

Tuesday, July 14th, 2009

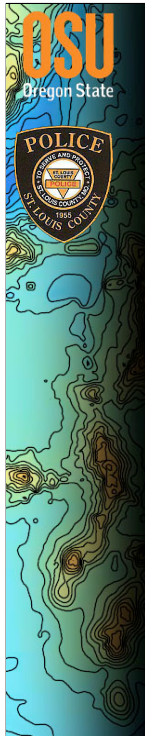
Advances in GIScience: Research Session 2

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

Brett K. Lord-Castillo*†, Dawn J. Wright,*
Bruce R. Mate‡ and Tomas Follett‡

**Department of Geosciences, Oregon State University, Corvallis, OR
†now at St Louis County (MO) Police, Office of Emergency Management
‡Marine Mammal Institute, Oregon State University*

Title Slide: 0:00/0:00

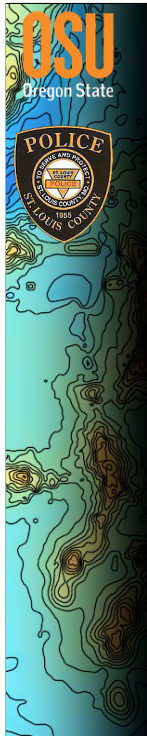


Tuesday, July 14th, 2009

Advances in GIScience: Research Session 2

The Case Study

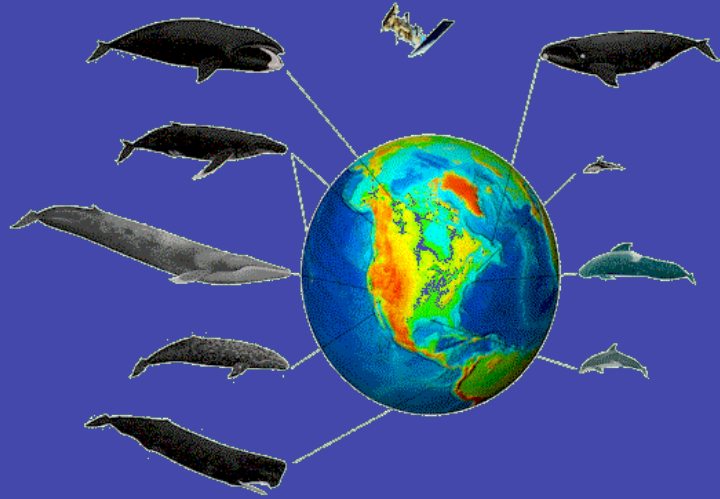
Title: 0:05/0:55



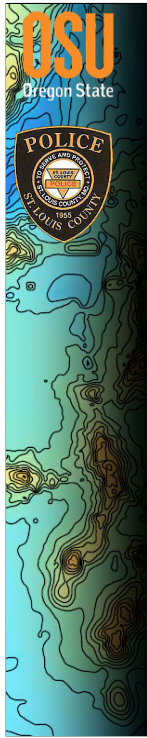
Tuesday, July 14th, 2009

Advances in GIScience: Research Session 2

Oregon State University- Marine Mammal Institute



MMI Intro 1: 0:30/1:25

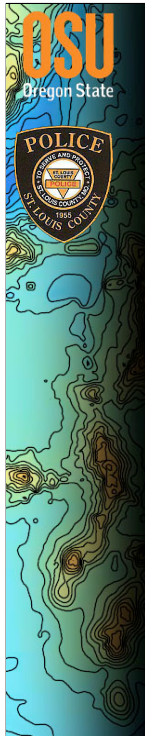


See also: Mate et al. 2007

MMI Intro 2: 0:30/1:55



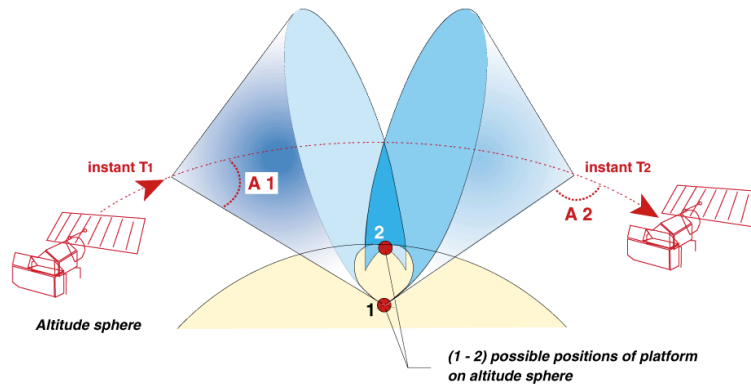
MMI Intro 3: 0:30/2:25



Tuesday, July 14th, 2009

Advances in GIScience: Research Session 2

Argos Telemetry

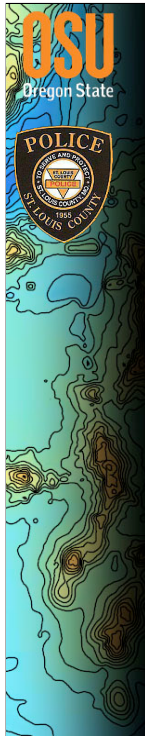


Adapted from Liaubet and Malardé, 2003

Austin et al 2003

MMI Intro 4: 0:30/2:55

Argos: Just mention, don't discuss



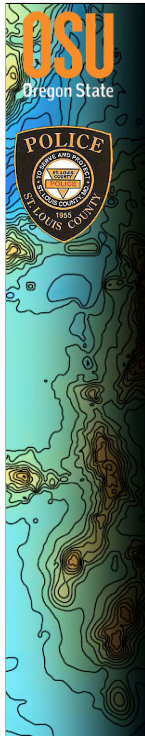
Tuesday, July 14th, 2009

Advances in GIScience: Research Session 2

Primary Question

How can the Arc Marine Data Model be customized to best meet the research objectives of the OSU MMI and the marine animal tracking community?

Question 1: 0:20/3:15



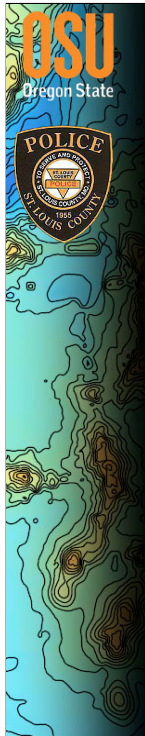
Tuesday, July 14th, 2009

Advances in GIScience: Research Session 2

Secondary Question

How can a geographic information system implementation enhance the key advantages of satellite telemetry?

Question 2: 0:15/3:30



Tuesday, July 14th, 2009

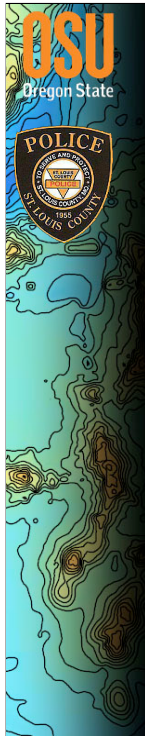
Advances in GIScience: Research Session 2

Key Advantages of Satellite Telemetry

- Continuous Coverage
- Autonomous Profiling
- Data relationships
- Timeliness

Lagerquist et al 2000; Block 2005;
Boehlert et al 2001.

