case study of sorts in the marine animal-tracking realm. It is unique in that it develops completely new feature classes, but also a large number of non-spatial *object classes* (with explicit relationships to the spatial feature classes). The purpose is to demonstrate the great utility of Arc Marine for satellite telemetry tracking of marine mammals (in addition to the use of ship and aircraft tracking data in Arc Marine as first described in Halpin et al. (2004)). For the purposes

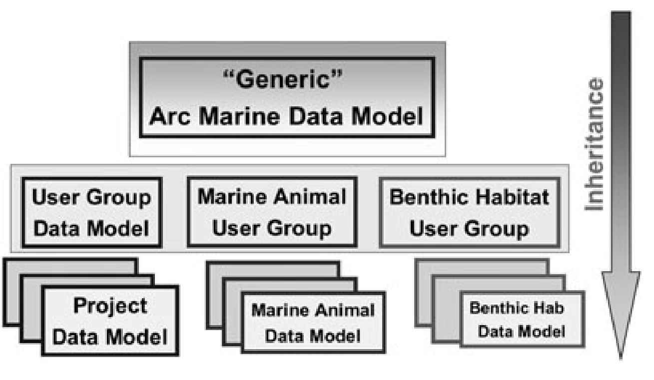

Figure 1 Design strategy for Arc Marine where the core data model is as generic as possible, but can be customized by various user groups (such as marine animal tracking or other applications) for use within their specific enterprise (after Wright et al. 2007)

of customization, the most important structural aspect of the data model is the availability of standardized classes to represent model entities and the relational joins, built into the model, which provide guaranteed relationships between data tables for complex querying.

Marine animal tracking is a central component of marine ecology research into the patterns of movement and distribution of economically important and endangered marine species. These movement patterns and distributions, coupled with habitat classification, drive the management of marine ecological resources, especially in the context of the Marine Mammal Protection Act (U.S. House, 110th Congress, 2007a, b). Marine animal tracking attempts to answer such critical questions as: What are the spatial and temporal distributions of key animal populations? How do species interact with each other across their ranges? What is the relationship between physical and biological processes and the distribution and movement of critical marine species? While answering these questions has always required an expertise in the biological sciences, knowledge of satellite remote sensing has becoming increasingly important.

Satellite telemetry through the Argos system (Block et al. 1998, Le Boeuf et al. 2000, Mate et al. 2007) has provided a major breakthrough to researchers seeking to examine the behaviors of marine vertebrates without time consuming and costly direct observation. Over the years, a wide range of Earth-bound methods (vessel surveys, aerial photogrammetry, surface trawling, marine acoustics) have been used to sample specific areas and generate home range analysis, population densities, stock assessments, and other area-based statistics (e.g. Boehlert et al. 2001, Duke University Marine Laboratory 2004, Block 2005, Block et al. 2005). The key advantages of satellite telemetry over ship and aircraft observation of animals is that it allows for continuous, synoptic coverage at the sea surface, autonomous profiling of the animals at depth, and unprecedented timeliness of observation (Lagerquist et al 2000, Boehlert et al. 2001, Block 2005). Rather than relying on a single observation via a survey, or even a time series, the satellite feed allows for the return of data from the animal in hours or even minutes with continuous updating. This makes it possible to track animal migration, feeding, and breeding patterns as of last week, and future weeks into the following seasons. With continuous coverage, one can begin to examine the movement of the animal to capture these aspects of the species life history.

However, continuous coverage brings very large spatial datasets, leading further to the basic question of how best to explore and integrate these data at a wide range of extents and grains, and how to find relationships between animal distributions and oceanographic processes. Currently such questions are only being explored at the level of data repositories and scattered pilot projects. Even at the largest such repository, the Ocean Biogeographic Information System – Spatial Ecological Analysis of Megavertebrate Populations project (OBIS-SEAMAP, Read et al. 2006), tools exist for mapping and animation of any dataset, but only four species-specific pilot projects have begun to relate marine biogeographic data to other spatial datasets. One of the key challenges is the development of a common set of computational and data management approaches (Block 2005), one of which is a GIS data model.

The Oregon State University Marine Mammal Institute (MMI) specializes in the use of satellite telemetry tags to track the movements of the "great whales" (i.e. the right, blue, humpback, fin, gray, and sperm whales), primarily along the Pacific and Gulf coasts of North America (Mate 1989, Mate et al. 1994). They are aided in this effort by a close collaboration with the Census of Marine Life's Tagging of Pacific Predators or TOPP

(Block et al. 2000). By mapping the distributions and abundance of whales throughout their migration, feeding and breeding activities, the MMI has filled unknown portions of the life histories of these species and hopes further to identify anthropogenic activities (e.g. collisions with ships) that have dwindled their numbers and to pose management solutions to aid in their recovery (Sherman 2006). Four whale species (blue, fin, sperm, and humpback) have been chosen for closer study by TOPP and MMI due to their significantly overlapping home ranges. Not only does this aid in data analysis, but these overlapping ranges allow also for the tagging of multiple species in deployment operations. The datasets produced are widely dispersed chronologically, seasonally and spatially, and are used with data from bathymetry and chemical, physical, and biological oceanography. Though much can be learned just from the descriptive aspects of this information base, discovering the distributions and movements of endangered species whose critical habitats are unknown for most of the year will require the consolidation and aggregation of these data into a unified spatial database. Hence, bringing the MMI satellite telemetry program together with Arc Marine generates the central research question: *What is the best customization of the generic Arc Marine data model to enhance the key advantages of satellite telemetry for mapping the distribution and movement of endangered marine mammal species?*

2 Methods

2.1 Argos Satellite Altimetry Data

The specific field methods employed in whale tagging for satellite altimetry are outside the scope of this article, but are covered in detail (including historical development of tag hardware and deployment techniques) by Mate et al. (2007). Briefly, in terms of data collection and geolocation, Argos consists of a network of four polar-orbiting satellites, which circle the Earth every 101 minutes. Argos platform transmitters resolve location using Doppler shift principles (Figure 2). Using the assumption of a stable transmission frequency and motionless platform, the system can use multiple measures of Doppler frequency shift to construct a circular solution set (intersections of spherical distance) around the satellite's path of motion. By collecting multiple transmissions (Service Argos requires a minimum of four passes), the location of the platform is determined with greater accuracy, resolving into the intersection of this circular solution with the elevation sphere of the platform (Figure 2).

2.2 Customizing Arc Marine

Arc Marine defines a structure for storing in a GIS a host of deep ocean and coastal features observed or gathered from multiple surveying platforms and instruments. The schema produces the standard, unified database mentioned above, and the associated community-wide advantages of data interoperability, while reducing analytic complexity and facilitating tool development. In this study, and as suggested by Wright et al. (2007), the core of the data model was kept intact. All core classes retain their original attributes to ensure compatibility with code from other sources. As such, class inheritance was used to create customized versions of core classes, which were still made as generic and universally applicable as possible to encourage reuse of the customization by other developers.

