

GEOinformatics

Magazine for Surveying, Mapping & GIS Professionals

3

April/May 2008
Volume 11



- Hydro Special
- ERDAS Interview
- Acquisition Sensors
- Mapping Cultural Heritage Underwater
- Arc Marine Data Model

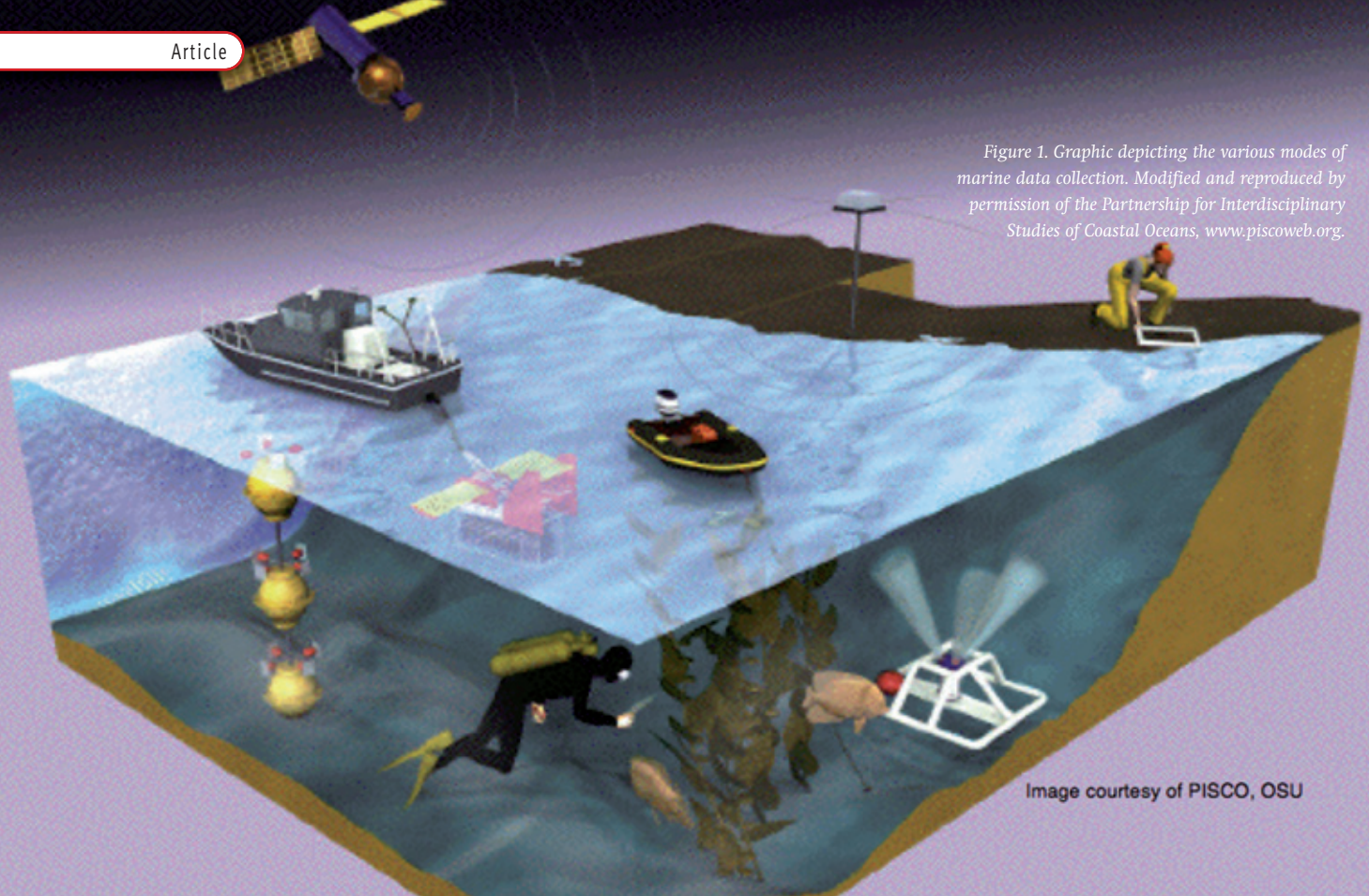


Figure 1. Graphic depicting the various modes of marine data collection. Modified and reproduced by permission of the Partnership for Interdisciplinary Studies of Coastal Oceans, www.piscoweb.org.