Key Advantages: 1:00/4:30



Tuesday, July 14th, 2009

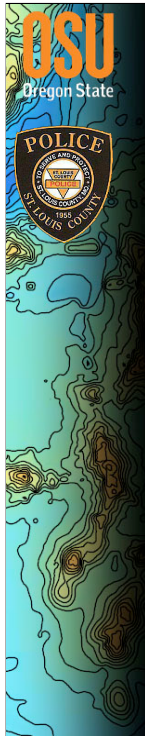
Advances in GIScience: Research Session 2

Area of Emphasis

Implementing Arc Marine in on-line analytic processing and data warehousing to enhance data exchange and provide access to high level data mining techniques

AoE:0:20/4:50

There is an additional implementation area of emphasis that emerged in this study, exploring how we can use Arc Marine to bring our research results into an on-line analytic processing and data warehousing environment so we can improve data exchange between researchers and utilize spatial data mining techniques.

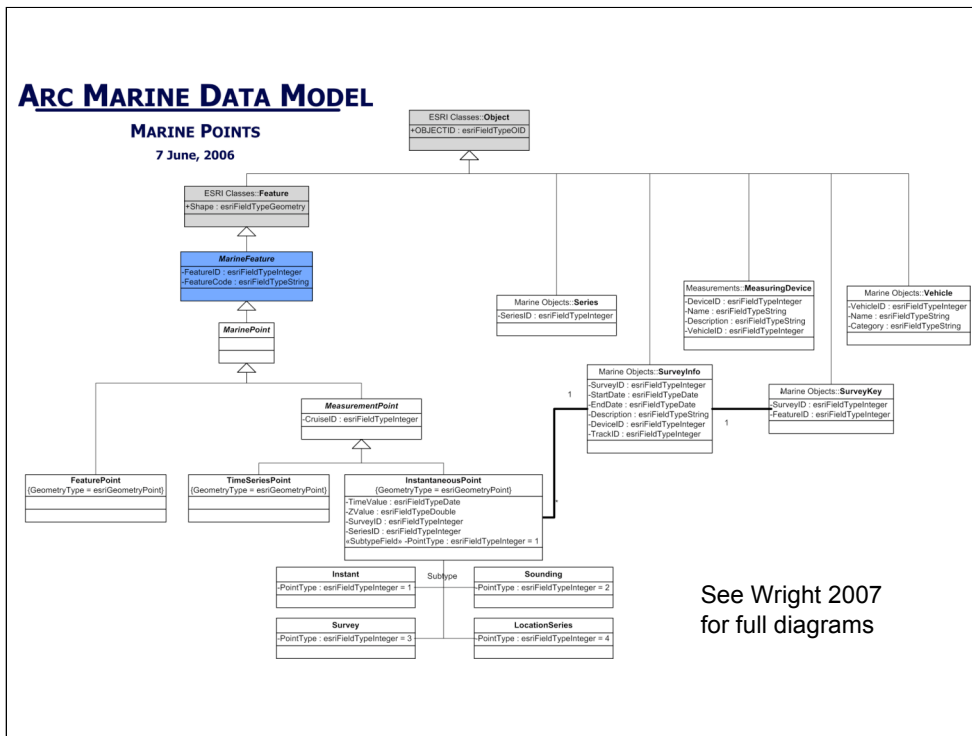


Tuesday, July 14th, 2009

Advances in GIScience: Research Session 2

Methods

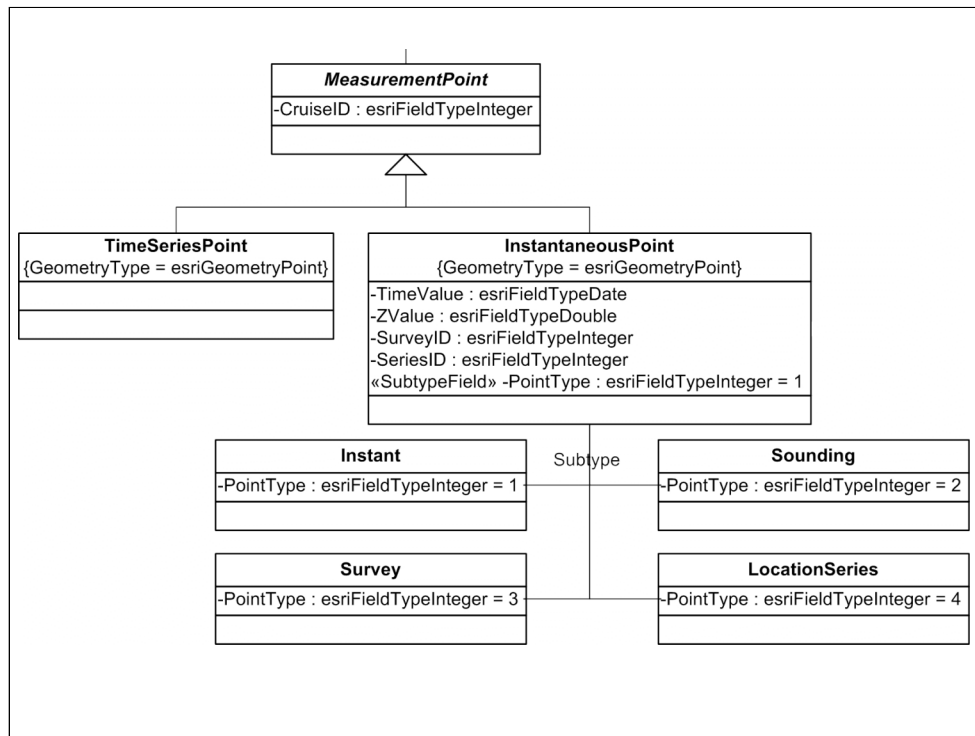
Title: 0:05/4:55



Quick Tour of Arc Marine 1: 0:20/5:15

First, we will dive into the sections of the Arc Marine Data Model used in this customization. Dawn Wright will be doing an Arc Marine Technical Workshop at noon today right after this session in Room 6C if you want to know more about the core Arc Marine Model.

We'll begin with the InstantaneousPoint feature class

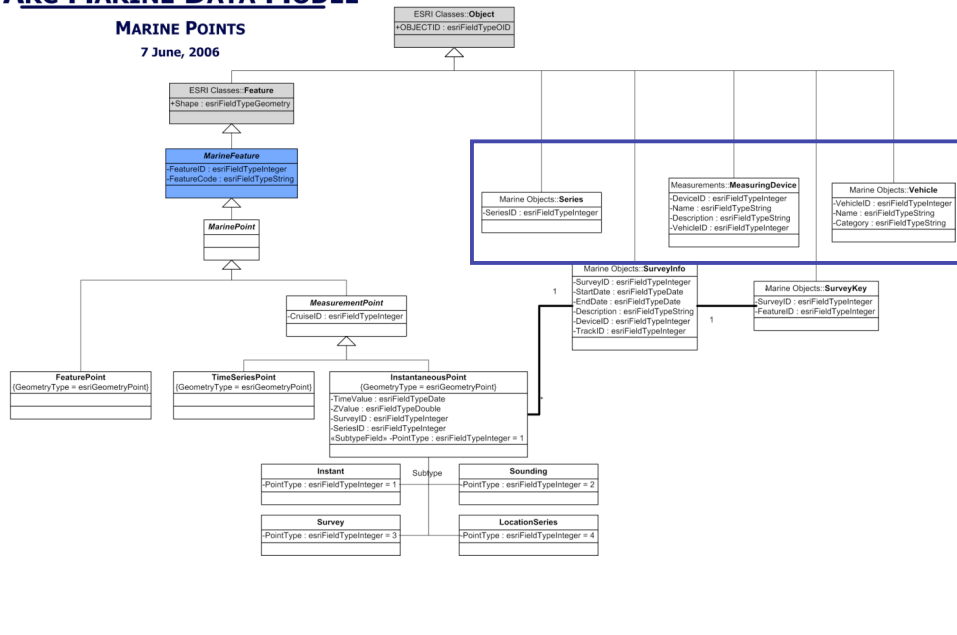


Tour 2: 0:10/5:25

One feature class serving multiple roles based on a domain-defined subtype. Throughout our discussion, we will be dealing with the LocationSeries subtype, a collection of locations ordered by time-stamp.

