# DIFFERENCES AMONG THE DATA MODELS USED BY THE GEOGRAPHIC INFORMATION SYSTEMS AND ATMOSPHERIC SCIENCE COMMUNITIES

Stefano Nativi, University of Florence, Prato, Italy;
and M. B. Blumenthal, J. Caron, B. Domenico, T. Habermann, D. Hertzmann, Y. Ho, R. Raskin, and J. Weber

Last Modified: November 3, 2003

## Abstract

In the Earth science disciplines, both the observational instrumentation and numerical forecasting technology used to generate data are improving so rapidly that the techniques available to manage and use the resultant datasets are struggling to keep pace. A notable example is represented by the atmospheric science discipline.

As observational and model output datasets in the atmospheric science community increase in resolution, there is an increasing demand to cross the boundaries between the GIS (Geographic Information Systems) and ASIS (Atmospheric Science Information Systems) communities. For example hydrologists, who traditionally use GIS, are interested in incorporating radar information into their GIS analysis and modeling systems. On the other hand, researchers and educators in the atmospheric science are interested in integrated views of terrain, infrastructure, and demographic data (typically in GIS data systems) with atmospheric data from forecast models, satellites, and radar data. Differences in the way the two communities think about their data can give rise to difficulties in integrated analysis and display of datasets from the two disciplines. For example, the atmosphere inherently has three spatial independent variables while the GIS community focuses on two. Furthermore, the atmosphere changes on time scales much shorter than those usually considered within the GIS community. Consequently, the atmospheric scientist thinks in a 4-dimensional space and requires a 4-dimensional data model. The paper presents a general, abstract view of differences between the data models of the two communities as well as a schematic description of where the data systems (traditional GIS, evolving systems based on Open GIS specifications, and traditional ASIS) overlap and where they are distinct from each other. Examples in each category are described. Finally, even for the datasets which seem to lie in the area of overlap, some of the difficulties inherent in the integration process are discussed along with solutions where they have been developed.

## GIS and ASIS abstract data models

In the Earth sciences, there are many conceptual models for the datasets in each subdiscipline. For the purposes of this discussion, the focus will be on atmospheric science because, in many cases, the data models differ dramatically from those in the GIS community.

In order to understand the ability of GIS data models to represent Atmospheric Science (AS) datasets, it is useful to consider the following questions:

1. How important is the geographic aspect for AS data?
2. How well is time modeled?
3. How much of AS semantics is captured?

## Geographic aspect

Geographic information can be defined as: *information concerning phenomena implicitly or explicitly associated with a location relative to the Earth* [ISO 19101]. Therefore, it is possible to distinguish two main topics: observed phenomena and Earth locations.

Due to the intrinsic nature of AS and the associated acquisition technologies (e.g. multiparametric remote sensing techniques), AS datasets primarily are used to capture and represent information related to complex observed phenomena. Conversely, the Earth location aspects of the datasets are traditionally kept as simple as possible. This approach stems, in part, from the fact that spatial resolution and geo-collocation have been inaccurate (as compared to GIS data) for many satellite datasets. The most used structures for representing observed phenomena are implicitly positioned over the Earth (e.g. regular grids); metadata describing dataset coordinate reference system (if any) rarely include values on geo-datum.

On the other hand, traditional GIS models consider the Earth location as important as the phenomenon itself. This is due to the need to enable complex and precise spatial queries. Consequently, most popular structures for representing observed phenomena (e.g. geometric entities, such as: point, line, polygon, etc.) are explicitly positioned over the Earth, enabling extremely complex topology-based functionality. Moreover, metadata about the coordinate reference system are extremely important.

## Time modelling

Time is essential for understanding AS phenomena. It can be expressed in units ranging from seconds (e.g. rainfall variations measured by a sequence of radar scans) to centuries (climatological variations calculated through complex models). Both *running clock* (e.g. experiment time) and *epoch based* (e.g. calendar time) approaches are commonly used. For AS data, time location and evolution of observed phenomena are as important as spatial location.

Historically GIS data are characterised by slow temporal evolutions (e.g. political boundaries, social statistics, infrastructure networks, etc.). Traditional GIS can manage time (e.g. as a layer attribute) but generally time series are not supported (e.g. visualisation of plume trajectories). The *epoch based* approach is generally the only one supported.

## AS semantics

In terms of semantics, AS datasets represent the measurements of the sensors from which they are collected [ISO 19129 DW2] or, in the case of synthetic data (e.g. the output of forecast models), the source from which it was generated. These aspects are not a typical component of traditional GIS data models.

Moreover, AS datasets are commonly characterised by aggregation structures. Determining the right granularity levels which characterise a complex AS dataset is a real issue when dataset semantics must be captured and modelled. Several extremely useful aggregation structures – used at different levels -- are:

- complex hierarchical tree (e.g. multiparameter complex datasets),
- simple trees (e.g. time series),
- grid cell aggregations (e.g. clusters, regions, topological sets, etc.)
- fiber bundles (e.g. multichannel satellite imagery), etc.

Aggregation structures characterising traditional GIS datasets are less complex, such as simple trees (e.g. views made up of thematic layers).

# Conceptual differences between GIS and ASIS data models

To develop a concrete understanding of the conceptual differences between GIS and ASIS data models, it is helpful to compare some of the most common abstract models for both systems. For ASIS, simplified schemas of the netCDF and the VisAD (function modelling) abstract models are described in the following sections. For GIS, a simplified schema of the *general feature* abstract model (used by OpenGIS and ISO TC 211) is considered. Content models stem from abstract models, defining metadata which are introduced in order to describe data model concepts and their relationships.

## VisAD abstract data model

The VisAD data model assumes that data objects are approximations to mathematical functions. VisAD data objects may be simple real number values, text strings, vectors of real numbers, arrays such as images or grids, or complex hierarchies of data.

*The central metadata of each data object is its mathematical type, which is a kind of data schema. The type defines names for primitive numerical and text values, groupings of these values, and functional relations between values. For example, an earth image may have the type: ((latitude, longitude) → radiance) which says that an image is a functional relation from (latitude, longitude) pairs to radiances. Other metadata associated with an earth image data object may include sampling geometry and topology for (latitude, longitude) pairs, units for latitude, longitude and radiance, a coordinate transformation relating (latitude, longitude) to some reference coordinate system (e.g., a standard Mercator map), missing data indicators for radiance values, and error estimates attached to latitude, longitude and radiance values. A time sequence of earth images may have the type:*
*(time → ((latitude, longitude) → radiance)) with additional metadata for time units and sampling* [Hibbard].

The following figure depicts a simplified schema of the VisAD functional model.

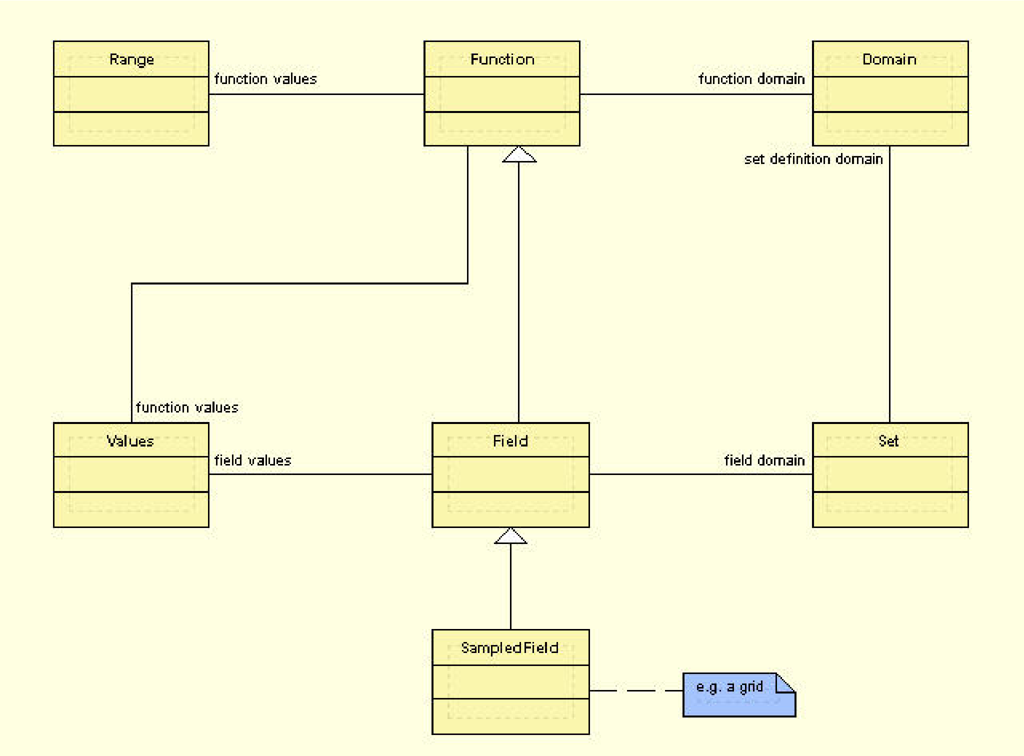

**Figure1**: Simplified schema of the VisAD functional model

## NetCDF abstract data model

The netCDF data model contains dimension, variable, and attribute objects which are all characterised by both a name and an ID value by which they are identified. These objects can be used together to capture the meaning of data and relations among data fields in an array-oriented dataset. This is shown in the following diagram.