Image courtesy of PISCO, OSU

Process and Product

The Arc Marine Data Model

Over the past few years ESRI, with a significant amount of user community input, has been engaged in the exercise of building “industry-specific” data models for ArcGIS. A new marine data model (Arc Marine) has been developed for those who apply GIS to the coasts, estuaries, marginal seas, and/or the deep ocean. This article provides an overview of the activities that led to the release of the data model and accompanying reference book (‘Arc Marine: GIS for a Blue Planet’), with an invitation for those interested to participate in the continued refinement and improvement of Arc Marine via the preparation of additional use cases.

By Dawn J. Wright

One of the most important lessons to be learned from collective experience in the application domain of marine GIS, both published and unpublished, is the importance of rigorous data modeling before attempting to implement a GIS database. Indeed, data models lie at the very heart of GIS, as they determine the ways in which real-world phenomena may best be represented in digital form. With regard to ESRI products, many marine and coastal practitioners and organizations have invested in the coverage data model. Although this has largely been successful, coverages have important shortcomings. For example, features are aggregated into homogenous collections of points, lines, and polygons with generic, 1- and 2-dimensional “behav-

ior.” And there is no way to distinguish between behaviors within feature classes: e.g., the behavior of a point representing a marker buoy is identical to that of a point representing a pulsing transponder; the behavior of a line representing a road is identical to the behavior of a line representing a dynamic shoreline.

In ESRI’s recent object-oriented data model called the geodatabase, GIS features are “smarter”, i.e., they can be endowed with real-world behaviors for individual objects within the same categories, and any sort of relationship may be defined among them. This has wonderful implications for marine and coastal applications, but important questions and concerns were still expressed

at the outset of the Arc Marine initiative in 2001. Given the existing investment, how and when should one make the transition to object-oriented data model and data structure (e.g., the geodatabase in ArcGIS)? How well are marine application domain requirements currently met in object-oriented data structures such as the geodatabase structure? What are the potential benefits? For users, any of the ArcGIS data models provide a basic template for implementing GIS projects (i.e., inputting, formatting, geoprocessing, and sharing data, creating maps, performing analyses, etc.), thereby helping to simplify and standardize the integration of data at various jurisdictional levels (i.e., local, state/provincial, national, global). For developers, they provide a basic framework

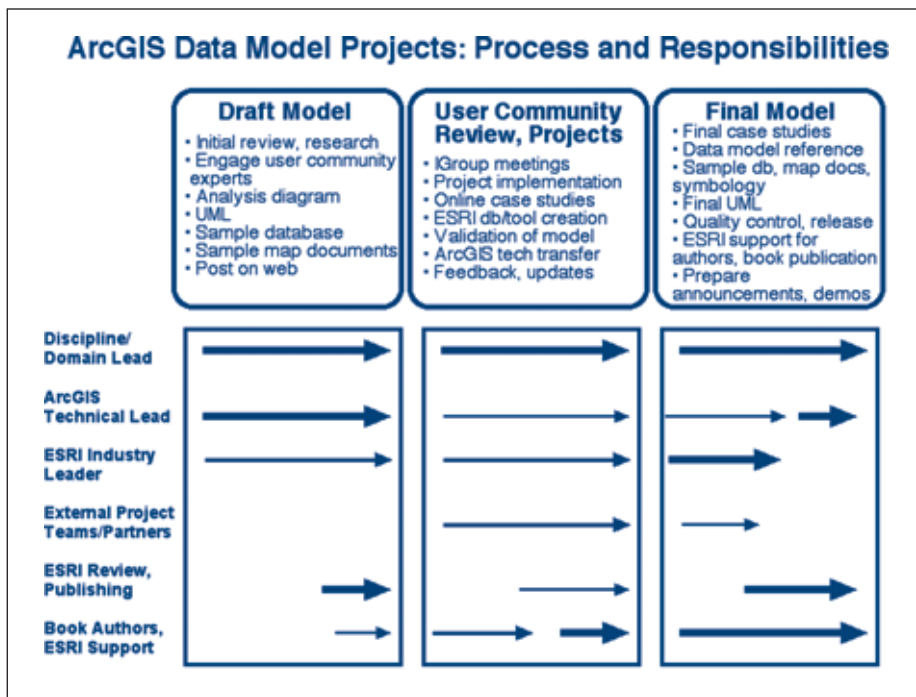


Figure 2. The Arc Marine development process from the standpoint of the core working group but also a broader community of external reviewers, partners, use case testers.