A

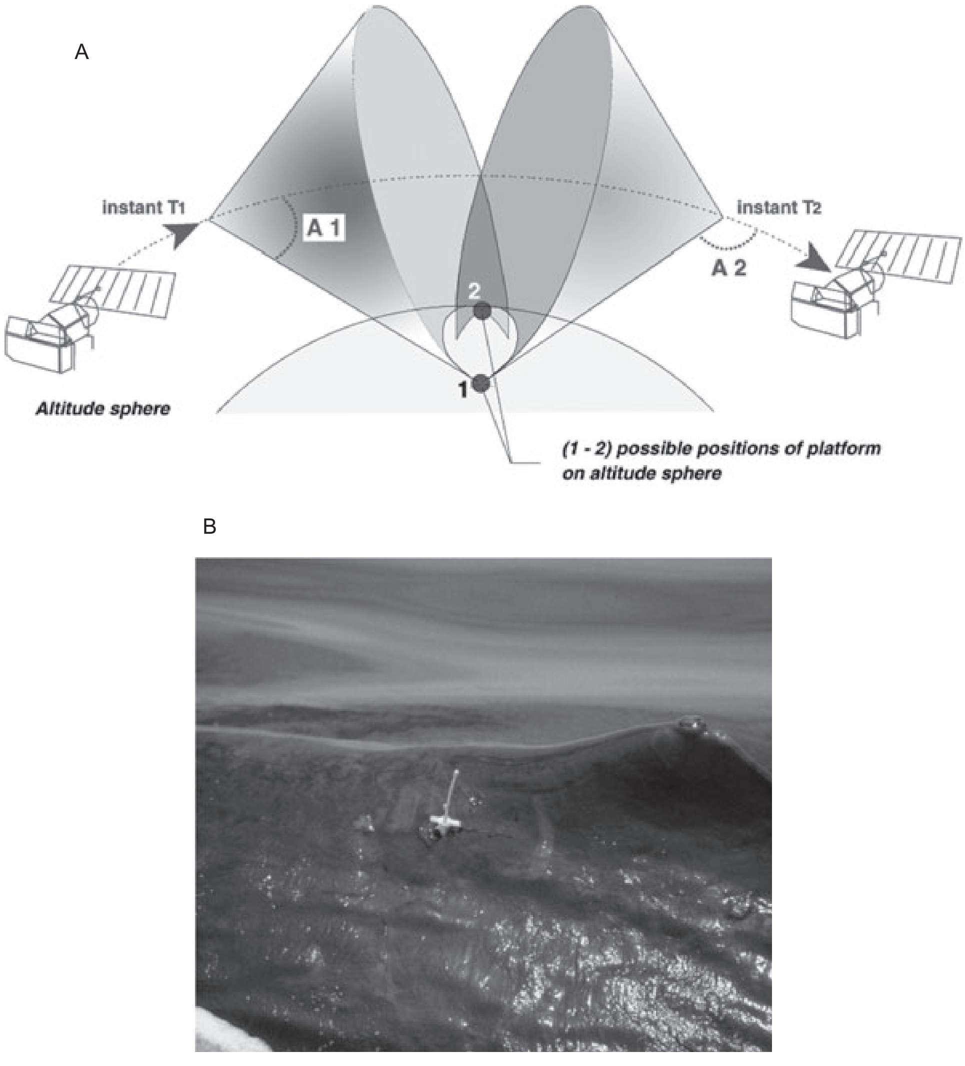

Figure 2 Top (A): Argos satellites and the solution pair generated by service Argos geolocation, as adapted from Austin et al. (2003) and Liaubet and Malardé (2003). Bottom (B): Photo of Argos satellite tag (~19 cm long) in a sperm whale immediately upon insertion (from Mate et al. 2007)

Arc Marine was originally written in unified modeling language (UML) using Microsoft Visio 2003 software, so the customization followed suit. New domains were added to the existing Domain layer, while new classes were added to additional layers in the UML. All resulting classes inherit behavior from ESRI Classes::Object. Implementation in UML addresses only the attributes of the new object classes and class extensions. Object behavior was implemented programmatically with further subclassing in VB.Net and Python 2.4 (further details and program code available in Lord-Castillo (2007)).

Within Visio 2003, the UML may then be exported to an XML Interchange file to be used as an ESRI geodatabase template. This XMI file must be used with computer-aided software design (CASE) tools in ArcCatalog (i.e. Schema Wizard) to create and populate a new personal file or ArcSDE geodatabase with associated classes. A test whale tracking geodatabase for the study was populated via data loading scripts developed by Lord-Castillo (2007).

A detailed diagram of the original Arc Marine schema is available in Wright et al. (2007) and on the web at http://dusk.geo.orst.edu/djl/arcgis (it takes a poster-sized document to depict the entire model in readable form). Of all the core Arc Marine objects, the entities most critical to this study are the InstantaneousPoint feature class, the Vehicle object class (Figure 3), and MarineEvent object class (Figure 4).

InstantaneousPoint (Figure 3) is a point feature representing a unique observation defined in time and space by geographic coordinates and a timestamp. LocationSeries, a subtype of InstantaneousPoint, allows for the spatial and temporal sequencing of a series of points moving through space. Again, each point represents a single unique observation. In this study, this subtype holds the critical geometry of satellite locations from the Argos telemetry messages. Animal tracks (as Track feature classes) are composited from the interpolation of movement between points. As such, tracks are calculated as on-the-fly features and not stored in the geodatabase. Typically, this class combined with the Series object class would define the movement of an animal.

An animal may be modeled in one of several different forms, taking on different classes depending on the animal's behavior compared to the object model. Most often, this representation is either as a MeasurementPoint representing a single observation of the animal in a survey or as a Series of LocationSeries points and associated Track representing the movement of a single animal (Figure 3). Yet in this study, point modeling was not deemed appropriate for representing a tagged animal. Complex multi-dimensional data are difficult to connect to single points, particularly when the data come from multiple sensors that are collecting between satellite fixes. Much of these data are associated with a time and an instrument rather than a specific location.

The Vehicle object class (Figure 3) is a less utilized class which generally stores information about a vehicle used during a survey run. Hence, the object class relates to both the MeasuringDevice object and Track feature class. Here, the important characteristic of the Vehicle object class is that it models a moving, instrument-carrying platform. Generally, this would be a survey vessel, but in this instance the moving, instrument-carrying platform is an animal.

MarineEvent (Figure 4) is meant to be used for linear referencing of time or distance mileposts (M-values) along linear features such as coastlines or ship tracks. As mentioned above, data collected from tag instruments are often associated with a timestamp rather than a location. Thus, MarineEvent in a time-stamp mode is a natural choice for subsequent dynamic segmentation of animal movement paths to create spatial locations for these time-stamped data.

In closely examining the Argos satellite telemetry data and the associated field operations at sea, it was determined that three more groups of objects (Animal, Telemetry, and Operations) needed to be added to Arc Marine for optimal user support. The Animal group develops and expands the base representation of animals as MarinePoints. The Telemetry group directly represents raw data and transformations of the data including location returns, data quality information, and data collected from tag sensors. An Operations group was developed in relation to the Cruise object class and Track feature class to provide auxiliary information about the events represented by those feature classes.