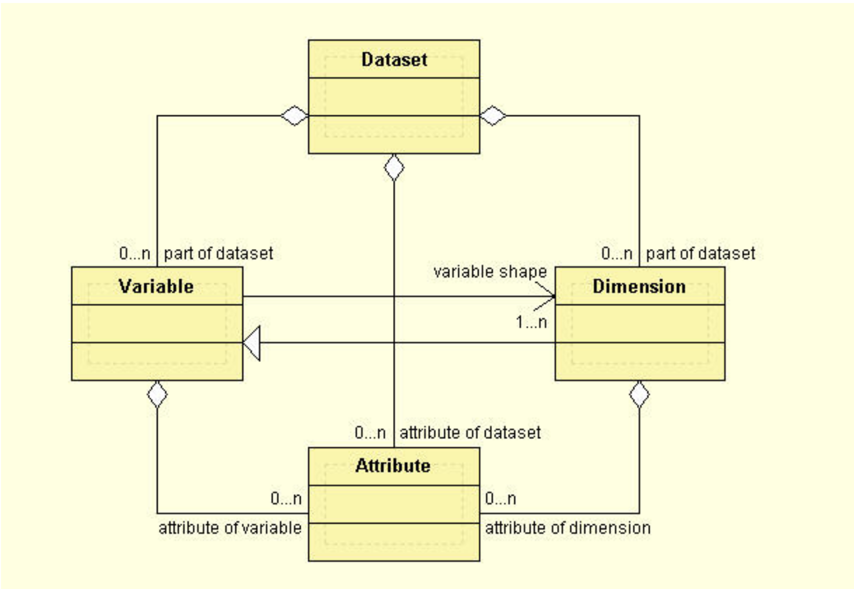

**Figure2**: NetCDF data model schematic

The following short netCDF example is taken from an official tutorial (by Russ Rew, Glenn Davis, Steve Emmerson, and Harvey Davies Unidata Program Center) and it is useful to illustrate the concepts of the depicted data model. The notation used to describe this simple netCDF dataset is called CDL (network Common Data form Language), which provides a convenient way of describing netCDF files.

```
netcdf example_1 {  // example of CDL notation for a netCDF file

dimensions:           // dimension names and sizes are declared first
        lat = 5, lon = 10, level = 4, time = unlimited;

variables:            // variable types, names, shapes, attributes
        float   temp(time,level,lat,lon);
                    temp:long_name     = "temperature";
                    temp:units         = "celsius";
        float   rh(time,lat,lon);
                    rh:long_name = "relative humidity";
                    rh:valid_range = 0.0, 1.0;      // min and max
        int     lat(lat), lon(lon), level(level);
                    lat:units       = "degrees_north";
                    lon:units       = "degrees_east";
                    level:units     = "millibars";
        short   time(time);
                    time:units      = "hours since 1996-1-1";
        // global attributes
                    :source = "Fictional Model Output";

data:                 // optional data assignments
        level   = 1000, 850, 700, 500;
        lat     = 20, 30, 40, 50, 60;
        lon     = -160,-140,-118,-96,-84,-52,-45,-35,-25,-15;
        time    = 12;
        rh      =.5,.2,.4,.2,.3,.2,.4,.5,.6,.7,
                 .1,.3,.1,.1,.1,.1,.5,.7,.8,.8,
                 .1,.2,.2,.2,.2,.5,.7,.8,.9,.9,
                 .1,.2,.3,.3,.3,.3,.7,.8,.9,.9,
                  0,.1,.2,.4,.4,.4,.4,.7,.9,.9;
}
```

The netCDF model has been extended in order to model a dataset’s location aspects and aggregation structures. The following picture depicts the abstract model of an extension.

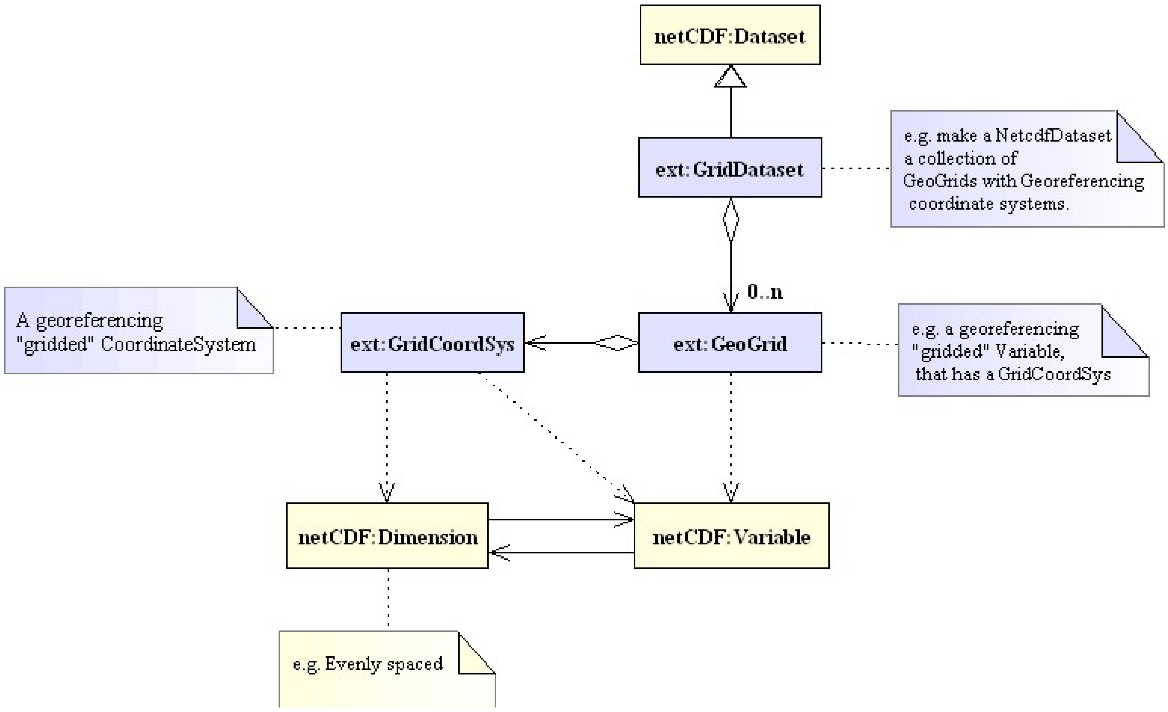


**Figure 3**: Extended netCDF data model

## The GIS General Feature abstract model

Both OpenGIS and ISO TC 211 specifications are based on the so-called *general feature* model; a simplified abstract schema of such approach is showed below.