for writing program code and maintaining applications. As such, the specific goals that drove the development of Arc Marine from the outset were:

- Production of a common structure, a “geodatabase template”, for assembling, managing, and publishing marine data in ArcGIS.
 1. For example, the model is specified in an industry-standard modeling notation called the Unified Modeling Language (UML). And because UML code is easily converted to an ArcGIS geodatabase, users can immediately begin populating the geodatabase, rather than having to design it from scratch.
 2. Users can produce, share, and exchange data in similar formats.
 3. Unified approaches encourage development teams to extend and improve ArcGIS software.
- Extending the power of marine GIS analyses by providing a framework for incorporating rules and behaviors in data, and dealing more effectively with scale dependencies.
- In the final analysis, aiding users in a fuller understanding of object-oriented ArcGIS, particularly if making a transition from the use of shapefiles and coverages. For instance, in the core ArcGIS data model, arc-node topology can be used in conjunction with other new and powerful data structures. Routes and regions are relegated to feature classes. And relationships between tables can be preserved,

maintained, and more efficiently managed.

It should be noted that, although Arc Marine was designed to be instantiated as a geodatabase in ArcGIS, its UML diagrams could be used to generate schema for other GIS packages or for other data management solutions. For instance, Arc Marine’s UML has been converted to an XML/GML (eXtensible/ Geography Markup Language) schema, as well as to an ontology in OWL (Web Ontology Language) that can be browsed in tools such as Protégé. Protégé is a free, open source ontology editor used to construct domain models and knowledge-based applications that can be used in taxonomies and classifications, database schemas outside of GIS, and semantic web services (<http://protege.stanford.edu>).

The Process of Building a Data Model

A marine data model working group was initiated in 2001 to begin designing the data model and to address the concerns above

with regard to the ESRI geodatabase. With ESRI’s great success in fostering a grassroots, user community approach to building customized data models for various applications, the working group was eager to accept their advice and consultation on how best to navigate from start to finish. This is illustrated in Figure 2 & 3.

The author served as project manager and, along with Pat Halpin of Duke University, as discipline/domain technical leads. Discipline/domain technical leads are deeply immersed in the subject matter, and are typically leaders in implementing GIS projects within that discipline. They provide technical leadership in the user community in order to help people understand what is being worked on and how it is relevant to their organization or research group. The discipline/domain technical leads need to understand the discipline well enough to talk about the big issues: i.e., current and evolving federal standards, typical projects, what should be important features in the data model design, etc. They are also the design partners of the ArcGIS technical lead.

Michael Blongewicz of DHI Water & Environment served as the ArcGIS technical lead. He developed and maintained the UML diagrams throughout the life of the project, and in collaboration with the discipline/domain leads, facilitated the technical design and reviews of the model, intimately understanding all of the technical issues involved, and able to explain the rationale behind modeling decisions.

Steve Grisé and then Joe Breman served the critical role of the industry manager/disciplinary experts from ESRI, balancing the perspectives of data producers, data users, and business partners within the marine community, while providing crucial input on the data model design and implementation. They also collaborated with discipline/domain leads to organize meetings and workshops, ensuring that key players were from within ESRI were

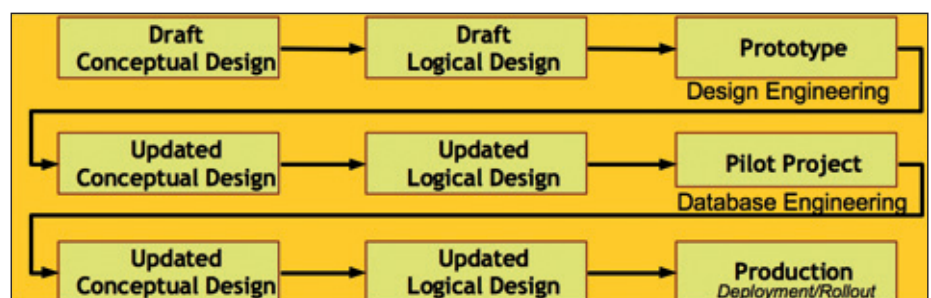


Figure 3. The development process presented from the standpoint of the core working group only, where the technical lead and the domain experts work together on the conceptual and logical designs.