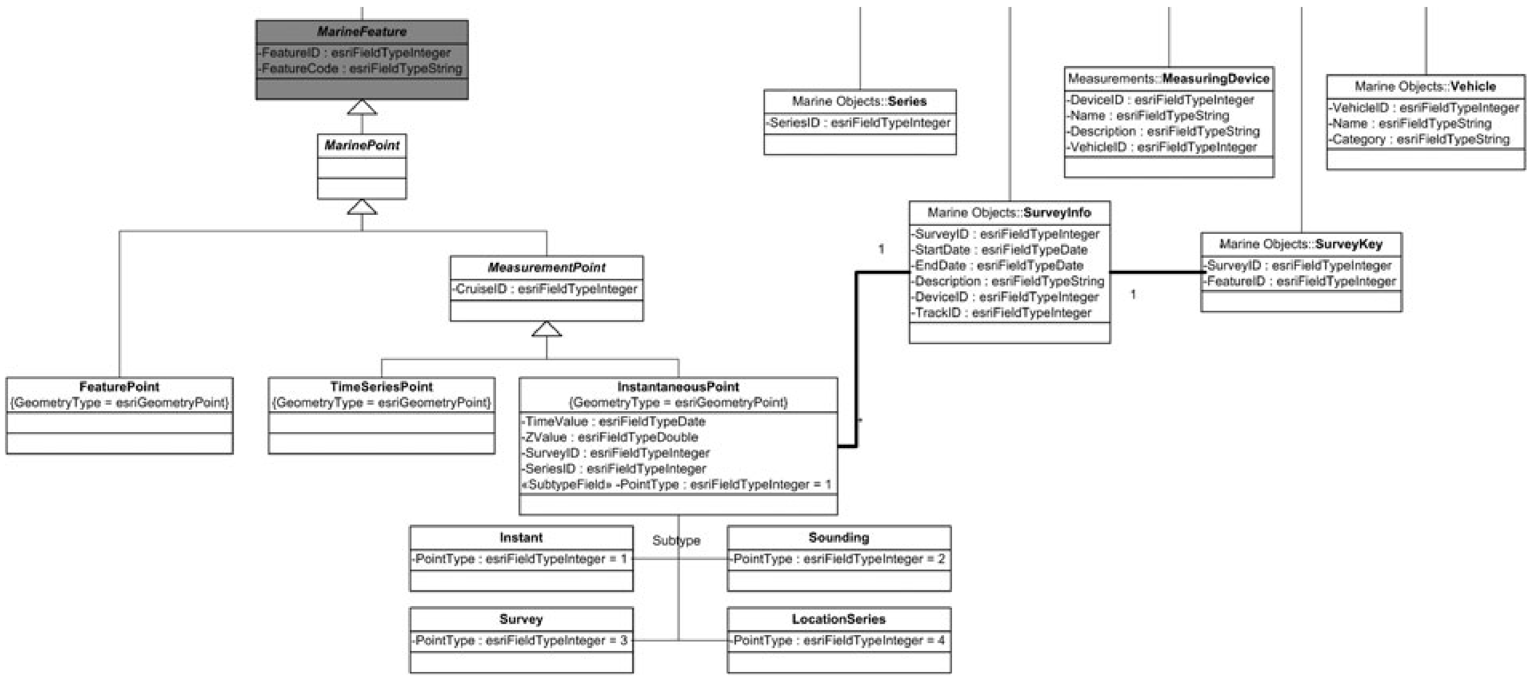

Figure 3 Fragment of the original Arc Marine UML, showing primary feature and object classes plus attributes for marine points, including InstantaneousPoints and Vehicle

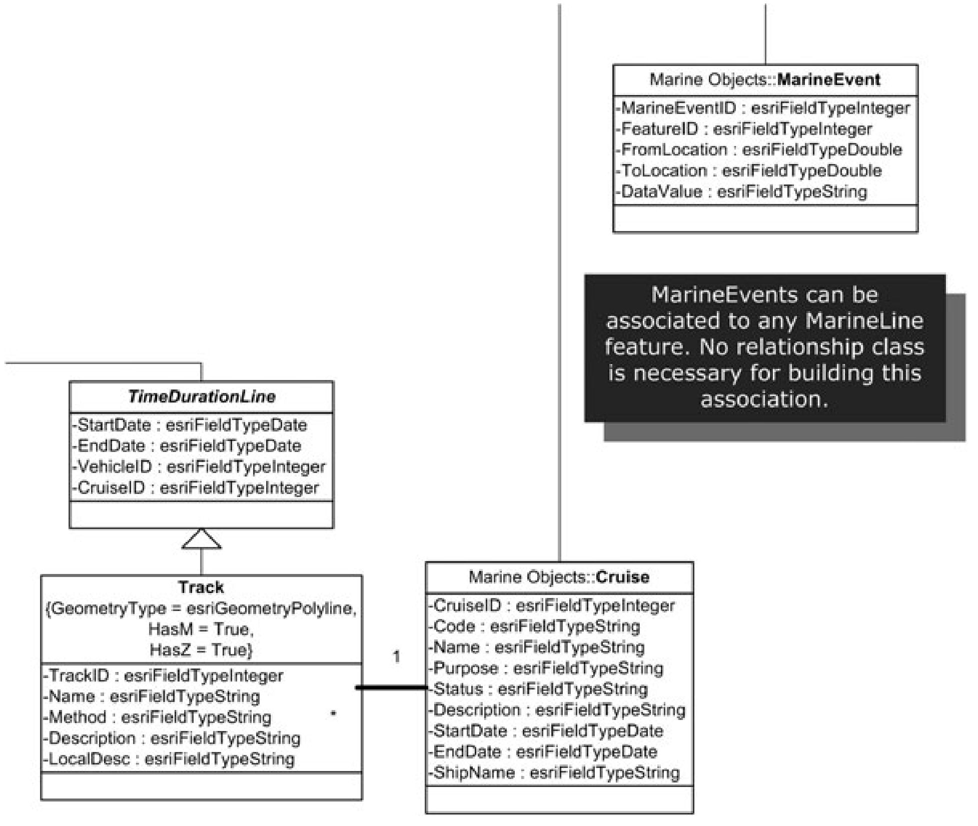

Figure 4 Fragment of the original Arc Marine UML, showing important feature and object classes plus attributes for marine lines, including Marine Event

Two new auxiliary entity-relationship groups were also devised: Tags (and hardware components) and Filtering (with object modeling of component functions to create a normalized filtering audit trail). The objects in the Tag group supply back-end information critical to the preprocessing of satellite telemetry returns, as well as the planning of hardware for future tag deployments. The objects in the Filtering group are designed to handle the front-end analytic task of selecting between Argos mirror points or indicating variable uncertainty, accuracy, or the exclusion of potential telemetry zingers in the data. These two groups were devised for developers to handle processing *outside* the Arc Marine data model. As such, we introduce them here only briefly and will not mention them further. The reader is referred to Lord-Castillo (2007) for a complete specification and associated computer code.

3 Results and Discussion

3.1 Animal

Figure 5 shows the new Animal class, which during the customization was created as a child class of the Vehicle parent class. In other cases, an animal might be better

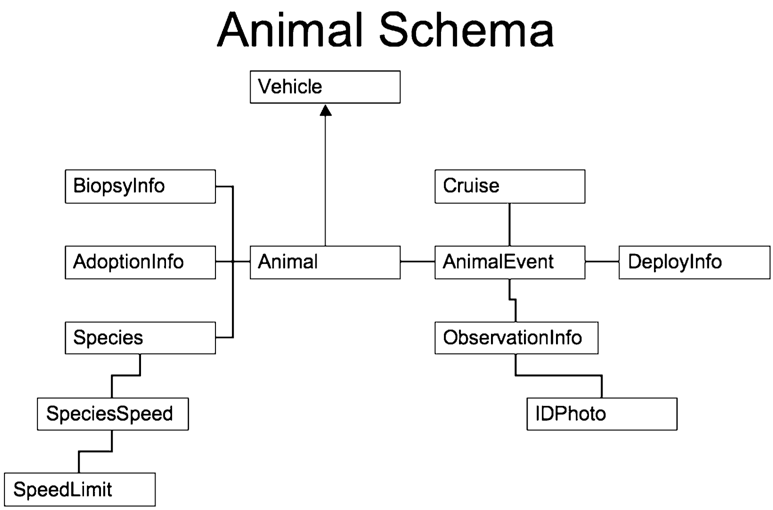

Figure 5 Generalized diagram of the new Animal group of object classes as customized for Arc Marine. Animal is a dimension table of BiopsyInfo, AdoptionInfo, and AnimalEvent, while Species is a higher-level hierarchy concept of Animal

represented as MeasurementPoint, possibly with related survey data. In this case though, the animal itself is not an observation but rather is its own instrument platform by way of the Argos tag, as the animal is carrying a collection of devices that measure a range of quantities including location, depth, temperature, salinity, and incident radiation, or in some cases surfacing rate or dive profile (Boehlert et al. 2001). This is analogous to a ship carrying a conductivity, temperature, and depth array with a GPS (though greatly miniaturized). Like a vehicle, an animal creates a Track (recorded by a measuring device) along which the attached MeasuringDevice array (the tag) records data.