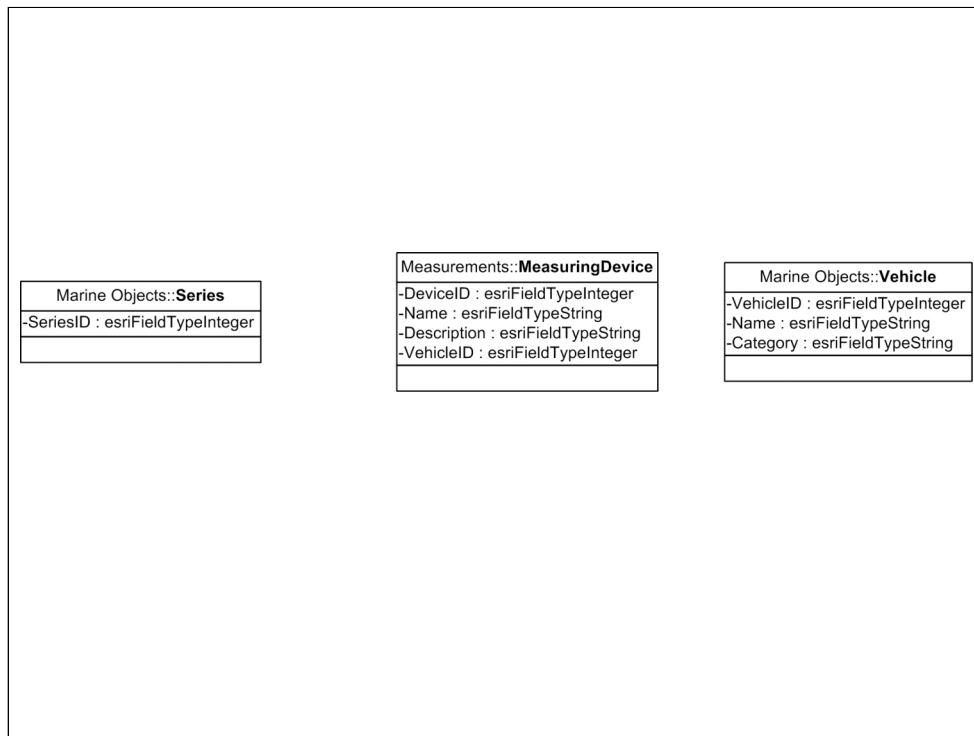
ARC MARINE DATA MODEL

MARINE POINTS
7 June, 2006



Tour 3: 0:10/5:35

To create and describe this collection of points, we'll customize these three core classes....

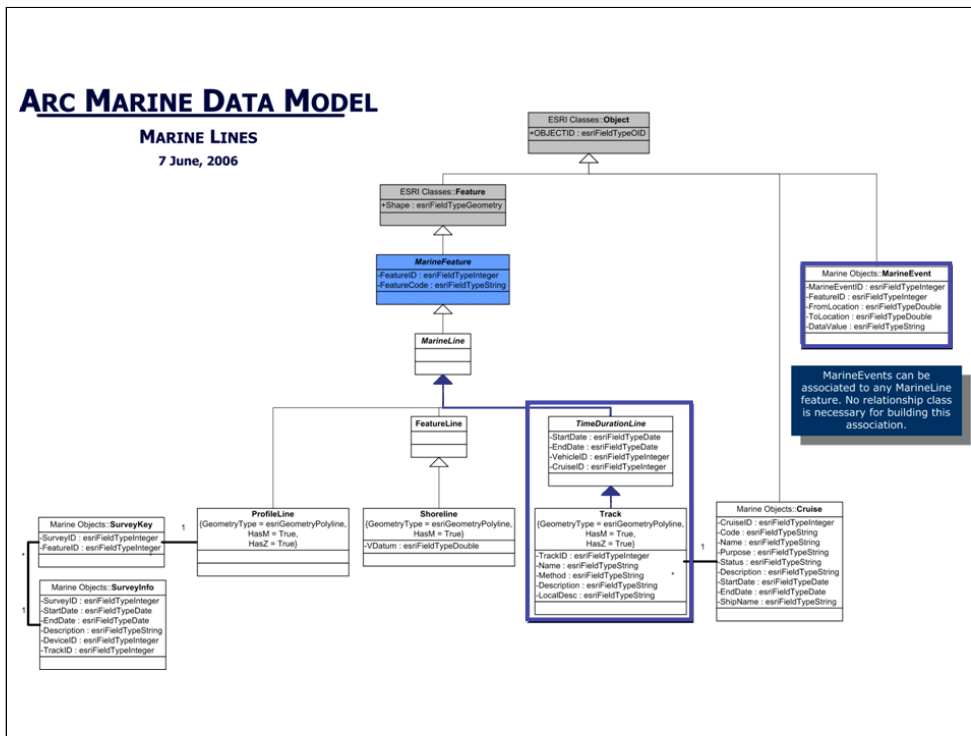


Tour 4: 0:15/5:50

Series, a container object for the collection. MeasuringDevice acts as a link to data and data parameters. Vehicle a platform for MeasuringDevices, and will carry a special role in this customization.

ARC MARINE DATA MODEL

MARINE LINES
7 June, 2006



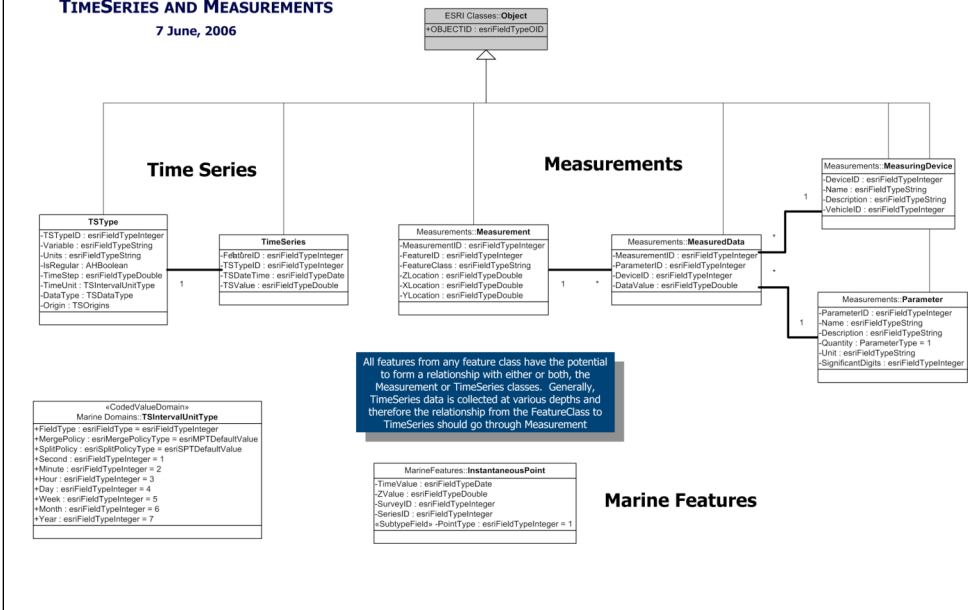
Tour 5: 0:15/6:05

A Vehicle, in addition to being associated with the measuring devices it carries, can be associated with a specific line feature type, a Track, which in turn can be dynamically segmented by m-value delineated MarineEvents, mileposts in space along the track line.