The entire process took six years.

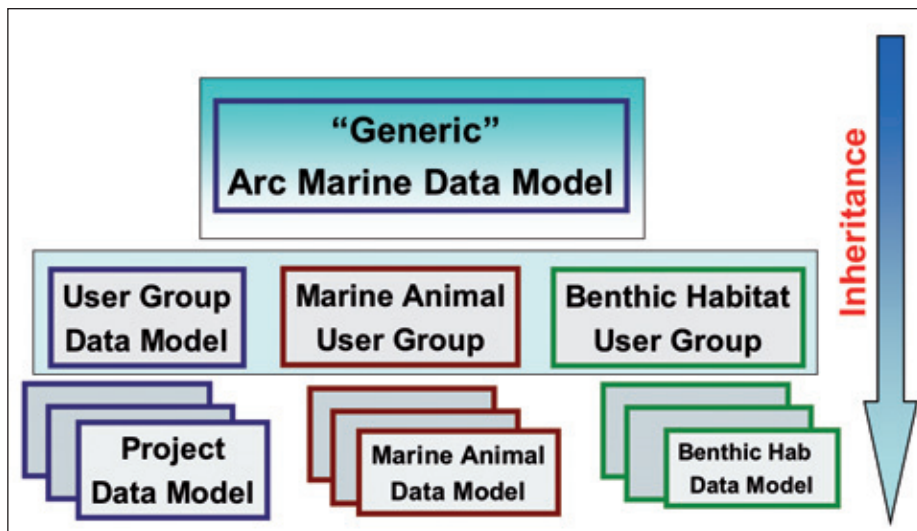


Figure 4. Design strategy for Arc Marine where the core is as exhaustive, temporally dynamic, and generic as possible, but can be customized by various user groups (such as marine animal tracking or benthic habitat mapping) for use within their specific enterprise.

applications. However, those elements that are common to many marine GIS users have been identified to make the core as exhaustive and generic as possible. Therefore, the model is a starting point upon which to build on and leverage the experiences of a broader range of practitioners. Practitioners may take the core, generic data model and add their own attributes, tables and relationships to suit their specific application needs (Figure 4). The generic core of the model should be left intact so that developers can code additional software tools based on the core that take advantage of the known entities in the database. For example, if certain attributes are removed from core feature classes or tables, new software tools (such as DHI's Time Series Manager) will not work properly.

informed and involved.

Over the course of the model's development from 2001 to 2007 (Figure 2 + 3, and sidebar of dusk.geo.orst.edu/djl/arcgis/about.html), there were three 2-day workshops hosted at ESRI headquarters, and at ESRI International User Conferences (UCs) from 2003-2007: three technical workshops, one paper session, and several presentations at marine special interest group meetings. This culminated in the publication of 'Arc Marine: GIS

for a Blue Planet', which was released at the 2007 ESRI UC.

Scope of Arc Marine

Although the focus of Arc Marine is on both the deep ocean and the coast - and it attempts to represent the essential elements for a broad range of marine and coastal data types and processes - it cannot include a comprehensive catalogue of objects meeting the needs of all user groups and

A data model for marine applications will undoubtedly be complex because of the many varied uses of the data. And modern marine data sets are generated by an extremely varied array of instruments and platforms, all with differing formats, resolutions, and sets of attributes (Figure 5). Not only do a wide variety of data sources need to be dealt with, but a myriad of data "structures" as well (e.g., tables of chemical concentration versus raster images of sea surface temperature versus gridded