As a specialized type of vehicle, an animal should also have a species, genotype, sex, social group, and length (the latter two based on the initial observation of the animal). Therefore, the Animal object class is also related to BiopsyInfo, AdoptionInfo, and Species object classes. The BiopsyInfo object class represents data on individual biopsies (and the related approach to the animal) and will eventually link to genetic information beyond the genotype as that part of the MMI program is developed. AdoptionInfo is an administration table related to fundraising that can be extremely helpful for such tasks as transmitting a tracking map for a specific whale to a donor who has adopted that whale. The Species object class not only avoids the redundancy of repeatedly storing genus, species and common name in the animal table, it also allows linking to species specific information such as maximum speed parameters (in the SpeedLimit table).

Since the animal is a specialized type of vehicle, it can carry MeasuringDevices (in this case, the satellite tags) that relate directly to MeasuredData. These measured data, though, are often derived from raw satellite telemetry data that can carry poor location accuracy or no location at all. Additionally, with Argos fixes there are two possible locations. The animal and the measured data are linked to these quality data and alternative locations through the AnimalEvent table that is the core fact table for much of the database (Figure 5).

3.2 Telemetry

AnimalEvent is the core relational table, or fact table, of the telemetry portion of the new database schema (Figure 6). It anchors the LocationSeries and Track feature classes to telemetry information stored in the extended database as well as tying together the animal, tag, and tag deployment (part of operations) in a star schema. AnimalEvent is similar to the MarineEvent class, but for time referencing rather than linear referencing, and for both object and feature classes. As noted by Wright et al. (2007, pp. 45–80), MarineEvent is intended to hold only a single value and cannot respond to the many parameters of an animal sighting. Similarly with the Telemetry group, a MarineEvent can tie a single value to a specified start and end location along a Track. AnimalEvent can relate complex parameters (through sub-dimension tables and a relationship to measured data) to start and stop points in time. Subsequent dynamic segmentation along a time-stamped Track fulfills the same geolocating purpose as MarineEvent.

AnimalEvent sub-dimension tables are context dependent (Figure 6). The table joined by AnimalEvent is dependent on the context of the event. Argos tag collection events link to ArgosInfo, tag deployments link to DeployInfo, field observations link to ObservationInfo (Table 1). Only the number of types of interactions with the animal limits the number of potential sub-dimensions. In particular, each new tag type links to a new sub-dimension table. As new tag types are added with different auxiliary attributes, new sub-dimension tables will be added.

These sub-dimension tables each carry a one-to-zero-or-one relationship with the AnimalEvent table; thus the joins from the animal to event information are context-specific (Figure 6). Though context-specific joins increase the complexity of query building, the cardinality of the sub-dimensions is significantly reduced (particularly for low frequency events such as DeployInfo). Note that attributes specifically needed for analysis are still stored in MeasuredData and spatial information is still stored in the geometry of feature classes. The AnimalEvent sub-dimension tables only provide access to auxiliary information related to a specific event.

3.3 Operations

The Operations group is divided into two areas, Cruise and Approaches (Figure 7). Cruise involves a small number of generic object classes to link field observations to the person making the observation. Approaches handle the specific operational situation of approaching an animal and deploying a tag.

Cruise is essentially a customization of the Arc Marine SurveyInfo object class tables. SurveyInfo links an InstantaneousPoint to a unique survey operation. This point may represent a sighting, photograph, deployment, telemetry location, or a wide variety of other features. When this point is linked to a survey though, that survey has a specific crew, identified by CrewKey, and specific crew members in that crew, identified by the Crew class object. Thus, a crew has crew members and carries out one unique survey. SurveyInfo is also a dimension of the ApproachEvent, a linking dimension table for the Approach object group.

ApproachEvent is a series of one-to-one related class objects which describe the specific instance of deploying a tag to an animal (Figure 8). This is an important special case, as this particular event ties together an Animal and MeasuringDevice to begin a Series. It is possible to completely omit the ApproachEvent and simply record which

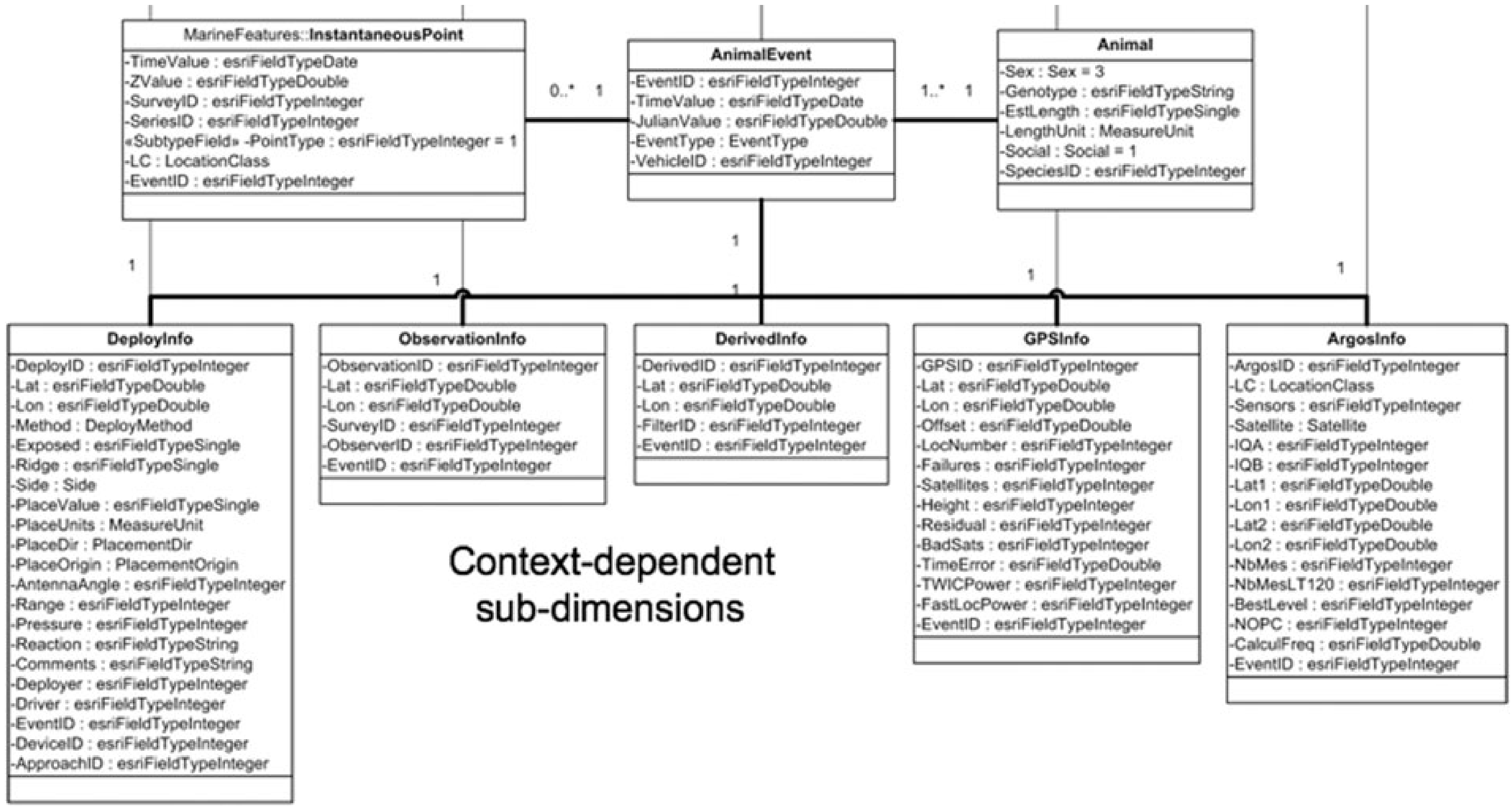

Figure 6 UML of AnimalEvent and context-dependent sub-dimensions in the new Telemetry group of object classes and associated attributes as customized for Arc Marine

Table 1 Context-dependent sub-dimensions of AnimalEvent

DeployInfo	Deployment of a measuring device onto an animal
ArgosInfo	Auxiliary information, Argos locations
ObservationInfo	Auxiliary information, field observations and photos
DerivedInfo	Auxiliary information, interpolated or derived location
GPSInfo	Auxiliary information, FastlockGPS locations

MeasuringDevice has been deployed to which Animal, but the significance of the event to marine animal tracking (particularly with the permitting requirements of marine mammal tracking) warrants specific inclusion in the database schema. DeployInfo is the linking table for this group. First, this table records a wide range of event parameters as an AnimalEvent sub-dimension. After all, deploying a tag to an animal is a rather monumental interaction in the study of that animal. This table also links to the specific tag deployed in MeasuringDevice and the ApproachEvent (which links to SurveyInfo and additional information about the specific approach). Thus, from Animal to AnimalEvent to DeployInfo to MeasuringDevice, the animal is initially linked to the tag instruments that it carries.