ARC MARINE DATA MODEL

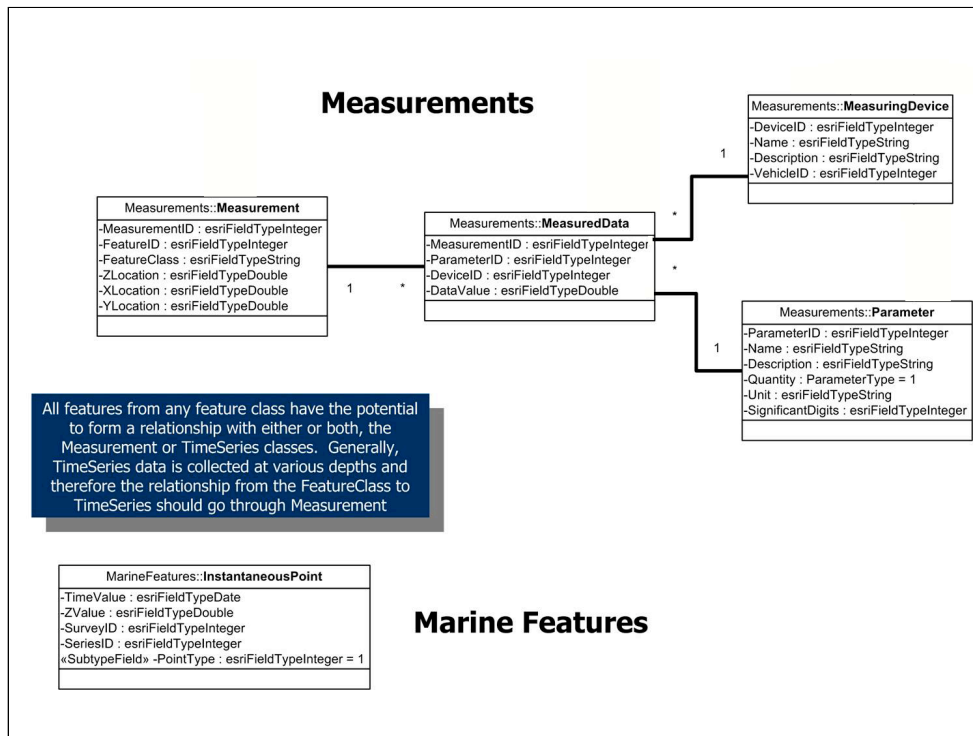
TIMESERIES AND MEASUREMENTS

7 June, 2006



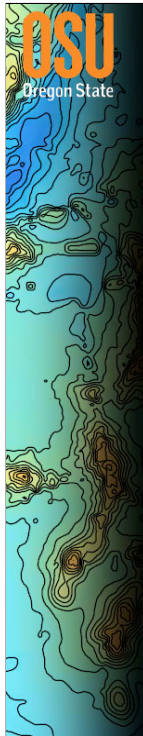
Tour 6: 0:10/6:15

The last Arc Marine core model pieces to our puzzle are the Measurements group classes.



Tour 7: 0:20/6:35

The Measurements group takes an atomic measure as a MeasuredData object, and provides it with measurement Parameters, provides spatial context through the Measurement class's linkage to feature classes, such as InstantaneousPoint, and finally links to instrumentation and instrumentation platforms (or Vehicles) through MeasuringDevice.



Tuesday, July 14th, 2009

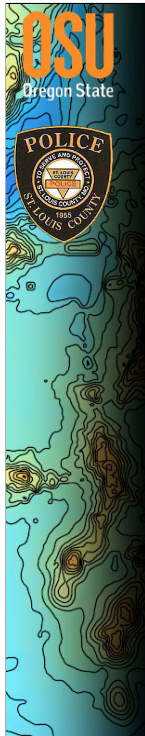
Advances in GIScience: Research Session 2

Programming Languages and Software

- Microsoft Visio 2003
- UML to XMI Export Utility
- ArcGIS 9.2
- SQL Server 2005
- MS Access
- VB.NET
- Python 2.4

Programming: 0:45/7:20

I also want to quickly address the software tools involved in this customization. Visio and the UML to XMI export utility are used for designing and exporting the database schema. The database itself was created using ArcGIS 9.2 schema tools and SQL Server 2005. Microsoft Access, VB.NET and Python all provided convenient structures for data loading and tool development, but a wide range of tool development environments can be supported. While the greyed out components can be swapped for upgraded or alternate tools, Visio 2007 is not compatible with the UML to XMI export tool, and the alternative to Visio, Rational Rose, may be cost prohibitive for most projects. So those two components, Microsoft Visio 2003 and the UML to XMI Export Utility are more difficult to swap out.



Tuesday, July 14th, 2009

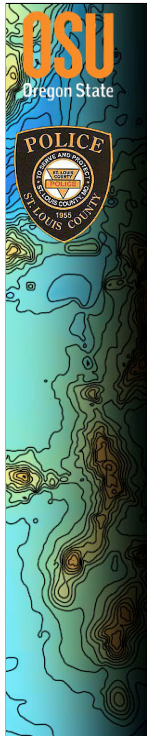
Advances in GIScience: Research Session 2

Customization Key Points

- No new spatial objects
- New spatial entities modeled as object classes related to feature classes
- All core classes retained
- Object behavior implemented programmatically (physical design dependent)

Key Points: 0:40/8:00

To keep customization in the logical and physical design phases of data model design, we adhered to a few key points. Use the exist core spatial objects to create spatial representations; new spatial entities are built by relating attributed object classes to feature classes. Even if not used, retain all the core classes in logical design to allow for integration with other Arc Marine templated projects. Removing empty classes can take place in physical implementation; similarly the programmatic implementation of object behavior should wait for physical design, but software design can be carried out knowing that attributes present in the logical design will be available in the physical design.



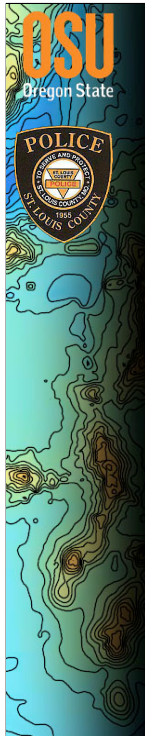
Tuesday, July 14th, 2009

Advances in GIScience: Research Session 2

Results

Title: 0:05/8:05

Moving on to the results of the customization case study.