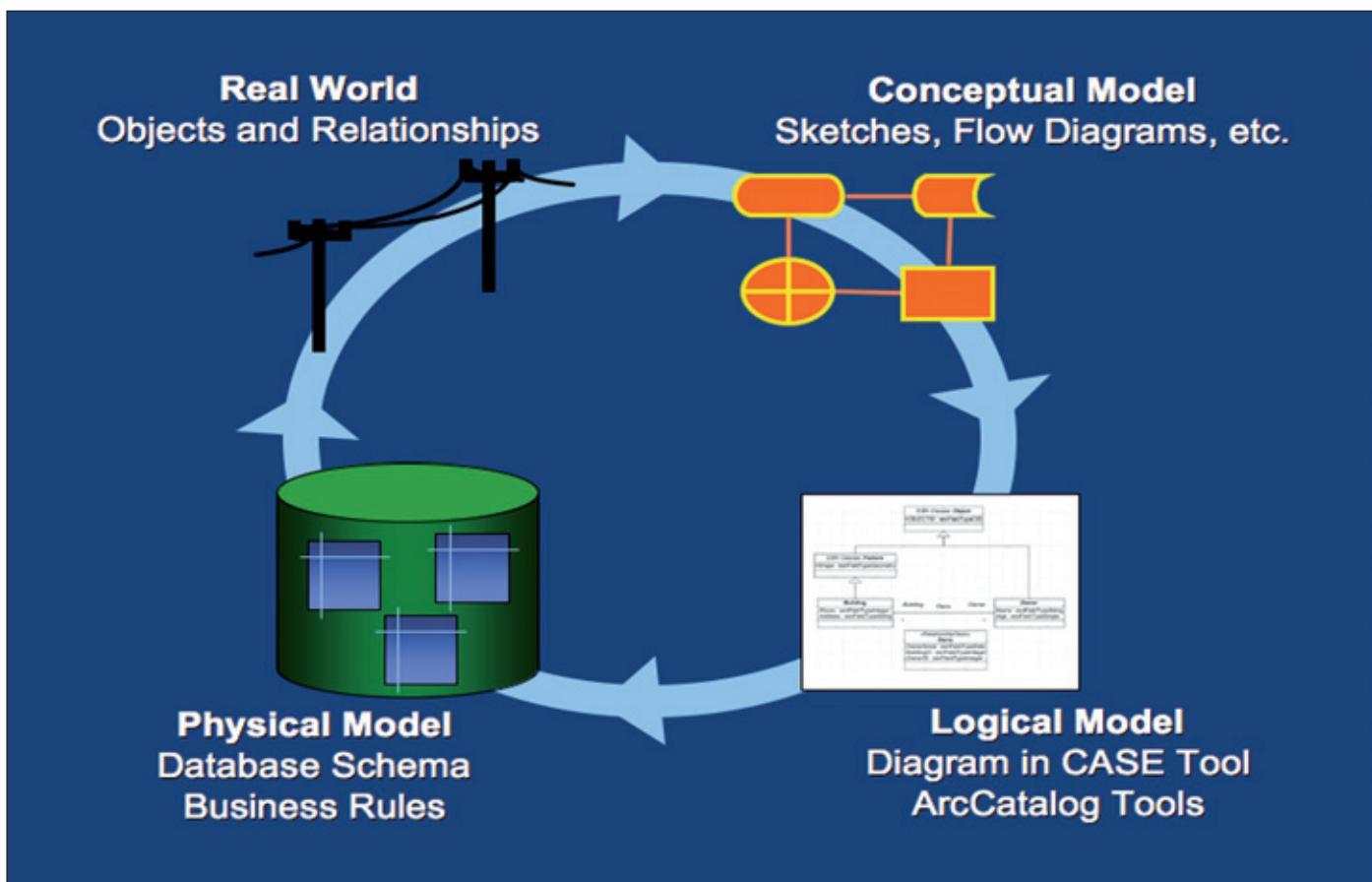


Figure 5. A typical lifecycle in the creation of a customized ArcGIS data model. Modified and reproduced by permission of ESRI.