3.4 Test Loading of Data

To test the customized Arc Marine data model, a personal geodatabase was created from the customized schema and sample MMI Argos telemetry data from the Gulf of Mexico were loaded to populate the InstantaneousPoint feature class and LocationSeries subclass, as well as the Animal object class and MarineEvent object class (data available from Lord-Castillo 2007). Figure 9 shows point locations of a large sample of whales as mapped from the test geodatabase, featuring sperm whales tracked with Argos satellite telemetry tags from August 2001 to February 2006. Figure 9 also shows the result of using the Flag table with LocationSeries point (subclass of InstantaneousPoint) to build linearly interpolated tracking paths of selected individual whales using Bezier interpolation in ArcGIS (and the approach of Tremblay et al. (2006)), as well as two commonly-used third party tools in analyzing animal movement: Hawth's Analysis Tools for ArcGIS (Beyer 2004) and XTools (DeLaune 2000). Either the "Convert Locations to Path" in Hawth's Tools or the "Make One Polyline from Points" in XTools can be used to load marine animal paths into the Track feature class. The SeriesID must be manually linked to Animal (Vehicle child class) to provide a relationship between LocationSeries points and the associated Track.

In addition, a small suite of Python tools is in development to automate the process of downloading Argos tracking data from the sensor and loading it into a geodatabase resulting from the customized Arc Marine data model. These tools are embedded in the processing flow depicted in Figure 10 (program code available in Lord-Castillo (2007)), which shows an application framework in future development for the MMI. All pieces of the framework center on the animal tracking customization of Arc Marine, including snapshot LocationSeries point sets within a potential "data warehousing" (as explained in the next section). In the upper left, two Python scripts designed in a modular sequence carry out an automated daily download and updating sequence from the Argos satellite.

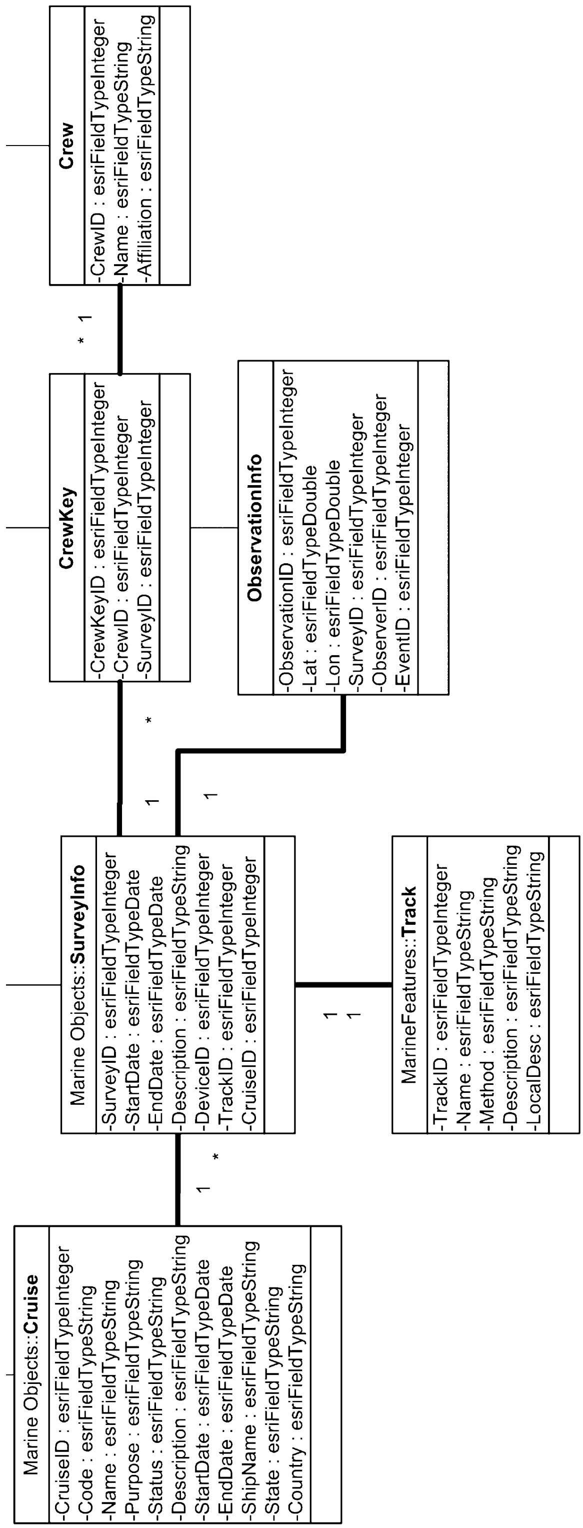

Figure 7 UML of the Cruise area within the new Operations group of object classes and associated attributes as customized for Arc Marine