Tuesday, July 14th, 2009

Advances in GIScience: Research Session 2

Three new object groups

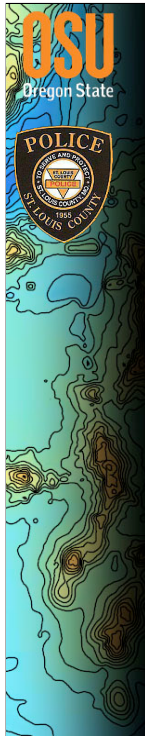
Animal

Telemetry

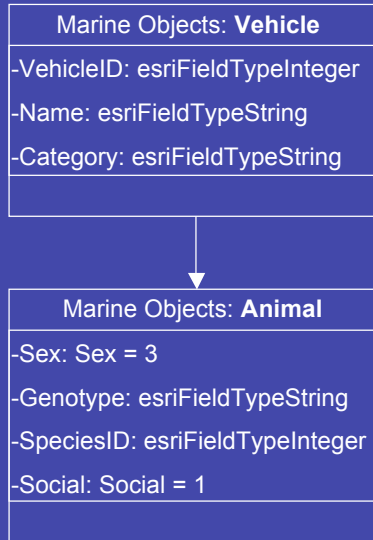
Operations

Object Groups: 0:20/8:25

Three new groups of object classes were introduced into the logical design to model the subjects, data, and operations of satellite telemetry tracking of marine mammals. I will discuss the subject, or Animal, group, and the data, or Telemetry, group, in this presentation. Refer to our paper for discussion of the Operations group and more detailed discussion of the two object class groups presented today.

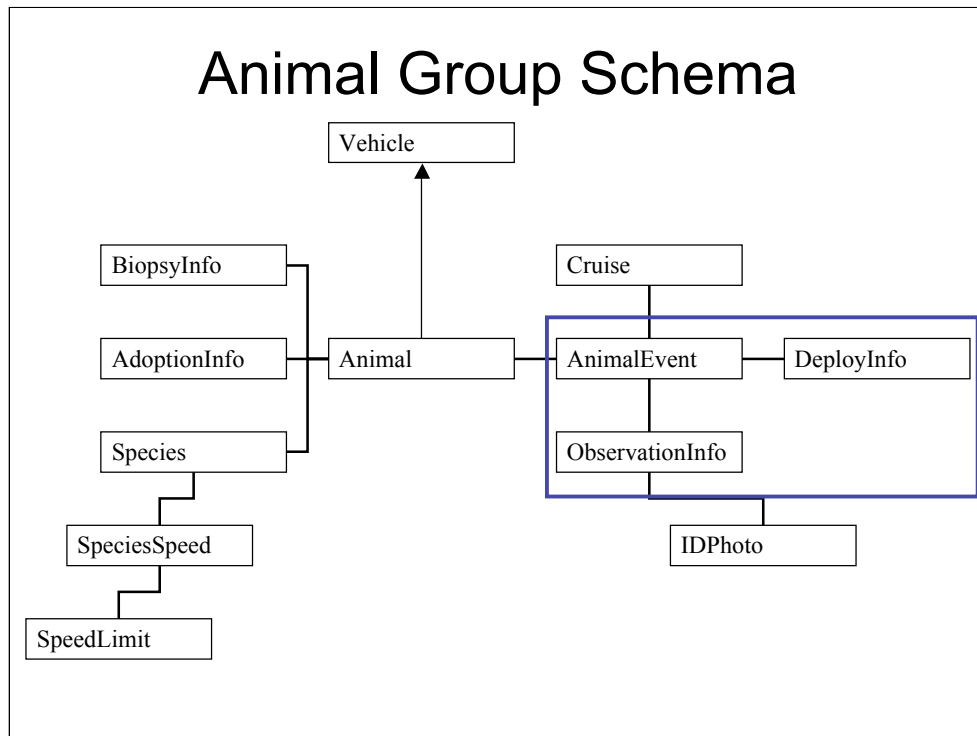


Animal: A type of vehicle



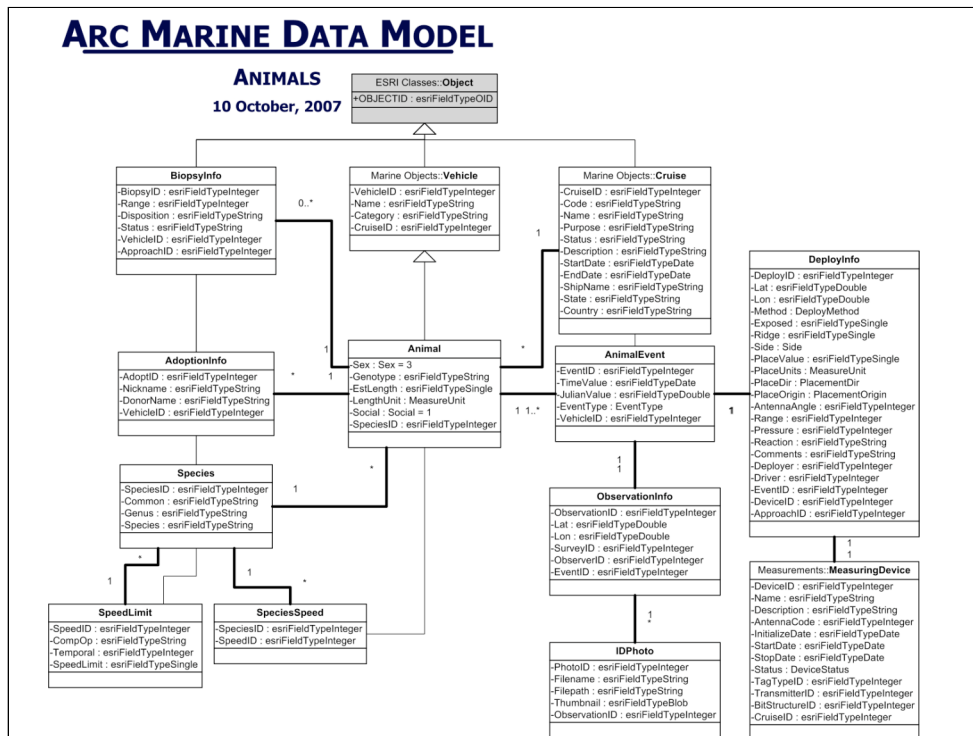
Animal as Vehicle: 0:30/8:55

The first significant choice was modeling the animals themselves. Unlike in a survey, where the animal is just an observation data point in a set of observations, in telemetry tracking the tagged animal is its own moving instrument carrying platform... a Vehicle! The Animal though has its own set of attributes too, not commonly seen on other ocean going research vessels. Sex, Genotype, Species, even social grouping. Though I suppose some research vessels have their own social grouping, that's not part of the core Arc Marine representation of Vehicles.



Animal Group Schema: 0:05/9:00

Here's a conceptual representation of Animal object group with its related dimensions.