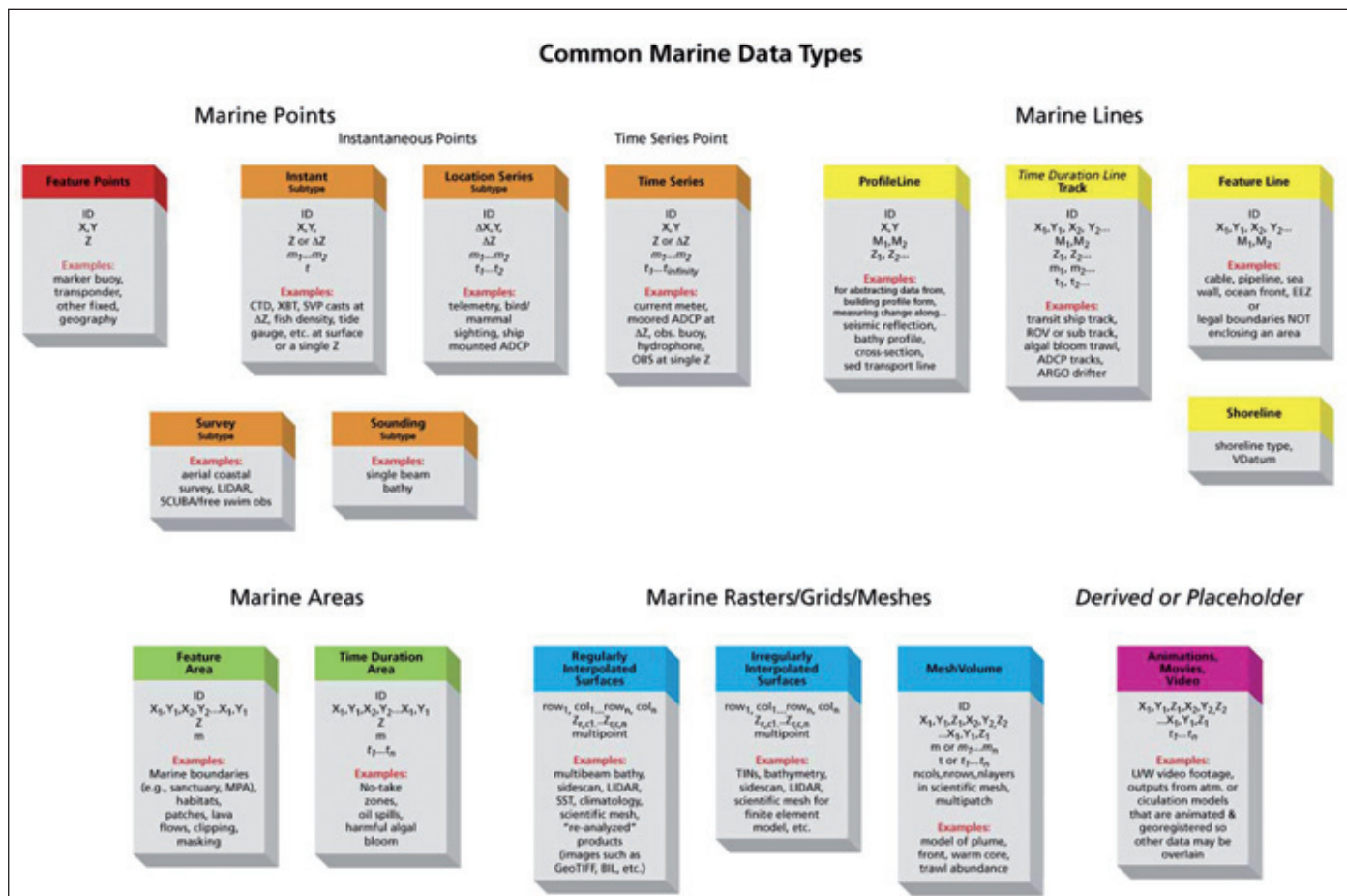


Figure 6. Common data types of Arc Marine from which feature classes, object classes, and relationships were eventually derived for the UML, and ultimately the geodatabase repository. A larger version of this diagram may be downloaded from dusk.geo.orst.edu/djl/arcgis/diag.html.

bathymetry versus four-dimensional data, etc.). It has become increasingly obvious that a comprehensive data model is needed to support a much wider range of marine objects, and that this is essential for advanced management, cartographic and analytical tasks. Arc Marine seeks to identify and organize these objects.