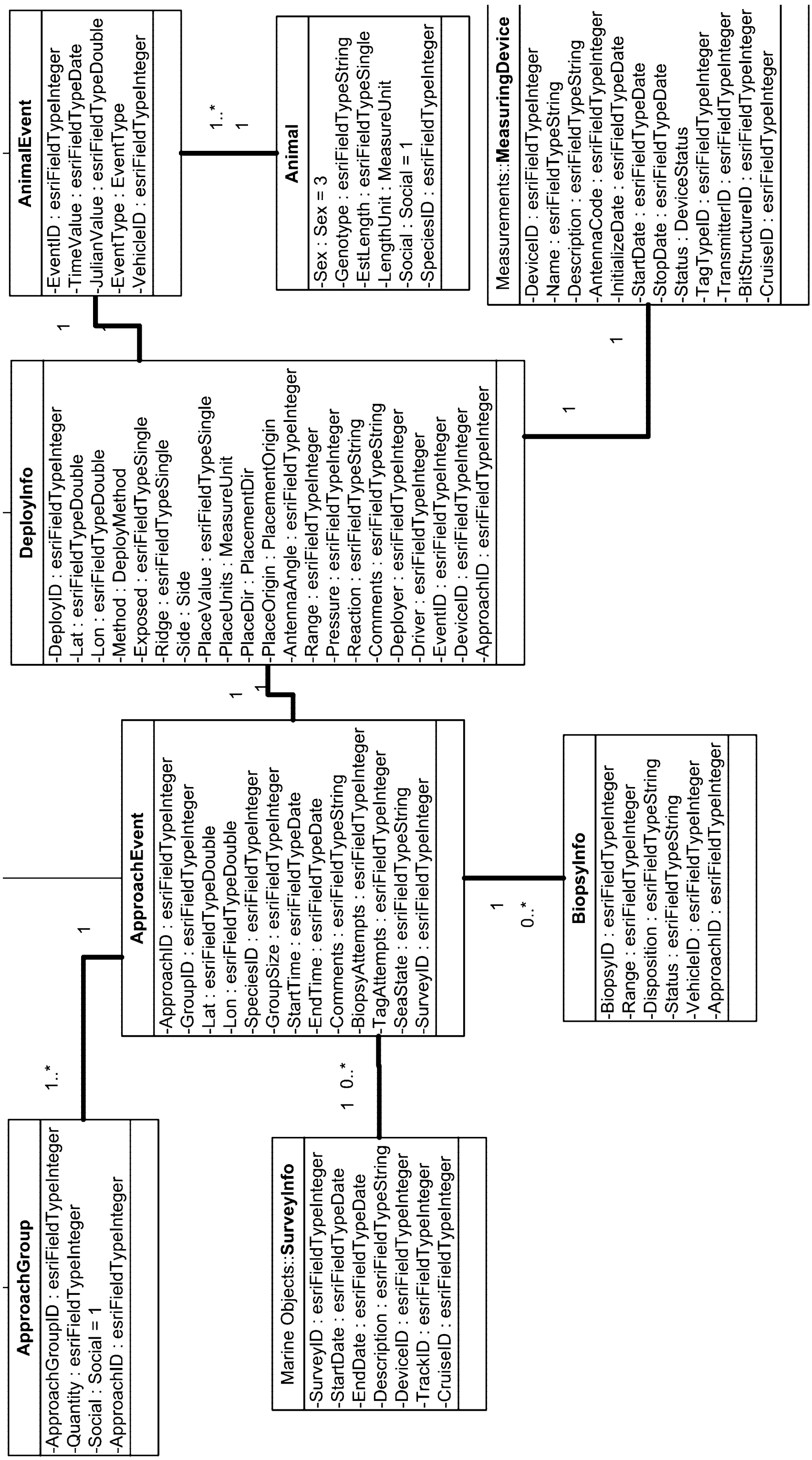

Figure 8 UML of the Approaches area within the new Operations group of object classes and associated attributes as customized for Arc Marine

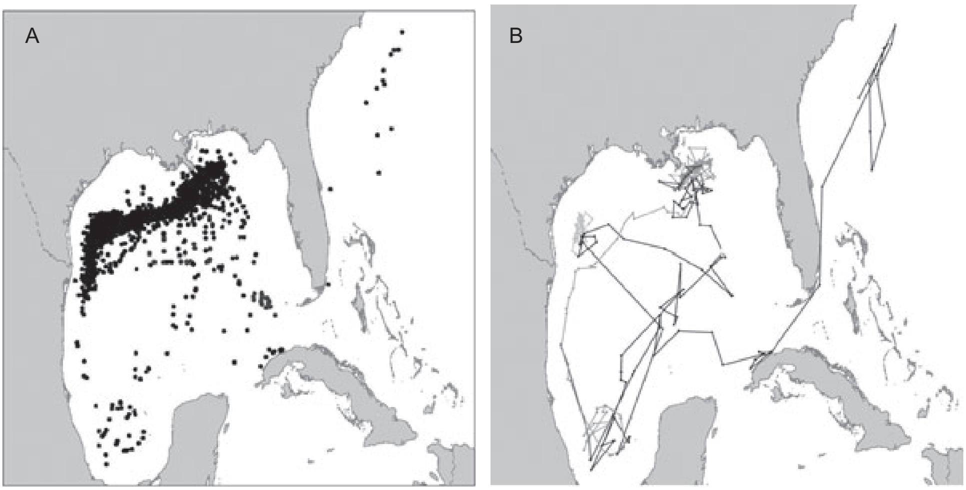

Figure 9 ArcMap screen shots of sample MMI Argos satellite telemetry data, August 2001-February 2006, loaded successfully into customized Arc Marine geodatabase. Left panel show point locations of a large sample of whales, while right panel shows interpolated tracks of selected individuals through time

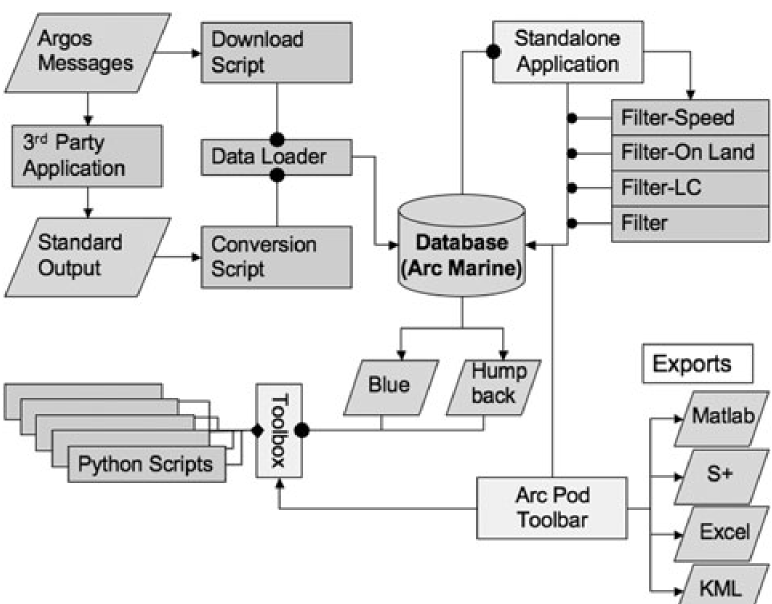

Figure 10 Conceptual diagram for a future application framework for MMI marine animal tracking data. The framework is centered on the customized Arc Marine database, download and data loader scripts, a standalone application with filter functions, an analytic toolbox with Python geoprocessing scripts, and an integrated ArcGIS extension

The first download script takes a series of connection parameters as an argument. That connection is used to download text results from Service Argos. The results are parsed to generate a Transmission container which holds, for each satellite transmission, PASS and DATA objects which respectively carry AnimalEvent (ArgosInfo) data and raw binary (MeasuredData). The second data loader script takes this Transmission container and iterates through the PASS and DATA objects to construct LocationSeries points from the PASS data and new MeasuredData records from the DATA object. MeasuredData records are then reconstructed from the new MeasuredData (as well as the appropriate AnimalEvents such as GPSInfo for binary encoded Fastlock GPS results) according to stored procedures based on tag type.

The standalone application depicted in the upper right of Figure 10, is a VB.NET application that implements protected access to feature class snapshots and non-spatial information tables outside of ArcGIS (e.g. the aforementioned Filtering group of objects). The lower left represents an analytic toolbox containing a series of Python scripts which rely on expected interfaces to Arc Marine objects. These tools can be shared with other researchers and applied to shared datasets as long as each data table implements the same standard interface, in this case an interface based on the InstantaneousPoint class. Finally, the lower right depicts the integrated application, or extension within ArcGIS to handle mapping tasks not handled effectively by the ArcGIS Engine or by geoprocessing scripts, and to provide menu-based access to common export formats.

4 An Arc Marine Data Warehouse Design Schema

Given the initial viability of the customized Arc Marine for satellite telemetry data, we propose here a possible "data warehouse" design schema for Arc Marine. Because Arc Marine is built upon atomic measures (scalar quantities, vector quantities, points, lines, and polygons) that are described by related dimensional tables (such as time, data parameters, tagged animal, or species) and concept hierarchies of different levels of generalization (e.g. tag < animal < social group < population < species), it should also be an appropriate target schema for the application of on-line analytical processing (OLAP) tools employed for data mining in the vast data warehouses of big business and genetics research (Chaudhuri and Umeshwar 1997). For a more extensive discussion of spatial data mining and the use of OLAP tools therein, see Miller and Han (2001, especially chapters 1, 3, and 4), and Han and Kamber (2006, chapters 3, 4, and 10), and Miller (2007). As a data warehouse, Arc Marine could be used for the spatial data mining of satellite telemetry tracking datasets, as well as integrated historical archives of other marine data types.