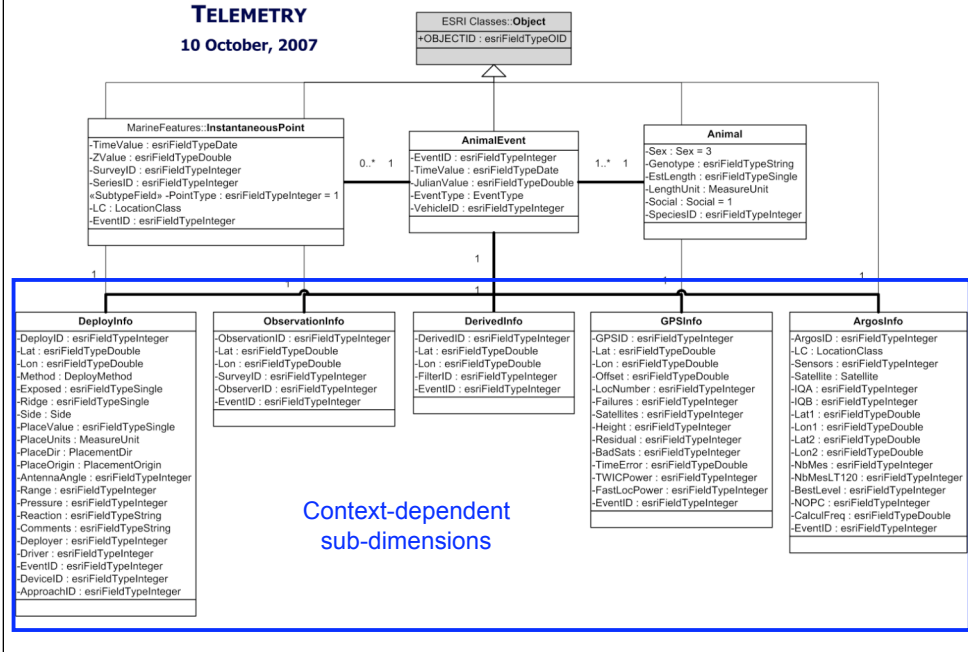


Animal-Logical Representation: 0:45/9:45

And here is its logical representation. The important thing to take away from this slide is that an Animal can be linked to more than just telemetry measurements. It can be linked to Cruise operations, Species specific information (like maximum travel speed), one time observations like biopsies and photographs, even administrative information like the Marine Mammal Institute's animal adoption program. One of the most important linkages though is to a model customization element that is basically a time stamp version of the MarineEvent, an AnimalEvent.

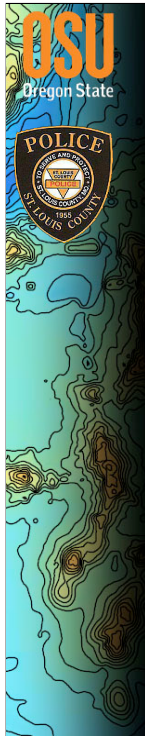
ARC MARINE DATA MODEL

TELEMETRY
10 October, 2007



Context-dependent Sub-dimensions: 0:45/10:30

AnimalEvent is how we tie together an Animal, a location, an event in time (such as an instrument deployment or instrument return), and all the metadata sub-dimensions that go with that specific event type. While each of these sub-dimensions is joined to the AnimalEvent table, which sub-dimension contains information on a particular event is dependent on the event type. In other words, these tables are context-dependent sub-dimensions. While I have listed some of the more common types here, (tag deployment, a field observation, a GPS return, and Argos return) the number of sub-dimensions is limited only by the number of unique data collection methods with unique metadata requirements.



Tuesday, July 14th, 2009

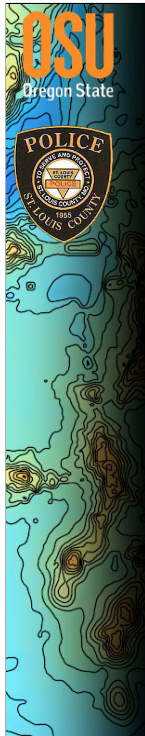
Advances in GIScience: Research Session 2

The Central Question

How can the Arc Marine Data Model be customized to best meet the research objectives of the OSU MMI and the marine animal tracking community?

Central Question: 0:15/10:45

So now that we have looked at the customization results, let's return back to our central question: "How can the Arc Marine Data Model be customized to *best* meet the research objectives of the OSU Marine Mammal Institute and the marine animal tracking community?"



Tuesday, July 14th, 2009

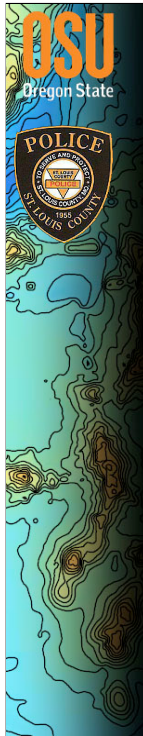
Advances in GIScience: Research Session 2

The Advantages of Satellite Telemetry

- Timeliness
- Relationships to environmental data
- Autonomous profiling
- Continuous Coverage

Advantages Subtitle: 0:05/10:50

To really address this question, we need to return back to those 4 identified advantages of satellite telemetry.



Tuesday, July 14th, 2009

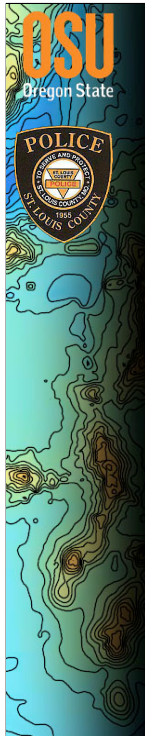
Advances in GIScience: Research Session 2

Timeliness

- Standardized feature classes
- Automated processing for a range of tag types
- Modularized data loading for implementing new data sources

Timeliness: 0:20/11:10

Timeliness of results: To take advantage of timeliness, standardized feature classes and atomic data structures allow for the development of automated processing for a range of tag types. The paper includes a discussion of a recommended application framework, not covered here, to modularize data loading and processing in order to adapt rapidly to new data sources and tag types.



Tuesday, July 14th, 2009

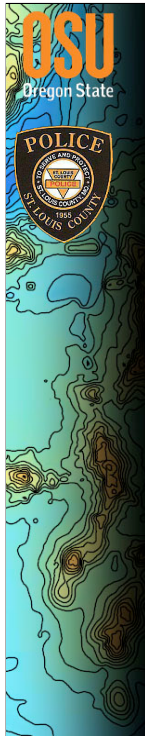
Advances in GIScience: Research Session 2

Data relationships

- Codified in data model
- Interoperability with results from other marine research communities
- Analysis tool interoperability – minimize ETL

Data relationships: 0:40/11:50

Data relationships: The relationship of data to their environmental context is codified directly in the model. Because we are building on top of core Arc Marine, we can move through object relationships to gain interoperability with results from other marine research communities such as physical oceanographers or biological modelers. This reliance on core classes means that an analysis tool built against the core classes will function with our customized classes with minimal or even no extract-transform-load operations. This makes it easier for research groups across disparate marine disciplines to collaborate in analysis tool development projects.



Tuesday, July 14th, 2009

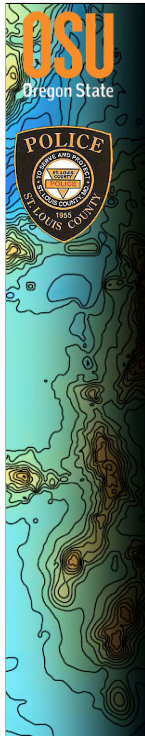
Advances in GIScience: Research Session 2

Autonomous profiling

- Animal as a Vehicle
- Multidimensional data collecting through Measured Data

Autonomous profiling: 0:15/12:05

Autonomous profiling: This advantage is directly addressed by the concept of the Animal as a moving instrument platform. Multidimensional data are collected as measured data by multiple measuring devices, which all converge at the Animal to form the profile of the Animal's environment.