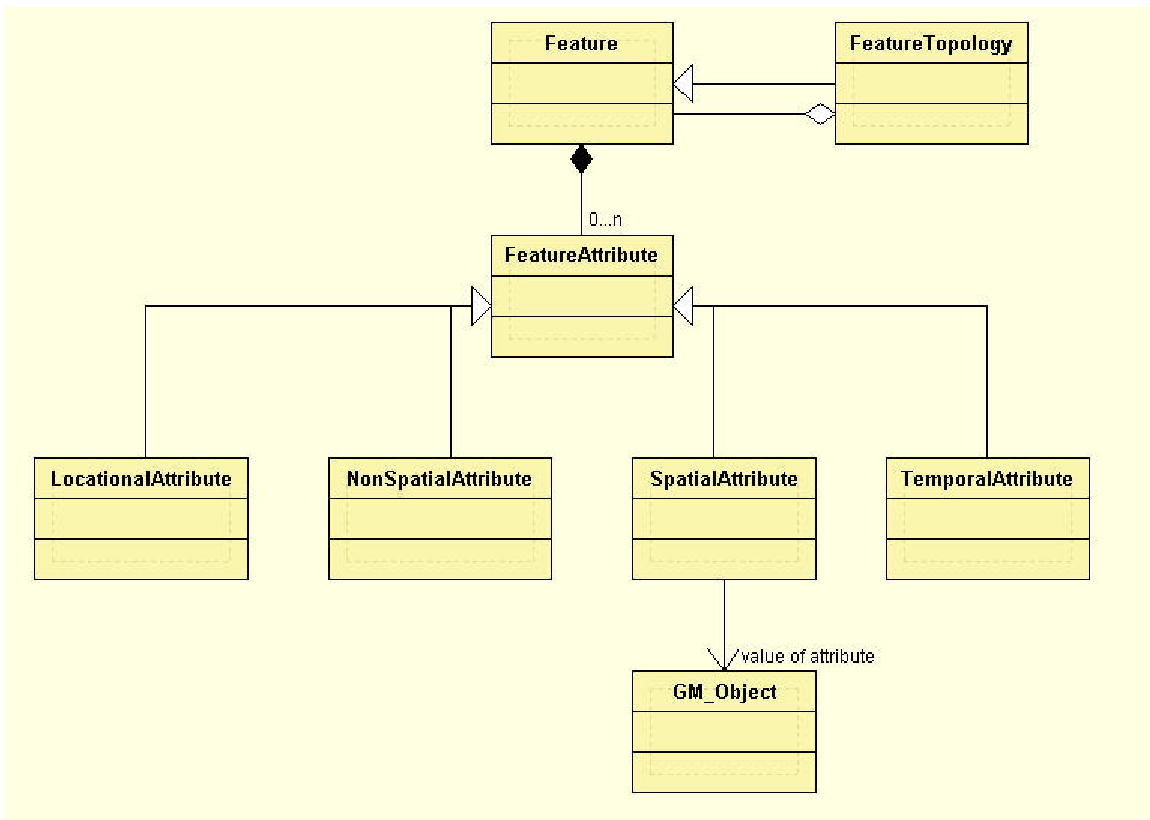

**Figure 4**: GIS general feature model

In order to understand the different philosophy that characterises this approach in contrast to the previously ones, it is useful to expand the geometry object (i.e. GM_Object). The next figure shows a simplified schema of ISO 19107 geometry basic types. These objects are different from the implicit geometries used by both VisAD and NetCDF abstract models for representing typical AS datasets (i.e. regular and irregular multidimensional grids, or sampled fields).

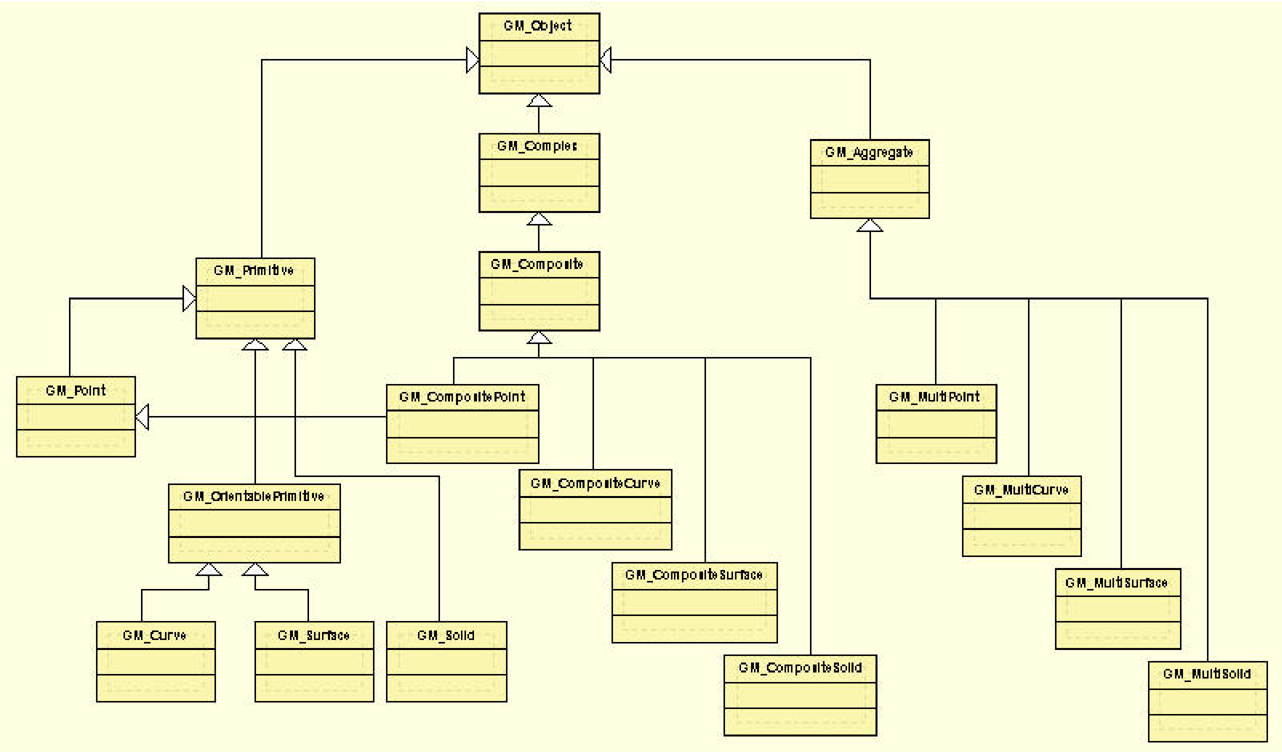


**Figure 5**: Simplified schema of ISO 19107 geometry basic types

Another important difference consists in the respective root concepts: *data objects* for AS models, *feature objects* for GIS model.

# GIS meets ASIS: the Coverage concept

With the advent of new and more powerful remote sensing techniques, and spurred by the Information Society's needs, the GIS community has been working on solutions for "importing" AS datasets. Hence GIS data models have been reshaped and extended to accomplish this ambitious task. International initiatives (e.g. ISO TC 211 and OpenGIS) have released geo-information standard models which conceived to support general interoperability. These efforts lead to the definition of "more general" models for geographic information.

Such models distinguish two kinds of geographic information: boundary and coverage data. Boundary data is often called "vector data" and is almost always feature-oriented. Generally, AS datasets fit into the coverage category and, in most cases, are grid-oriented.

In GIS, the coverage concept can be defined and explained as: *a feature that acts as a function to return one or more feature attribute values for any direct position within its spatiotemporal domain* [ISO 19123].
*GIS coverages (including the special case of Earth images) are two- (and sometimes higher-) dimensional metaphors for phenomena found on or near a portion of the Earth's surface. Fundamentally, coverages (and images) provide humans with an n-dimensional (where n is usually 2, and occasionally 3 or higher) "view" of some (usually more complex) space of geographic features…. the "view" will be geospatially registered to the Earth… A coverage is a special case of (or a subtype of) feature* [The OpenGIS™ Abstract Specification Topic 6: The Coverage Type and its Subtypes].

According to these definitions, it is clear that coverage is a key concept for bridging the gap between GIS and AS data models. Nonetheless, the coverage concept is part of the GIS semantics. In fact, a coverage is defined as a *feature* subtype. Hence, it must be characterised by explicit geometries. Meanwhile, AS datasets are generally characterised by implicit geometries, such as regular or irregular grid matrices.

In GIS, grids are modelled as a set of geometric objects (e.g. GM_Points, GM_Curve, GM_Surface, etc.); besides, *DiscreteCoverageFunction* objects map each geometric object to a tuple of attribute values.

The following schema is a simplified view of the abstract model introduced by ISO 19123 (i.e. Coverage geometry and functions specifications) for modelling grid matrixes.