There are three common data warehouse design schemas: star, snowflake, and fact constellation (Han and Kamber 2006). A *star schema* is the most common form and is characterized by a normalized central fact table containing the atomic measure and dimension table keys, and a set of denormalized dimension tables. The structure is termed a star because the schema graph is typically displayed with the dimension tables in a radial pattern around the fact table. The *snowflake schema* differs from the star schema in that the dimension tables are normalized into further sub-dimension tables. This normalization reduces redundancy, saves space, and is easier to maintain. The tradeoff though is a greater number of joins in query execution. When the dimension tables are small relative to the fact table (the most common case), the advantages of normalization are minimal. The *fact constellation* is a schema containing multiple fact tables sharing dimension tables. As an example, a data warehouse containing historical fact tables for shipping, inventory, and sales would share location and product dimension tables between the three fact tables. This schema can be thought of as a collection of star schema, hence the terminology fact constellation.

For this study, LocationSeries Point and MeasuredData are the central fact tables, with LocationSeries point serving as part of the concept hierarchy for the Measurement dimension of Measured Data. MeasuredData represents the atomic measure within the schema, while features, in this case LocationSeries Point serve as part of the concept hierarchy for Measurement. LocationSeries (as well as other features) is also an atomic measure of its own fact tree (sharing a fact constellation with other feature classes) when it is used for purely analyzing spatial distribution or movement.

For example, surfacings are a quantity commonly measured by satellite telemetry tags on whales. In this simple example, surfacings are just a count of the number of times the animal reaches the surface (there are other ways to measure this metric). A data view of animal surfacings focuses on measured data from the tags and may or may not have a spatial component. When a spatial component is used, it merely represents an aggregating spatial area for a count of surfacings. That count is still contained within the measured data themselves, and would rely on a data view based on the MeasuredData star. Meanwhile, a kernel density or home range analysis relies only on animal locations and no elements of measured data. There may be dimensions to the animal locations (animal, species, location quality), but the atomic measure used in the analytic calculations is the point feature. Hence, the fact table for kernel density would be the LocationSeries Point table and the data view would be based on that table's star schema.

4.1 The MeasuredData Star

MeasuredData is the fact table of star schema represented by a three dimensional data cube of Measurement, MeasuringDevice, and Parameter (Figure 11). These dimensions alone do not present interesting levels of aggregation, but the concept hierarchies for MeasuringDevice and Measurement introduce significant analytical aggregations, while Parameter defines data of common types. The MeasuringDevice hierarchy roles up from MeasuringDevice to Vehicle to the feature class Track. Meanwhile the Measurement concept hierarchy rolls up to multiple parallels. One of these parallels is the TimeSeries object class which subsequently rolls up to the MarineFeature classes. The remainder of these parallel hierarchies are the feature classes themselves which embody spatial and temporal quantities as well as rolling up to higher aggregations such as surveys, cruises, and series. In the context of the MMI customization, a fourth dimension is added in the form of AnimalEvent (Figure 12). The MeasuringDevice concept hierarchy develops a more significant aggregation by substituting Animal (and hence Species and higher levels of Animal) for the Vehicle concept level. AnimalEvent not only allows another route to aggregation by Animal or feature classes, it also opens up a route to aggregation by the wide array of context-dependent sub-dimensions. Denormalization relative to specific sub-dimensions can create multiple sub-cubes by tag type that will allow the relation of MeasuredData all the way back to raw data messages.

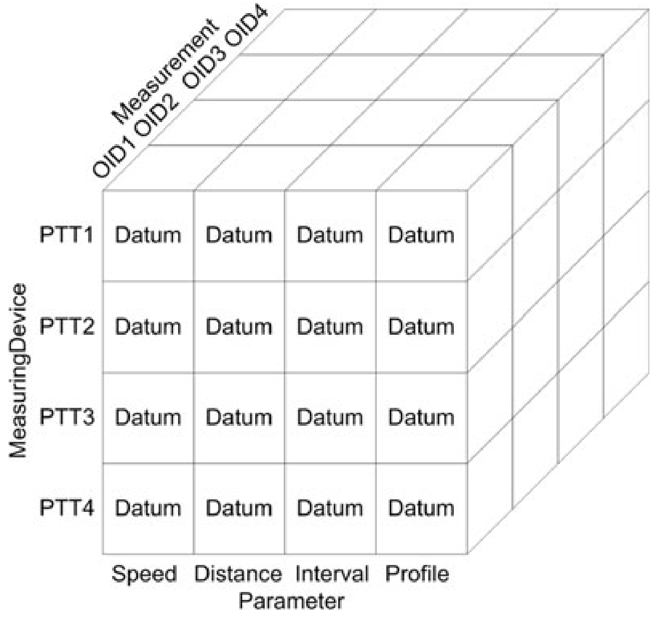

Figure 11 A MeasuredData data cube model for the core Arc Marine data model

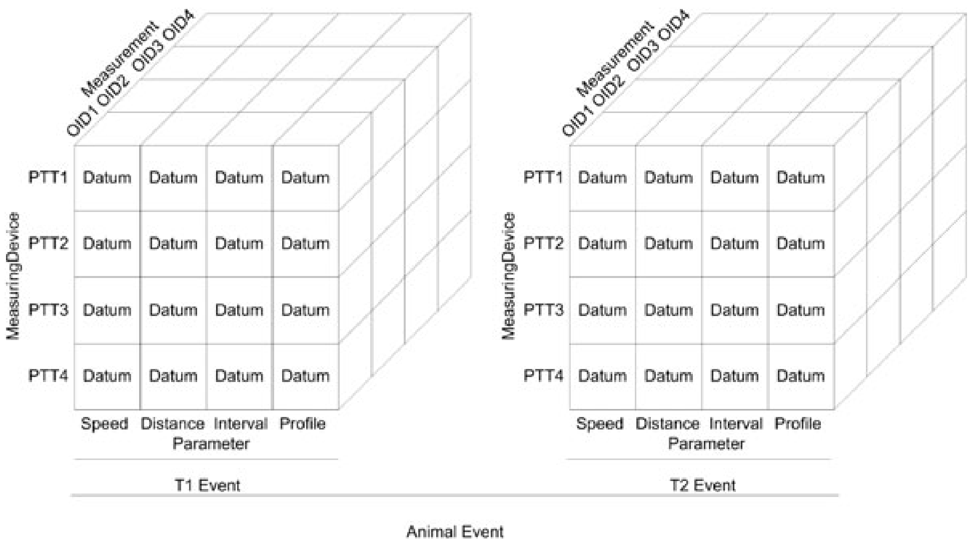

Figure 12 MeasuredData data cube in the MMI customization. Note that this is a four-dimensional cube, with the fourth dimension, AnimalEvent, represented as multiple cubes

One extremely important aspect of the spatial data warehouse is the ability to aggregate spatially. Any concept hierarchy which can roll up to a feature class can further roll up to spatial generalizations and even take on additional dimensions from spatial joins with environmental rasters as defined in spatial data mining techniques (see Miller 2007 for an extensive discussion of spatial OLAP operations).

4.2 The LocationSeries Point Star

This discussion of the MMI customization of Arc Marine as a spatial data warehouse closes with an exploration of the spatially oriented LocationSeries point star. In the MMI customization, LocationSeries point carries only two object class dimensions, Animal Event and Animal (as linked through Series). The implications of each of these classes in a concept hierarchy have been discussed above. It should be particularly noted that the multidimensional cube of this fact star can be drilled down, for example from Animal to MeasuringDevice to MeasuredData. Yet, LocationSeries point, as a spatial feature class, carries an additional spatial dimension as defined by its spatial geometry. The concept hierarchy of geographic space has potential levels limited only by the grain and extent of the dataset. As mentioned above, this spatial dimension also contains spatial joins to other spatial datasets introduced into the data warehouse. As a result, the potential for data mining expands out to any form of marine data linked to Arc Marine, and with it, the potential to more deeply explore the central question of the MMI program, what is the relationship between physical and biological process and the distribution and movement of endangered whale species.