Tuesday, July 14th, 2009

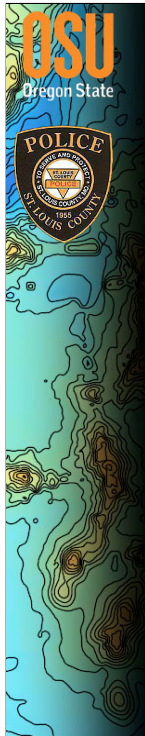
Advances in GIScience: Research Session 2

Continuous Coverage

- Fact tables retain full range of data
- Data warehousing
- On-line Analytical Processing (OLAP)

Continuous Coverage: 0:20/12:25

And finally Continuous Coverage: In the core Arc Marine Data Model, as well as in this customization, data are stored in an atomic form in central fact tables. We can retain the full range of data, not just filtered views, and make it available to data warehousing and on-line analytical processing, this study's area of emphasis. But first we need to address some terms.



Tuesday, July 14th, 2009

Advances in GIScience: Research Session 2

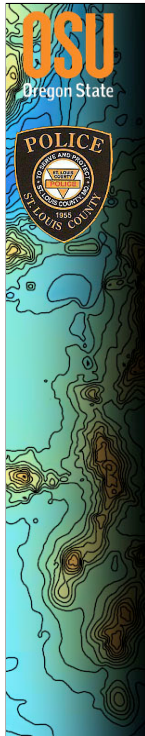
OLTP vs. OLAP

- OLTP:
On-line Transactional Processing
Traditional database function
- OLAP
On-line Analytical Processing
Research database function

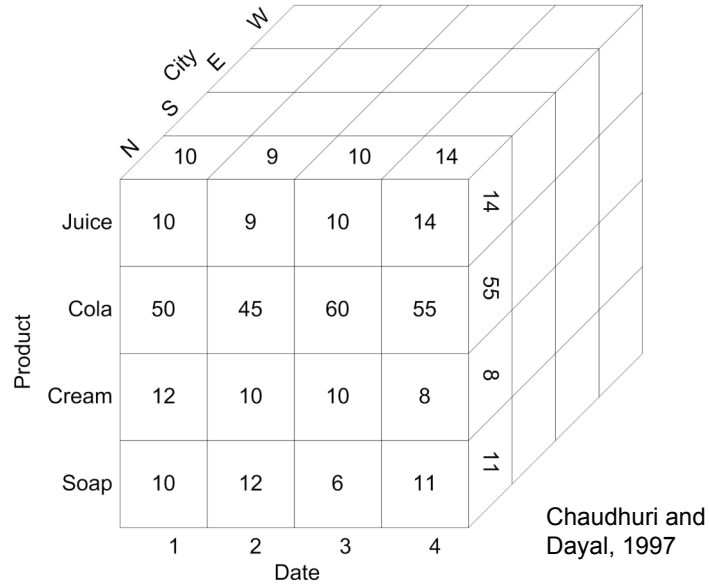
OLTP vs. OLAP: 0:45/13:10

OLTP, On-line Transactional Processing, is the traditional business function of databases. Short transactions, add a record, delete a record, update a record. You want to make a large number of simple queries quickly while maintaining data integrity among multiple editing and viewing access channels. Views are snapshot views of the current state of the data.

OLAP, On-line Analytical Processing, works against the consolidated data produced by OLTP. Data has high dimensionality, is heavily indexed, and historical data is readily available. Queries are complex with multiple aggregations. This is the realm of data mining and the avenue for posing research questions against a database.

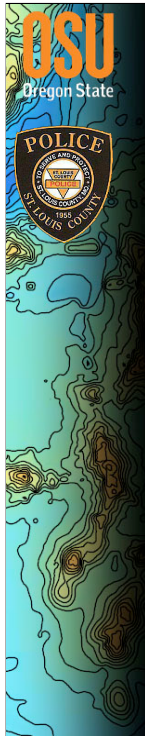


Multidimensional Model

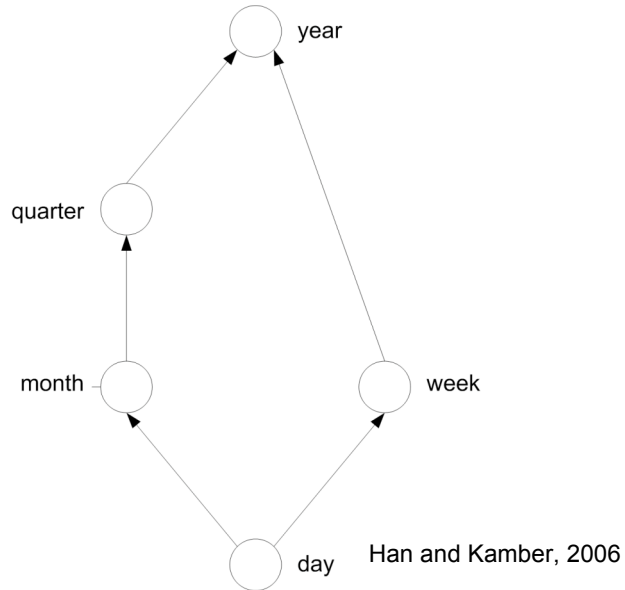


Multidimensional Model: 0:15/13:25

A high dimensional OLAP system can be modeled as a multidimensional cube. Aggregations are represented by dialing up the sides of the cube. Product might be dialed up from specific categories, Juice, Cola, Cream, Soap, to aggregates, Food and Personal Products, following a concept hierarchy.

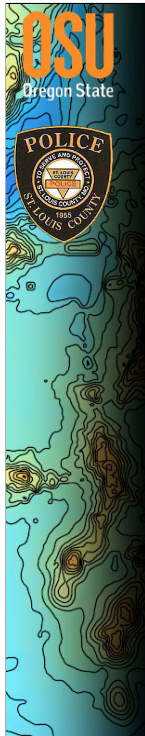


Concept Hierarchy



Concept Hierarchy: 0:20/13:45

The concept hierarchy represents the levels of aggregation for our atomic dimensions. Days roll up to Months which roll up to Quarters which roll up to Years, Decades, Centuries. And a concept hierarchy can be multi-pathed, such as here where day moves to week on to year or day moves to month, quarter, and then year.



Tuesday, July 14th, 2009

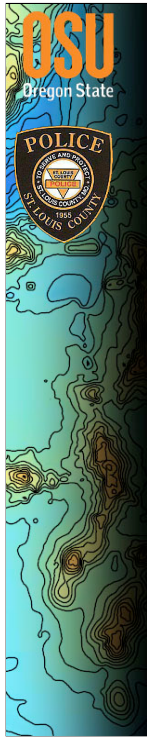
Advances in GIScience: Research Session 2

Arc Marine Star Schemas

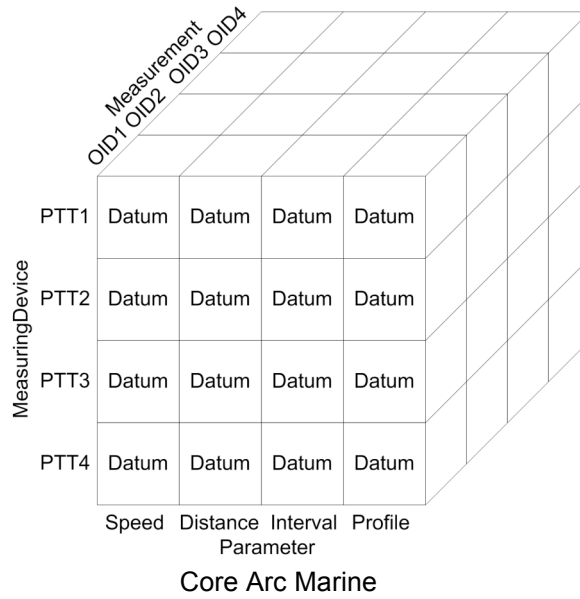
- LocationSeries Point
Spatial atomic measure
- MeasuredData
Scalar atomic measure

Star Schemas: 0:25/14:10

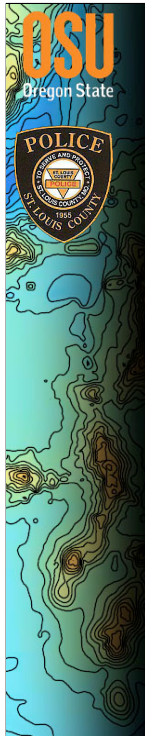
There are two main fact tables in this Arc Marine Data Model customization, each of which is the center of a star schema. The LocationSeries Star has a spatial atomic measure at its center, while the MeasuredData star holds the atomic form of all of our other instrument measured scalar quantities. These are not the only fact tables available in the core Arc Marine Data Model, just the two used in this customization.