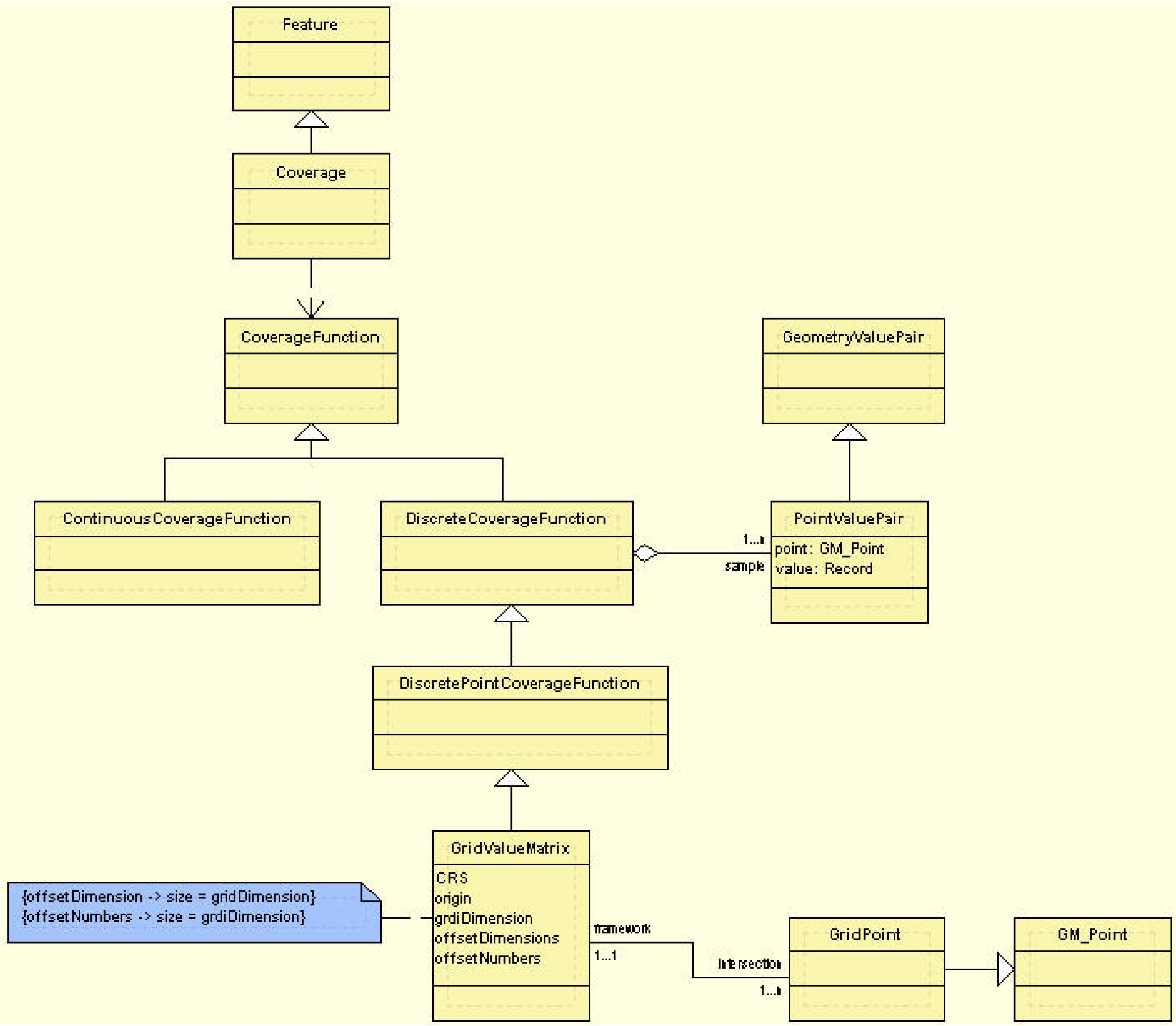


**Figure 6**: Simplified view of the abstract model introduced by ISO 19123

Referring to the coverage model, a grid may be defined only with respect to a particular coordinate reference system.
In GIS, the most common grids occur in a two-dimensional space, and are themselves two-dimensional; three-dimensional grids are not unusual [ISO 19123]. In other words, a simple grid is the most common data structure used for storing coverage information in real GIS data models.

AS datasets can be classified as hyperspatial data (i.e. *space comprising more than the three standard x, y and z dimensions* [Oracle's server Spatial Cartridge Glossary]) or multidimensional discrete data (MDD). It is also possible to call them Hyperspatial Structured Data (HSD). Generally speaking, AS datasets are characterised by the following attributes:

- high dimensionality (e.g. four or more dimensions: space, time, pressure, etc.);
- continuous parameters (i.e. the functions composing a dataset) in nature, but sampled parameters in acquisition and storage;
- multivalued parameters: characterised by a tensor rank; for example, a brightness temperature parameter can be characterised by several values -one for each of the N channels of acquisition- therefore, its rank is equal to N, and the number of its elements are dimensionrank.

AS data models, such as the VisAD and NetCDF models, are generally able to represent such data complexity. On the other hand, the only way to reconcile such complex data structures with the simple grid model, as presently supported by GIS, it is to convert AS data back to a simple grid [ISO 19130].

In summary, the coverage structure – supported by GIS -- seems to be the best solution available to bridge GIS and ASIS data models, but it doesn't capture the entire complexity of AS datasets. To be fully interoperable in the GIS data environment, AS data must be simplified so they can be represented as coverage objects. In essence this can be seen as a projection from the more complex AS data space to the simpler GIS data space.

Regarding semantics, ISO 19129 WD2 -dealing with Imagery Data Framework specification- introduced the following *Image and Gridded data structure model elements*:

- Geometric structure (e.g. grid metadata), with associated spatial and temporal reference systems (e.g. Coordinate Reference Systems metadata);
- Representational structure (e.g. dataset encoding metadata);
- Metadata (e.g. source sensor metadata)

Up to now, this metadata doesn't seem to be sufficient or appropriate for all types of AS datasets (e.g. complex forecast model output).

# Implementing abstract data models

Both GIS and ASIS implement abstract data models (entirely, or just useful profiles of them) in order to achieve three essential service/funtionality types: data management (i.e. storing and querying), data processing and data visualization. The advent of the Web era, and the new needs posed by the Information Society have brought to the fore the importance of another kind of characteristic of online data services: interoperability. The Web and underlying Internet have given rise to new, alternative facilities:

- service-oriented approaches to interacting with data,
- new data models (e.g. semistructured data models), and
- augmented communications technologies and protocols (e.g. SOAP/WSDL/XML).

Exchanging data in a neutral, standard and self-descriptive way has become necessary in order to exploit the expanding system of distributed by interconnected computing platforms that are rapidly becoming ubiquitous. Semantic interoperability is becoming more important for ensuring optimal synergy among these systems.

It is our opinion that interoperability frameworks (i.e. services and data models) will play a fundamental role in the future of GIS and ASIS. data models must represent both the structure of the datasets as well as how they will be used.
The following section briefly covers these topics with respect to GIS and ASIS. First traditional and new data models technologies and environments are considered. Then, common solutions for managing and visualising GIS and AS data are introduced. Finally the data model interoperability challenge is discussed.

# Data Model Technologies

## Highly-structured database models

For many decades, business applications for computers involved underlying database systems. Each of these systems tended to have its own model for the data. In this context, a data model is an unambiguous and neutral view of data that consists of a set of concepts used to organize the data -- describing its structure so it is comprehensible to the computer applications programs.

Traditionally, these data models have been studied in database research and development where two main classes of data models are distinguished: object-oriented and record-oriented models. In addition, there exist different abstraction levels at which data models can be introduced: the interoperability/application level, the conceptual/logical level, and the physical level. These three levels of data abstraction make it possible to achieve both logical and physical data independence.

“Standard” data models for databases are:

- Entity-Relationship model (conceptual model);
- Relational model (record-based model, SQL standard, logical models);
- Network and Hierarchical model (legacy models, physical models);
- Object-Oriented models (e.g., ODMG's object model).

## The WWW semistructured data model

The advent of the World Wide Web introduced a world of data and information that was not organized according to such precise and formal mechanisms. While the traditional data models are useful in dealing with structured data (e.g., data organized in formal databases), they are not as successful for representing the Semistructured Data (SSD) of the sort that is found on the web.

The SSD Web data has the following characteristics [Ramakrishnan]:

- It cannot be represented completely by formal schemas or types;
- It generally has an irregular structure and is heterogeneous and deeply nested;
- Its structure evolves -- often without notice;
- It can only be accessed through limited capabilities;
- It has links among datasets stored on distributed servers.

As the success of the Web has shown, SSD provides a convenient and flexible format for exchanging and querying data and information. Moreover, SSD arises in important application domains such as scientific data collections and digital libraries where there is a strong demand for information and data integration in spite of the fact that the underlying data are not formally structured along the lines of the traditional database models. In the language of computer science, the data model for the Web’s SSD consists in a rooted, edge-labelled, directed graph (e.g. OEM, UnQL). In layman’s terms, this describes the connections established by the pointers among the many documents on the Web. If you take any one document and follow the pointers in that document, then in turn follow the pointers in those documents, you trace what starts out looking like a hierarchy of documents, but is not a strict hierarchy because the links can circle back on themselves – thus the more general term directed graph is used. Clearly this sort of constantly evolving maze is not easy to represent in a formal database schema.