Design Stages

Data model design consists of at least three stages, increasing in abstraction as one goes from human-orientation to implementation in a computer:

1. conceptual design (where the real world is simplified according to application requirements, and an initial model is populated with spatial objects and attributes in the form of entity-relation (ER) diagrams)
2. logical design (where ER diagrams are converted to some kind of schema, such as a table)
3. physical or internal design (where the functions of actual hardware and software must be considered for actual implementation of the model, in the case of ArcGIS as a geodatabase repository, which can then be populated with the user's data)

The first stage in the data modeling process is to define the overall scope and content of the model. From an external design standpoint, this involves the challenging task of identifying the common, essential "things" that are modeled in most GIS projects within an application domain. From a conceptual standpoint, it involves the creation of an analysis or in the case of Arc Marine, a "common marine data types" diagram (Figure 6), with the identification of major thematic groups and an initial set of object classes within these groups. The common marine data types diagram was therefore at the "conceptual design" stage in the data modeling process for Arc Marine, and was created and refined at the first two user workshops hosted at ESRI headquarters in Redlands, California.

Moving to the logical design stage, the common themes from the common marine data types provided the major feature and object classes, and relationships that were defined in UML (e.g., Figure 7). These were organized into major groupings containing a set of feature classes and supporting object classes, with the relationships between those classes. Each feature and object class consists of a descriptive name and a set of defining

attributes. All marine feature and object classes inherit properties from one of the basic feature and object types that are provided in ArcGIS. The major groupings hold related or complementary sets of essential feature classes and objects within the larger data model, related by function or type. Arc Marine groupings include: Marine Points, Marine Lines, Marine Areas, Time Series & Measurements, and Mesh (used for storing the results of numerical modeling of marine phenomena, particularly finite element modeling).

With UML in hand, a user may participate in the physical design stage by taking advantage of an existing collection of CASE (computer aided software engineering) tools in ArcGIS in order to generate their own schema. For example, an empty geodatabase can be created in ArcCatalog, and the Schema Wizard used to translate to only those portions of the UML that are needed for a particular project to an XMI (XML Interchange) template or *.mdb repository. This in turn allows the user to populate that geodatabase with their own data for use in a specific GIS project, with all the necessary feature classes, attributes, and relationships from the data model intact.