5 Conclusions

In seeking the best customization of the generic Arc Marine data model to enhance the key advantages of satellite telemetry for marine animal tracking, the study demonstrates the flexibility of the LocationSeries subtype of InstantaneousPoint in handling a wide array of data types and tag types. As import/export tools develop, the LocationSeries feature class should evolve into a system for transfer between geodatabases, as well as an archival form for data warehouses.

The customization has revealed two key concepts that can help guide the future development of Arc Marine for animal tracking and as an enterprise on-line analytical processing structure. First, intended multidimensionality of the model can indeed be extended effectively with additional dimensions (such as time through AnimalEvent), and broader levels of hierarchy concepts, from time-stamped data acquisition events to aggregated spatial regions. This multidimensional view of marine data opens a pathway to the implementation of higher level analytical tools including data warehouses, OLAP techniques, and spatial data mining.

Second, Arc Marine creates an expandable platform to drive community application development. By defining a tracking community framework with the MMI customization, researchers and programmers from different projects can develop compatible tools and share compatible datasets. This will speed data extraction from online repositories such as OBIS-SEAMAP. The additions of cross-platform Python scripting and database abstraction will help separate physical implementation decisions from processing tool choices. Tools developed for the back-end framework will facilitate data loading and ease the transition to Arc Marine. Tools developed for the front-end will open up more powerful analytical techniques and make the adoption of Arc Marine more attractive across the marine animal tracking community.

New definitions of programmatic and data management frameworks, from loading to warehousing to analysis, will provide the structure for a high speed and accurate automated workflow from satellite download to deep end user analysis. The encompassing object definitions of the core Arc Marine classes provides a standardizing framework, pointing towards common paths of import and export between research initiatives that will be able to publish and subsequently share faster than ever before. Finally, a fully-developed multidimensional framework allows for the development of analytical tools across a limitless range of environmental variables and into the narrowest and broadest scales of the spatial concept hierarchy crossed by tracks of these critical species.

Acknowledgments

The support of NOAA Grant # NA04NOS4730181 (NOAA Coastal Service Center Education and Research Opportunities in Applying GIS and Remote Sensing in Coastal Resource Management) to Dawn Wright is gratefully acknowledged, as it supported the M.S. thesis of Brett Lord-Castillo. The reviews of an anonymous referee and *Transactions in GIS* editor John Wilson significantly improved the manuscript.

References

Austin D, McMillan, J I, and Bowen W D 2003 A three-stage algorithm for filtering erroneous Argos satellite locations. *Marine Mammal Science* 19: 371–83

Beyer H L 2004 Hawth's Analysis Tools for ArcGIS. WWW document, http://www.spatialecology.com/htools

Block B A 2005 Physiological ecology in the 21st century: Advancements in biologging science. *Integrative and Comparative Biology* 45: 305–20

Block B, Bograd S, Costa D, and Stevens J 2000 Tagging of Pacific Predators. WWW document, http://topp.org

Block B A, Dewar H, Farwell C, and Prince E D 1998 A new satellite technology for tracking the movements of Atlantic bluefin tuna. *Proceedings of the National Academy of Sciences USA* 95: 9384–9

Block B A, Teo S L H, Walli A, Boustany A, Stokesbury M J W, Farwell C J, Weng K C, Dewar H, and Williams T D 2005 Electronic tagging and population structure of Atlantic bluefin tuna. *Nature* 434: 1121–7

Boehlert G W, Costa D P, Crocker D E, Green P, O'Brien T, Levitus S, and Le Boeuf B J 2001 Autonomous pinnipeds environmental samplers: Using instrumented animals as oceanographic data collectors. *Journal of Atmospheric and Oceanic Technology* 18: 1882–93

Chaudhuri S and Umeshwar D 1997 An overview of data warehousing and OLAP technology. *ACM SIGMOD Record* 26: 65–74

DeLaune M G 2000 XTools ArcView Extension Oregon Department of Forestry. WWW document http://www.odf.state.or.us/divisions/management/State_forests/XTools.asp#xtools

Duke University Marine Laboratory 2004 Duke North Atlantic Harbor Porpoise Tracking OBIS-SEAMAP. WWW document, http://seamapenvdukeedu/datasets/detail_test/83

Halpin P, Best B, Fujioka E, and Coyne M 2004 Marine animal analysis applications. In *Proceedings of the Twenty-fourth Annual ESRI International User Conference*, San Diego California

Han J and Kamber M 2006 *Data Mining: Concepts and Techniques*. San Francisco, CA, Elsevier

Lagerquist B A, Stafford K M, and Mate B R 2000 Dive characteristics of satellite-monitored blue whales (*Balaenoptera musculus*) off the Central California Coast. *Marine Mammal Science* 16: 375–91

Le Boeuf B J, Crocker D E , Grayson J, Gedamke J, Webb P M, Blackwell S B, and Costa D P 2000 Respiration and heart rate at the surface between dives in Northern Elephant Seals. *Journal of Experimental Biology* 203: 3265–74

Liaubet R and Malardé J 2003 Argos location calculation. In *Proceedings of the Argos Animal Tracking Symposium*, Annapolis, Maryland
Lord-Castillo B K 2007 Arc Marine as a Spatial Data Infrastructure: A Marine Data Model Case Study in Whale Tracking by Satellite Altimetry. Unpublished M.S. Thesis, Oregon State University (available with program code at http://marinecoastalgis.net/brett07)
Mate B R 1989. Watching habits and habitats from Earth satellites. *Oceanus*. 32: 14–8
Mate B, Mesecar R, and Lagerquist B, 2007 The evolution of satellite-monitored radio tags for large whales: One laboratory's experience. *Deep Sea Research II: Topical Studies in Oceanography* 54: 224–47
Mate B R, Stafford K M, Nawojchik R, and Dunn J L 1994 Movements and dive behavior of a satellite-monitored Atlantic white-sided dolphin (*Lagenorhynchus acutus*) in the Gulf of Maine. *Marine Mammal Science* 10: 116–21
Miller H J 2007 Geographic data mining and knowledge discovery. In Wilson J P and Fotheringham A S (eds) *Handbook of Geographic Information Science*. Oxford, Blackwell: 352–66
Miller H J and Han J 2001 *Geographic Data Mining and Knowledge Discovery*. London, Taylor and Francis
Read A J, Halpin P N, Crowder L B, Best B D, and Fujioka E (eds) 2006 OBIS-SEAMAP (Spatial Ecological Analysis of Megavertebrate Populations): Mapping Marine Mammals Birds and Turtles. WWW document, http://seamapenvdukeedu
Sherman L 2006 Tracking the great whales. *Terra* 1(2): 2–8
Tremblay Y, Shaffer S A, Fowler S L, Kuhn C E, McDonald B I, Weise M J, Bost C, Weimerskirch H, Crocker D E, Goebel M E, and Costa D P 2006 Interpolation of animal tracking data in a fluid environment. *Journal of Experimental Biology* 209: 128–40
US House 110th Congress, 1st Session 2007a HR 1006: Marine Mammal Assistance. WWW document, http://www.gpoaccess.gov/bills/index.html
US House 110th Congress, 1st Session 2007b HR 2327: To Amend the Marine Mammal Protection Act of 1972 to Strengthen Polar Bear Conservation Efforts and for Other Purposes. WWW document, http://www.gpoaccess.gov/bills/index.html
Wright D J, Blongewicz M J, Halpin P N, and Breman J 2007 *Arc Marine: GIS for a Blue Planet*. Redlands, CA, ESRI Press