A very popular approach for integrating database technology into the Web consists in implementing a wrapper module which, on demand, generates SSD in the form of XML documents or HTML Web pages from the structured data managed by a traditional database. This approach introduces a

SSD model at the interoperability level which implements an export view of the conceptual/logical model of the database.

In the example of the Boulder map in the figure above, the various features specified in the legend are in fact stored in a relational database and the map is generated by the "Geography Network Explorer" which transforms the features into layers in a form viewable within a Web browser.

**Specifying a structure for data accessed via Web services**

While the specific modelling of data using formal database schemas is powerful in the context of targeted database computer applications, a more general form of data modelling -- independent of specific applications technologies – is needed to facilitate integration and interoperability among datasets and computer applications that access and use them in the context of the Web.

A powerful and standard technology for encoding and exchanging information among computer programs on the Web is XML. While it was originally defined as a textual language rather than a data model, it is now very popular for data representation and exchange of data among computer programs on the Web. However, while XML is an extremely versatile and valuable data transport format, experience has shown that -- despite high hopes for it -- XML is mediocre to poor as a data storage and access format.

(While not exactly a data model, a standard Document Object Model (DOM) for XML has been defined to enable XML to be manipulated by software. The DOM defines how to translate an XML document into data structures and thus can serve as a starting point for an XML data model.)

In general, the results of data integration, achieved through systems/applications interoperability, are semi-structured. For example, Web-services and e-business technologies use SSD in the form of XML documents.

# Geographic data management and visualisation

## Traditional GIS data in relational databases

In the realm of geolocated datasets, data stored in Geographic Information Systems (GIS) are highly structured and most often stored in an underlying relational database. While this may be a gross simplification, GIS datasets typically consist of "features" on the surface of the Earth that can be represented by points, lines and polygons. An example is a county plat which can show natural features such as streams and rivers, infrastructure like roads and bridges and buildings, and plots of land such as towns, lots, and so forth. The attributes of these features lend themselves to storage in the tables of a relational database. There can be a table for the roads, another for the towns, yet another for the rivers, etc. Each specific feature is a record in a table which provides a very useful way of keeping track of the characteristics of each instance of each feature.

Visualization is conceptualized in terms of a set of "layers." In the physical world, transparent mylar sheets are often used to overlay various sets of features on a given base map. The same idea is used for manipulating the visualization of the classes of features in GIS visualization systems

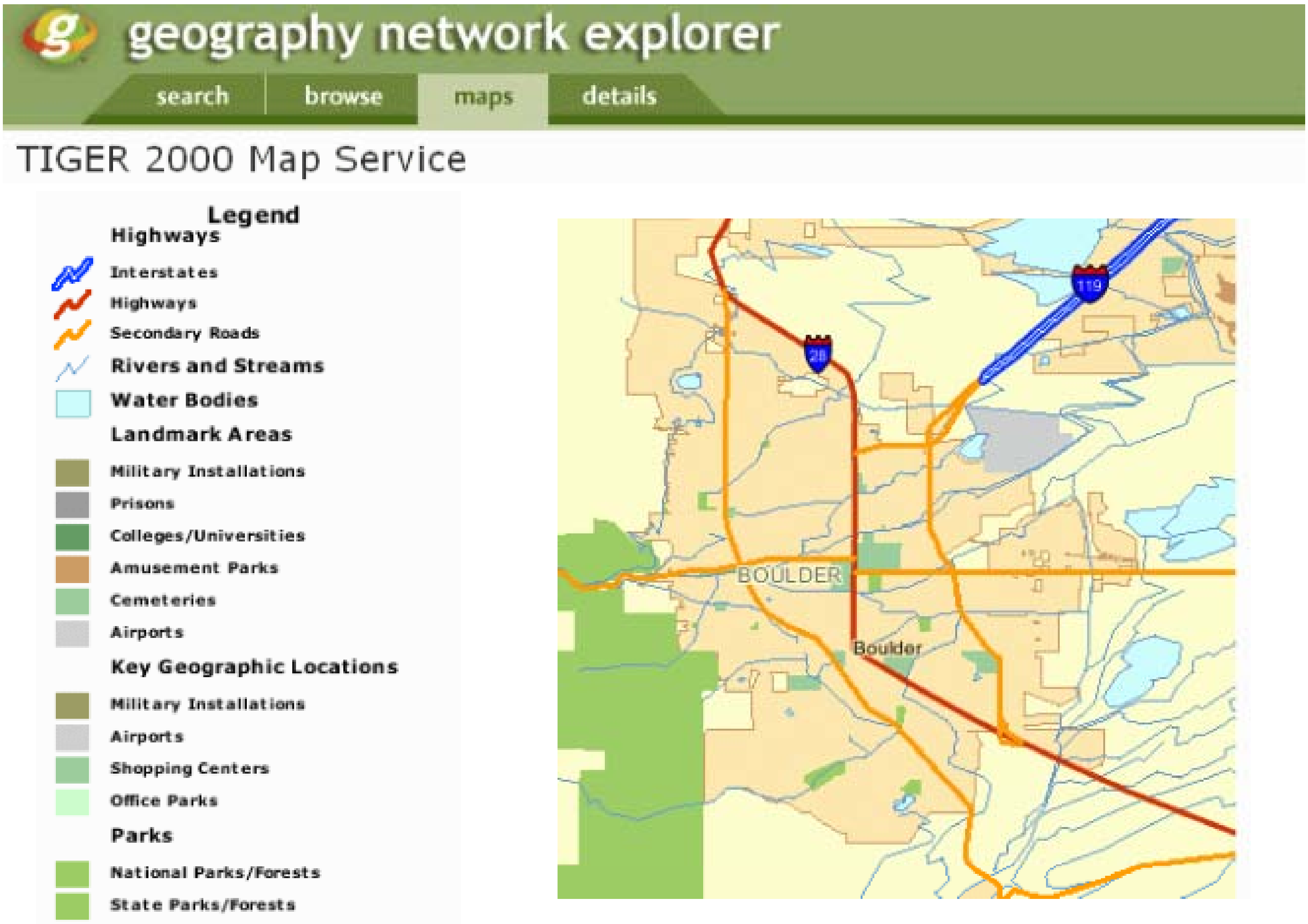


**Figure 7:** The example above is a map of Boulder, Colorado displayed from a Web site which is part of the GIS-based Geography Network. The legend on the left lists the various classes of features that can be displayed on such maps. In this particular case, the underlying GIS technology is an ESRI-based system called the Geography Network Explorer, but the data model of overlays of features on the surface of the Earth is common to all GIS-based systems. Nearly all such systems also have a relational database in which the features are stored according to a specific schema.

### Atmospheric science datasets

Among Atmospheric Science data, one of the most extreme cases is the voluminous datasets output by supercomputer forecast modelling programs. While these are not data in the sense of observations, they are among the most important datasets for the climate and weather forecasting communities.

Since these datasets are the results of highly-sophisticated numerical simulations run on powerful computers, they are highly structured. But their structure is quite different from GIS layers one, and cannot be successfully modeled by the formal database structures discussed earlier. And the scientist's conceptual understanding of the datasets differs dramatically from the typical conceptual model associated with the traditional GIS collections. In fact, the forecast modeling community does not think of data in terms of features on the surface of the Earth but rather the data are discrete points in a continuous function space where many parameters (e.g., temperature, pressure, wind speed and direction) vary in three spatial dimensions and time. This is a natural viewpoint because the datasets themselves are generated by numerical solutions to the complicated (e.g., the Navier Stokes) equations of fluid dynamics.