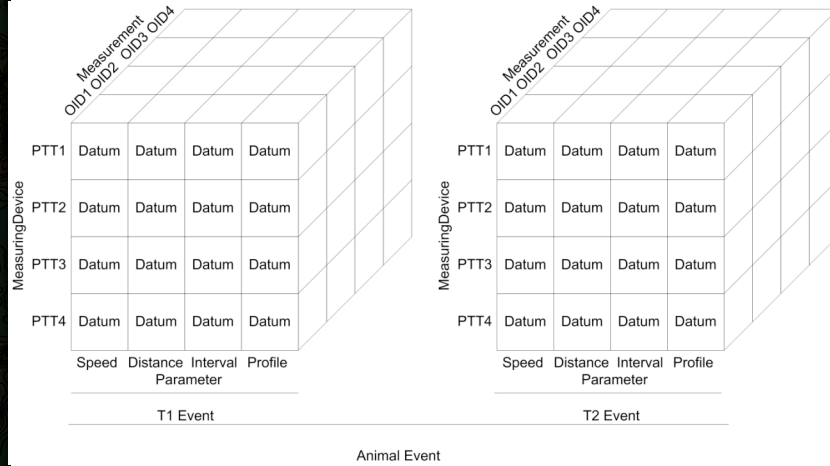
The MeasuredData Star



Core MeasuredData Star: 1:00/15:10



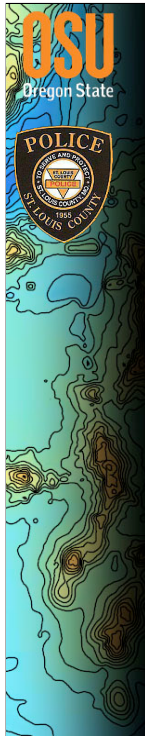
The MeasuredData Star



MMI Customization of Arc Marine

MMI MeasuredData Star: 0:40/15:50

In the MMI Customization, we have added an additional dimension in the form of the AnimalEvent. Timestamp, of course, is available in core ArcMarine, but AnimalEvent ties in those context dependent sub-dimensions that we discussed earlier. Now, through the sub-dimension attributes of these events, we have added an additional layer of dimensionality to the MeasuredData Star allowing for deeper data mining on the basis of the context-dependent measurement metadata. This also more clearly illustrates how each event is a burst of data measurement that can encompass anywhere from a single device measurement to a data stream from every measuring device on the tag.



Tuesday, July 14th, 2009

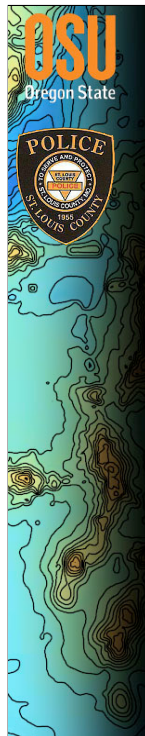
Advances in GIScience: Research Session 2

LocationSeries Point Star

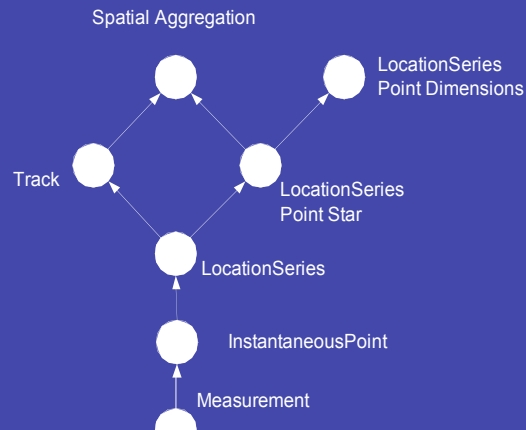
- Two dimensions
 - Animal
 - AnimalEvent
- Spatial Aggregation
- Part of the concept hierarchy of Measurement

LocationSeries Point Star: 0:15/16:05

The LocationSeries Point Star, with a subtype of InstantaneousPoint as its atomic measure, is really a multidimensional square with two dimensions with deep hierarchies, Animal and AnimalEvent.

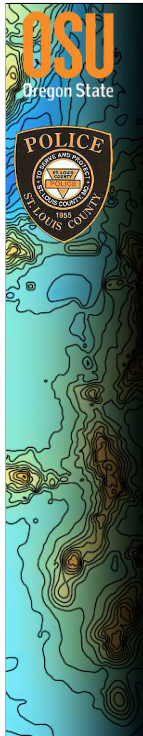


Measurement Concept Hierarchy



Measurement Concept Hierarchy: 0:40/16:45

As we can see here, LocationSeries serves as part of the concept hierarchy for the Measurement dimension of the MeasuredData Star. Thus, atomic scalar measures can be aggregate by point features, which are aggregated by LocationSeries. LocationSeries, in turn opens up the concept hierarchies of the Animal and AnimalEvent dimensions. Finally, LocationSeries and Track lines built off the LocationSeries sequence can ultimately be aggregated along all the different available hierarchies of point and line space.



Tuesday, July 14th, 2009

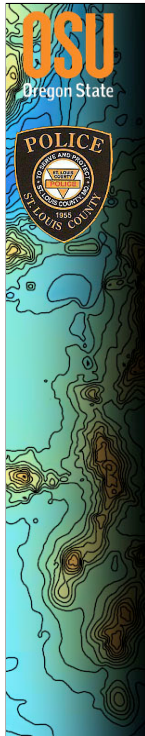
Advances in GIScience: Research Session 2

Key Concepts

- Arc Marine takes advantage of the multidimensionality of the geodatabase object model.
- Arc Marine creates an expandable platform to drive community application development.

Key Concepts: 0:50/17:35

I would like to finish with two key concepts that emerged in this study, concepts which present goals for any community-driven data model project, not just Arc Marine. As I just discussed over the last few slides, Arc Marine takes advantage of the multidimensionality of the geodatabase object model. Atomic scalar and spatial measures are linked to multiple object tables to provide dimension and context that drive complex queries to address research questions. The second concept is more directly addressed in our paper (available here today in the ESRI UC special edition of Transactions in GIS), but hopefully I presented some supporting evidence today. The core Arc Marine Data Model presents an expandable customizable platform that supports the development of interoperable data exchange and analysis tools across the marine GIS community.



Tuesday, July 14th, 2009

Advances in GIScience: Research Session 2

Questions?

Questions: 5:00