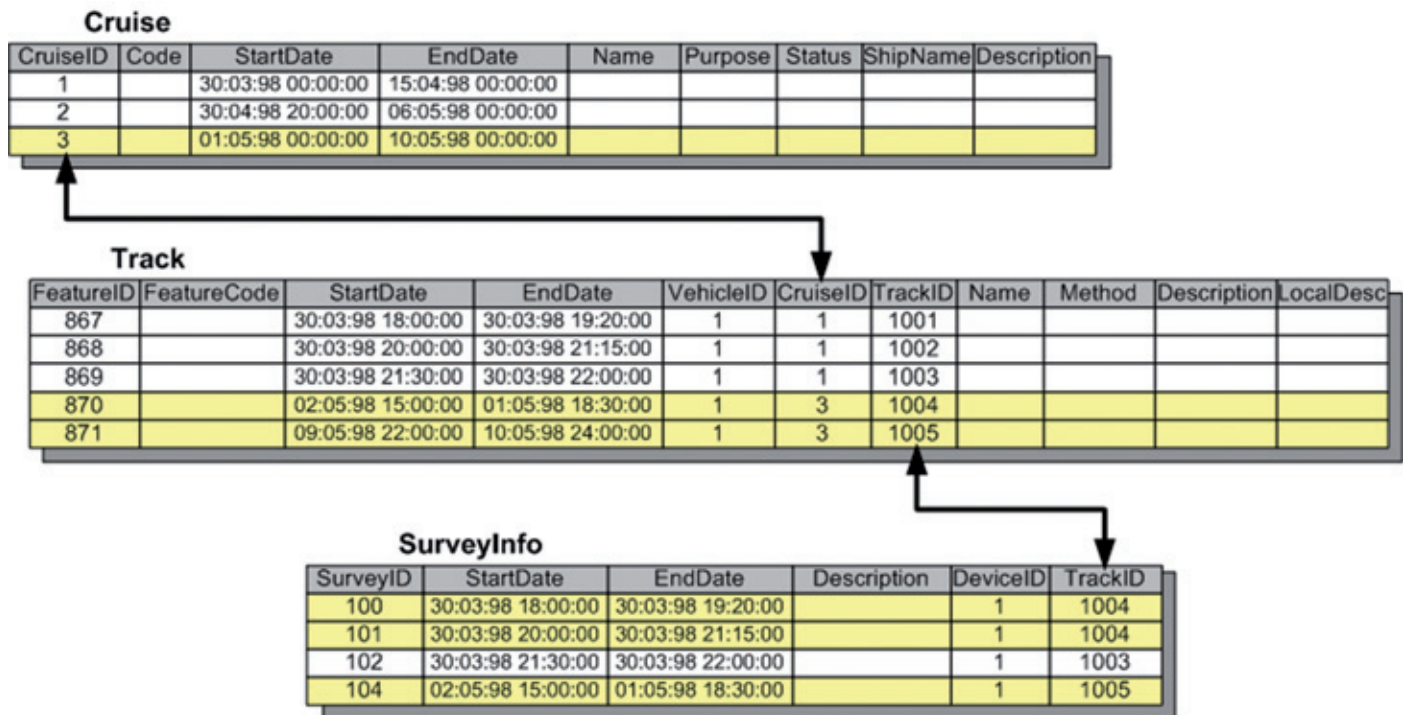


Figure 7. A diagram representing a fragment of UML from Arc Marine, showing the connectivity between data collected on a Cruise to a track and multiple surveys which may be conducted with any number of vehicles or instruments within said cruise.

Conclusion: Final Data Model Content, Purpose and Use

The model has now reached initial maturity with the following products available to the international marine community:

1. A final reference book, 'Arc Marine: GIS for a Blue Planet', and companion web site at dusk.geo.orst.edu/djl/arcgis describing and explaining the data model, demonstrating its implementation in ArcGIS, and providing thirteen use cases of various marine applications (along with feature and object class glossaries, lessons learned and helpful tips/tricks). The related items below are all available on the companion web site.
2. The Arc Marine data model itself, specified in UML (via Microsoft Visio software) and further as an XMI (XML Interchange) template or *.mdb geodatabase file, that can be used by most marine application development teams as a starting point for structuring marine data. The physical UML is a detailed view of the design, providing specifications of data types, relationships, and other details. The specific data needs of a given user may require modifications of, or extensions to, the basic data model.
3. A large reference poster that presents the basic, conceptual structure of the data model in UML. Project teams can use this to review the data model with users and stakeholders.
4. An Arc Marine tutorial for new users, available on the companion web site.

5. Several case study data sets that can be used to demonstrate the efficacy of the marine data model. The data sets will be composed of public domain, marine data that may be freely used to support specific applications. Current case study databases and descriptive information are available on the companion web site. Helpful tools accompany some of the case studies (e.g., tools to import various data sets into Arc Marine, to manage and analyze time series data, to visualize data).

The user community is now invited to further refine this model by preparing additional applications of Arc Marine using their own data sets in order to test its viability and usability. For instance, the reference book features use cases in seabed mapping and characterization, marine animal tracking, sediment transport and shoreline evolution, coastal planning, coral reef monitoring and habitat analysis, water column chemistry, numerical modeling of physical oceanographic parameters, and more. But the preparation of additional use case in these and other areas (e.g., ocean fisheries, marine pipelines, marine protected area design, species population viability, etc. are welcome and may be coordinated into additional ESRI UC sessions or addenda to the reference book.

It is also useful to understand what relationships may exist to related ArcGIS data

modeling efforts such as Arc Hydro, the S-57 Data Model (for electronic nautical charting), the Atmospheric Data Model, and even the ArcGIS Parcels Data Model for coastal land development. Once a data model is designed and initially implemented, then the many ways it will be used will become clearer, and it can be refined accordingly.

Dr. Wright is professor of geography and oceanography at Oregon State University. She has for many years been interested in solving problems related to the management and spatial analysis of oceanographic data, with an emphasis on application issues for GIS.

Visit her Home Page at: dusk.geo.orst.edu
 Davey Jones' Locker Lab – dusk.geo.orst.edu/djl
 Main Arc Marine Web Site – dusk.geo.orst.edu/djl/arcgis

Acknowledgements: The author gratefully acknowledges her collaborators on Arc Marine, Michael Blongewicz, Pat Halpin, and Joe Breman, as well as Steve Grisé and Simon Evans of ESRI, and international and case study team listed at dusk.geo.orst.edu/djl/arcgis/people.html, all of whom are deeply thanked for providing helpful guidance, critique, and, at times, good humor, in creating a successful data model.