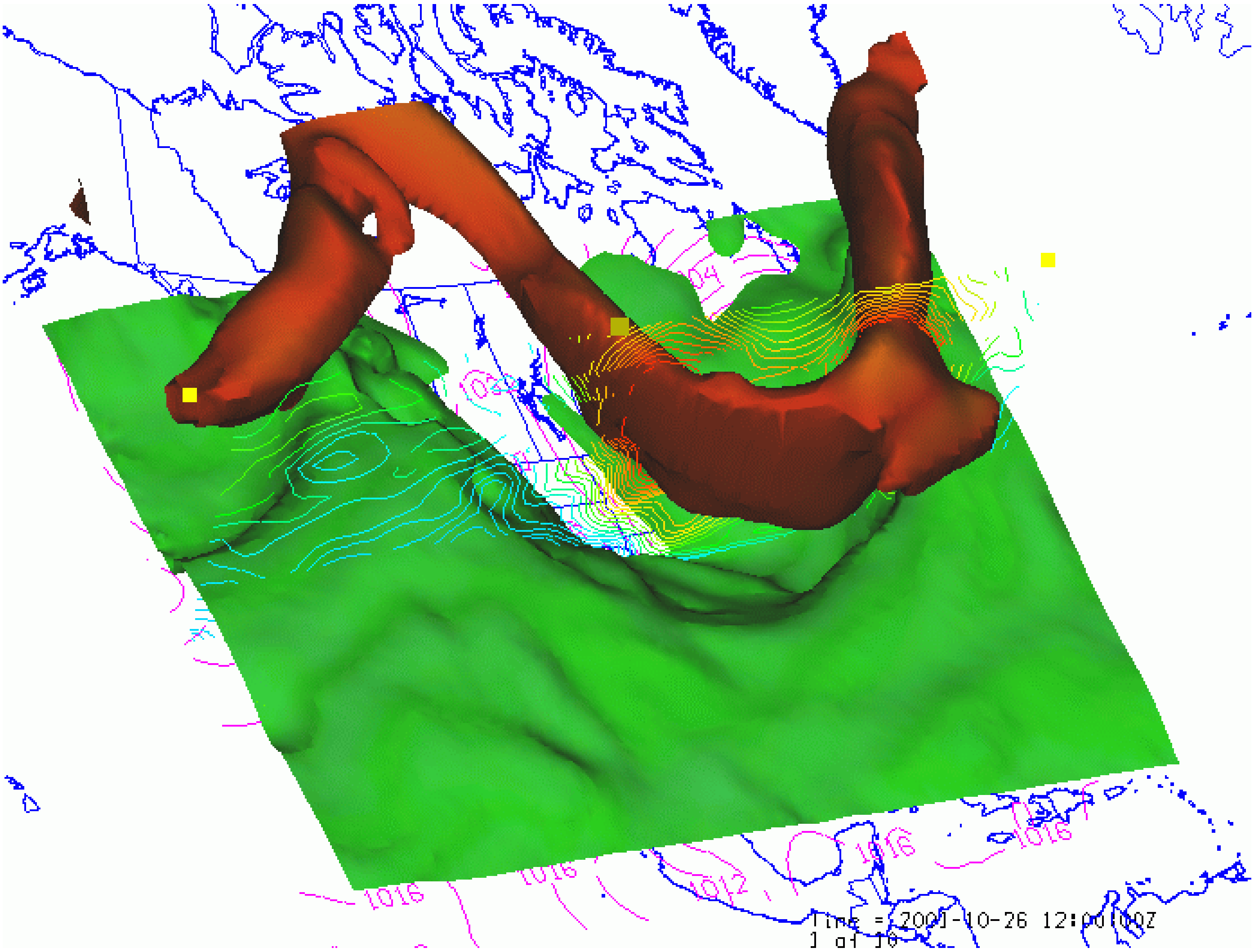


**FIGURE 8:** The visualization of the jet stream development predicted in the output of a numerical forecast model shown above illustrates many of the special characteristics of this type of data. It shows many atmospheric parameters varying not only in three spatial dimensions, but in time as well. Moreover there is more than one time dimension in that these the forecasts themselves are generated at regular time intervals and each such forecast has an associated forecast time scale -- which of course extends into the future if it is to be of practical use. In this case, the underlying maps are actually generated from GIS shape files. This illustrates integration of GIS data into a visualization of data characteristic of the atmospheric sciences.

(Note that forecast model output datasets are only one of many classes of data that could be chosen to illustrate the challenge of data interoperability in the Earth sciences. Among the others, are satellite images, data from radars, balloon- and aircraft-borne soundings, data from individual weather observing stations, swath data from polar orbiting satellites and from shipboard observation systems. For this discussion, the forecast model output has been chosen because, in many ways, the conceptual models are the most distinct from those of the traditional database community.)

These data are structured but are more complex than traditional structured data; as a matter of fact, they can be classified as hyperspatial data, or Hyperspatial Structured Data (HSD).
Generally speaking, AS datasets are characterised by the following observational nature:

- high dimensionality (e.g. four or more dimensions);

- multivalues parameters, characterised by a tensor rank (e.g. a brightness temperature parameter can be characterised by several values, one for each acquisition channel; its rank is equal to the number of acquisition channels);

Functional mappings are defined and used by scientists to map parameters elements and then visualise them. For instance, a four dimensional parameter of rank three is characterised by $4^3$ quantities [Treinish], which could be visualised and cross analysed.

Another important aspect -especially for GIS interoperability- is the base geometry used to map dataset parameters to earth coordinate systems. These geometries are generally gridded based, but also ungridded ones are possible. Among gridded geometries, regular, deformed and irregular grids are present. These grids may be explicitly or implicitly positioned. Topological relationships among grid cells are also important.
Data type are very heterogeneous, e.g. real, complex, integer, etc. Useful data aggregations are often utilised: grid cells patterns, multi-grids, time series, fiber bundles, dataset hierarchical trees [Treinish].

Traditionally, scientific communities have been using file systems to manage HSD, instead of DBMS. Scientific communities use DIFs (Data Interchange Formats), such as: CDF, HDF, CIF etc. to store and manage HSD, though such formats were intended for data exchange. Several experts still consider such approach more appropriate. As a matter of fact, several GIS are able to directly "interface" file systems based on common DIFs.

A research subject for Database science is the utilisation of traditional database models (i.e. relational, network and hierarchical) to model HSD, achieving database schemas which can be directly used by GIS. Naturally, such technology would be extremely useful to store, query and manage HSD in a GIS framework, as well as in any other application context.

Another promising research subject is related to modeling HSD (managed by either DBMS or file systems) by means of rooted, edge-labelled, directed graphs (e.g. encoding HSD in XML). Such point is important to work out interoperability models and, therefore, enable applications interoperability in the field of Atmospheric Science. Eventually, it is of paramount importance to generate web data (e.g. "documents") from HSD, and publish them through DL applications.

# The challenge: data models interoperability

As the technology of web services accessible by computer programs evolves, the challenge for those studying the Earth from an interdisciplinary perspective is to develop interoperable data models that can span the specific data models employed in individual disciplines. Moreover, these interoperable models have to be integrated with the semistructured framework of the Web itself. Only in this way will it be possible to develop visualization applications that afford the user an integrated view of datasets from different disciplines overlaid on societal and infrastructure impacts information from traditional GIS databases. Furthermore, it is imperative for the systems to evolve in such a way that the datasets themselves can be embedded into that semistructured but extremely accessible and useful directed graph of documents on the Web.

### Display integration

Interoperability among datasets can be achieved at several levels. A basic level of interoperability will allow two datasets to be visualized in a common view. The results of this level of interoperability can be seen in the two integrated images that follow. The first image -- generated using GEMPAK (GEneral Meteorological Packag) shows atmospheric science radar data overlaid with infrastructure information imported from GIS shapefiles.

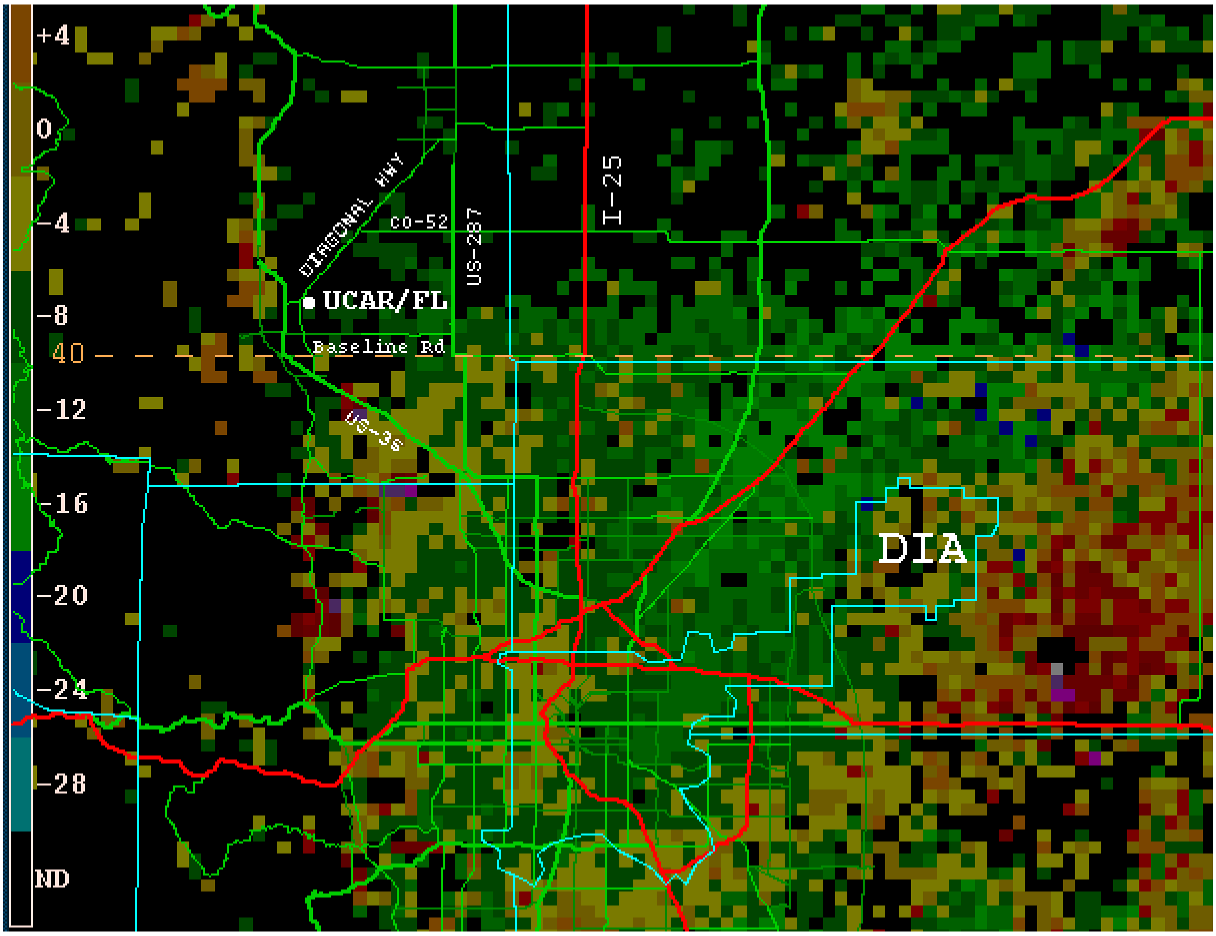


**FIGURE 9**: In this visualization, both the radar data and the GIS infrastructure data appear as if they are "features" on the surface of the Earth. In this case, the visualization was actually created by GEMPAK, an atmospheric science analysis and display application. But a similar display could be created by a GIS tool because the data are in a common form.

The following image shows integration in the other direction. In this case a slice of the meteorological forecast model output has been transformed so that it can be displayed as a layer in the ESRI ArcMap display tool.

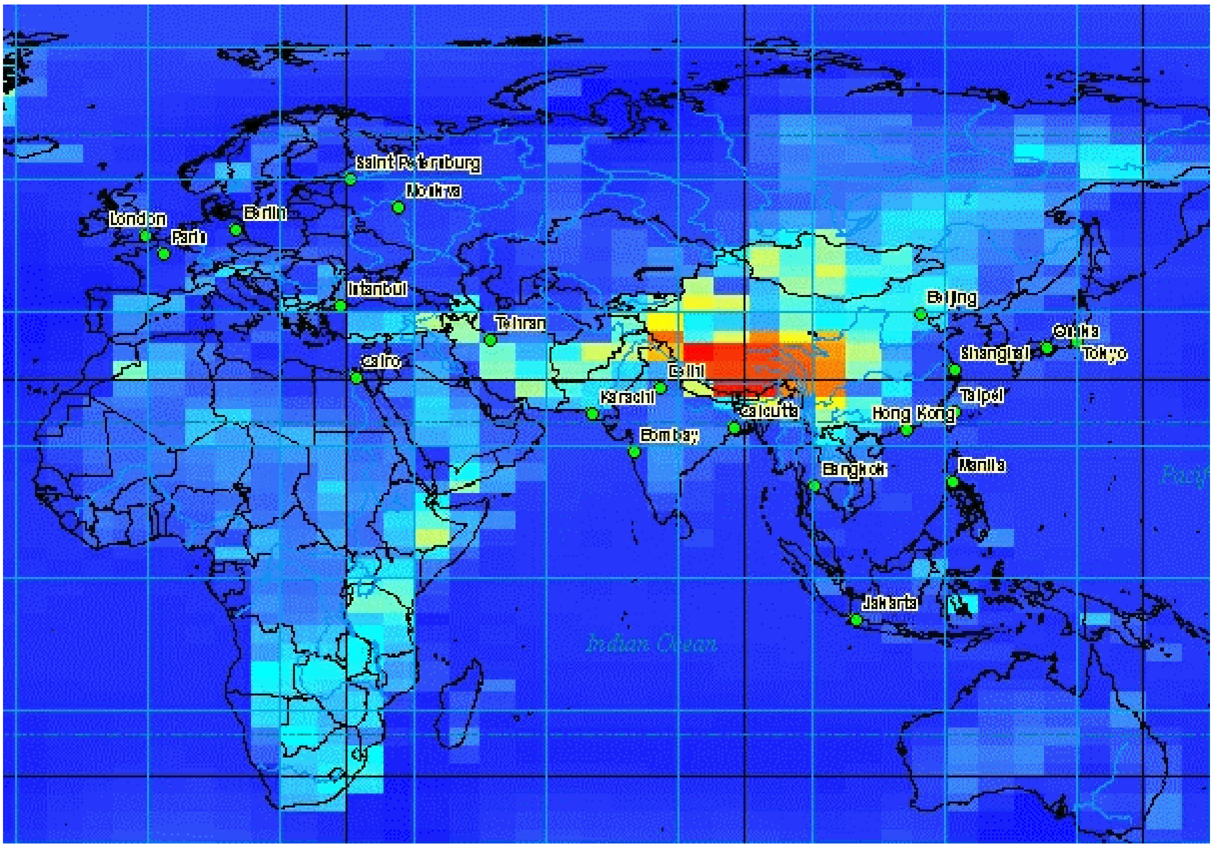


**FIGURE 10**: This image shows the output of a numerical weather forecast model (the AVN or Aviation) model visualized within the ESRI Arcmap GIS application. In this case, the integrated view was made possible by a transformation program that takes a horizontal slice of the model output and converts it into a geoTIFF form that can be displayed as a layer on the underlying GIS map.

## Application analysis interoperability

### GIS data analysis

However, integration at the display level does not ensure interoperability for data analysis. The image below illustrates the type of analysis that can be done if all the data are integrated into a GIS structured database form. In this case, the output of a "slosh" model forecasts the storm tide associated a hurricane. The storm surge heights are integrated into the GIS system as features on the Earth's surface. Combined in the database with information about the number of schools in the affected region, one can envision how questions about which counties are forecast to have a storm surge of 5 feet or greater. Beyond that, a decision maker can query the database to learn how many schools would be affected if those counties closed their schools. With the structured database view of data, these inquiries are quite natural.

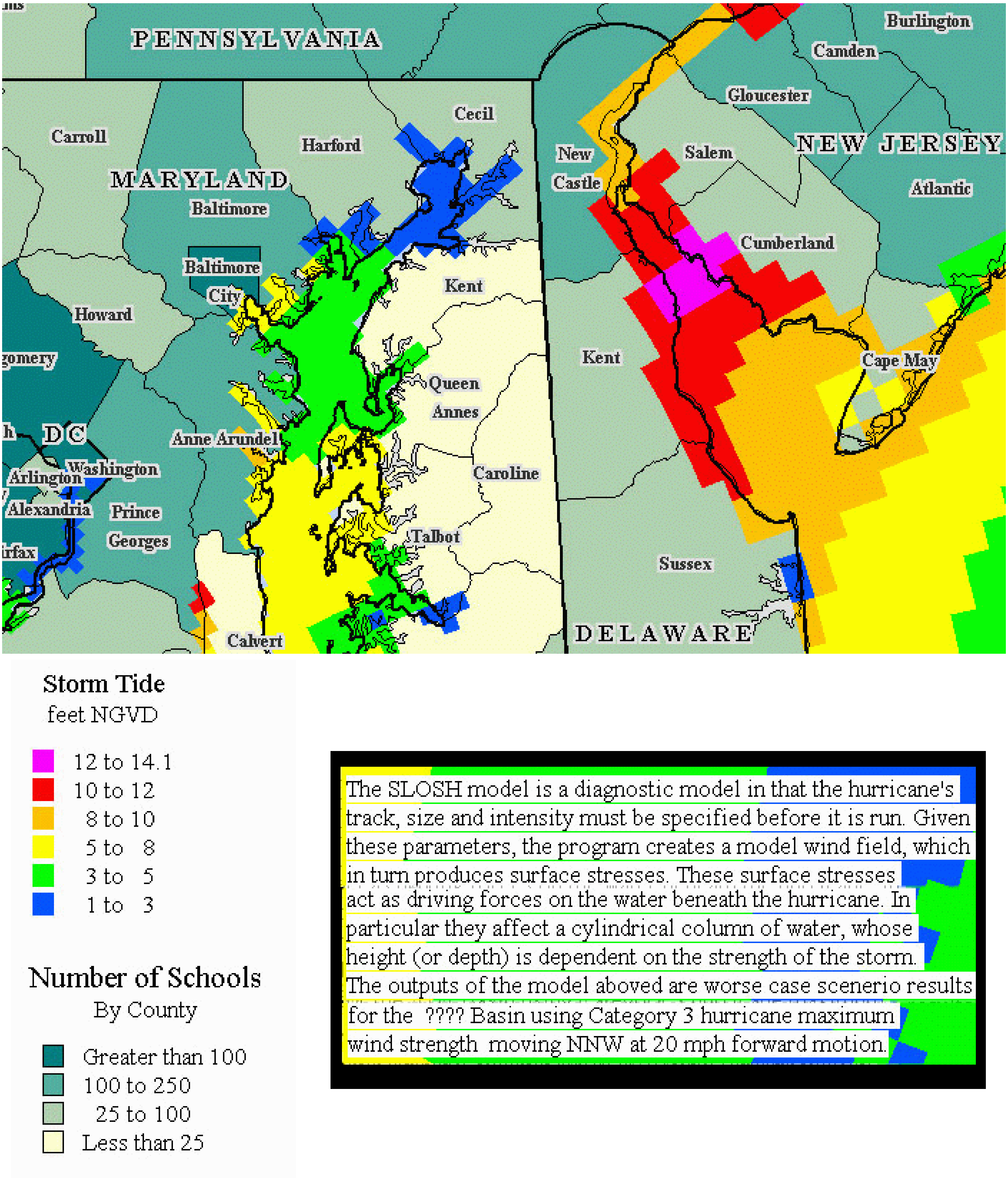


**FIGURE 11**: These images were taken from a FEMA Mapping and Analysis Center web page at: http://www.gismaps.fema.gov/2003graphics/storms/isabel/schools.jpg

**Atmospheric science analysis**

The images in the figure below illustrate a type of analysis possible using the tools of the atmospheric sciences where data are seen as discrete points in a multi-variate function space that varies in three spatial dimensions and time. In this case, the Vis5D analysis tool has been used to trace the trajectories of individual air parcels in time. With the function space data model, these foreward and backward trajectory calculations are quite natural, as is the vertical cross-section shown in the image on the left.

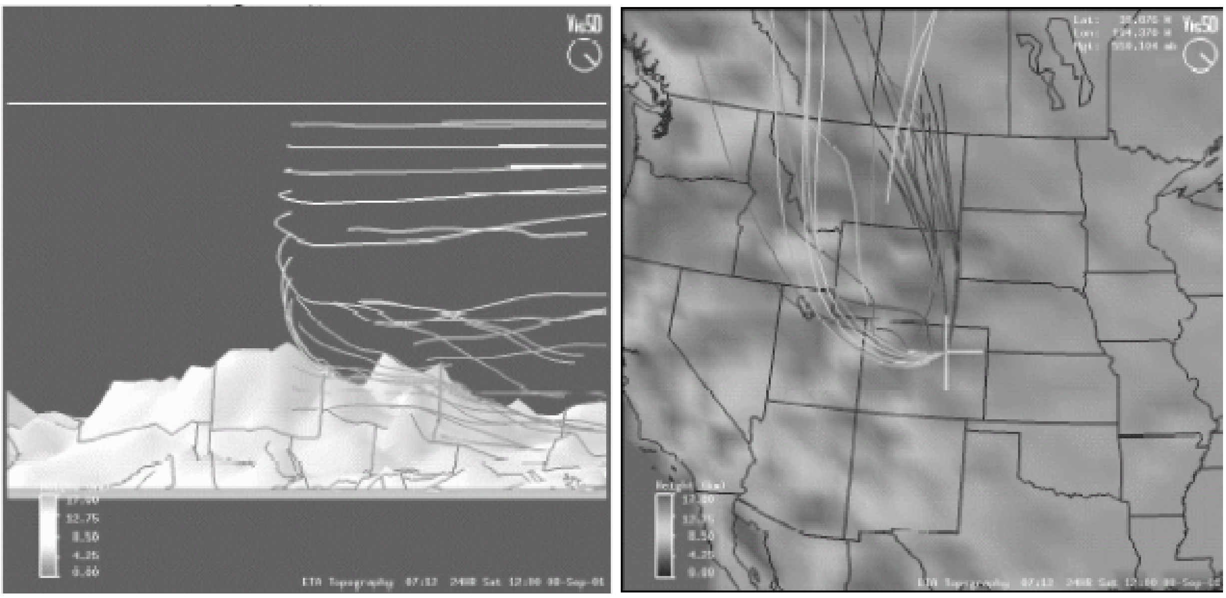

**FIGURE 12**: Visualization of ETA forecast model output showing computed air parcel trajectories viewed in vertical (left) and horizontal (right) cross-sections. These trajectories can be calculated because, as noted earlier, the atmospheric sciences forecast output is a set of discrete points in a mathematical function space. The data can be overlaid on a display that includes GIS-type map data (as in the illustration on the right), but the vertical cross section on the left is not straightforward in most GIS systems.

## Conclusions and future directions

Traditional GIS data models are fit for managing and visualising feature-based datasets, characterised by relatively slow time variations. They don't seem to be particularly fit for Atmospheric Science data visualisation or management.

As a matter of fact, it is possible to distinguish different data models which are more effective for different functionalities (or services); hence, for example, NetCDF and HDF5 data models are fit for data storage and exchange; VisAD data model for dataset visualisation, etc.

New kind of services (functionalities) are getting more and more important: interoperability services. They require data models suited for enabling web service to "understand" and easily exchange Atmospheric Science datasets. These data models must be particularly accurate on metadata and encoding model, which enable the effective sharing of data content and meaning, and hence the real system interoperability. For example, ESRI's ArcXML, OpenGIS GML (Geography Markup Language), ESML and NcML/NcML-G are examples of encoding language, based on a semistructured data model, for enabling ASIS and GIS interoperability.

In our vision, such data models are extremely useful to achieve applications interoperability (in particular GIS and ASIS interoperability). They are useful to interconnect sibling services implemented on heterogeneous applications. Besides, they are useful to interconnect services/functionalities at different levels -e.g. implemented in a distributed service framework- which require diverse and specialised data models; for example, data management, exchanging and visualisation services.

Leveraging such middleware, it could be possible to achieve a distributed service framework which uses:

- VisAD model for datasets visualisation services;
- HDF 5.0 or NetCDF with extensions for dataset exchange and storage;
- WCS 1.0 and GML 3.0 for GIS interoperability.

# References

Bill Hibbard "VisAD: Connecting People to Computations and People to People", 1998, http://citeseer.nj.nec.com/cache/papers/cs/2512/http:zSzzSzwww.ssec.wisc.eduzSz~billhzSzcompgrap.pdf/hibbard98visad.pdf

Russ Rew, Glenn Davis, Steve Emmerson, and Harvey Davies, "NetCDF User's Guide
An Interface for Data Access Version 2.4", Unidata Program Center, 1996, http://www.rzg.mpg.de/~dpc/netcdf_guide/guide_toc.html

ISO/WD 19123.3.3, "Geographic information — Coverage geometry and functions", ISO TC 211/WG2/N138, 2000.

ISO/TC 211/WG 1/WI 19129, "Geographic information - Imagery, gridded and coverage data framework", ISO/TC 211 N 1176, 2001.

ISO/WD 4 19130, "Geographic information – Sensor and data models for imagery and gridded data", ISO TC 211, 2003.

OpenGIS Consortium, "The OpenGIS™ Abstract Specification Topic 6: The Coverage Type and its Subtypes Version 4", 1999.

L.A. Treinish, "Scientific Data Models for Large-Scale Applications", IBM T.J. Watson Research Center, New York,
http://www.research.ibm.com/people/l/lloydt/dm/home.htm

R. Ramakrishnan, "Database Management Systems: Introduction to Semistructured Data and XML", http://www.cs.wisc.edu/~sekar/1

D. Barjenbruch, Operational Applications of Three Dimensional Air Parcel Trajectories using AWIPS D3D,
Interactive Symposium on AWIPS 18th International Conference on IIPS, American Meteorological Society Annual Meeting, January 2002.