International Hydrographic Review, Monaco, LXXII(2), September 1995

A METHOD FOR THE ASSESSMENT OF OBJECT-BASED DATA BY MEANS OF A REFERENCE MODEL

by Jan HERBERG [1]

1. Abstract

This contribution provides an overview of a method developed at the Institute of Ship Operation, Maritime Transport and Simulation (ISSUS) for checking the consistency of object-based data by means of a reference model. The method was designed in the course of the research project "BANET - Baltic and North Sea ECDIS Testbed" and was used to check object-based hydrographic data. In the following, the principles and application of the method are described.

2. THE PROBLEM

With the increasing use of object-oriented modelling techniques in many fields of technology, the demand for suitable test methods for object-based data structures is becoming ever more prominent. This applies in particular to large, complex-structured data quantities such as those occurring in modern geo-information systems. On the one hand, such quantities of data can no longer be checked manually, and on the other hand, the object-oriented approach offers the possibility of performing comprehensive consistency checks, which were not capable of being used on conventional data structures. In the case considered here, the distinct safety-relevance of the data to be checked was a further reason for the development described in the following.

Within the framework of BANET, there arose the specific question of checking object-based hydrographic (chart) data. The background of the project is the desire - which has been growing for several years - to replace the conventional paper chart on the ship by a special geographic information system (ECDIS - Electronic Chart Display and Information System). ECDIS integrates the various

[1] ISSUS, Elbchaussee 23, 22765 Hamburg, Germany.

sensor values of the ship, such as the values from the position sensor, gyro compass and radar display, together with the digitised chart data within a single system, thus providing the officer of the watch with all of the information needed for track control. One of the main components of the system consists of the underlying chart data, which are stored in object-based form and permit detailed requests for information, e.g. about water depths, sea marks or lights.

The data stored in ECDIS contribute directly to the situation assessment and decision-making performed by the officer of the watch; the quality of the data therefore has a direct influence on the ship's safety. The aim was to provide a method that would make it possible to check the logical consistency of ECDIS data, so that objects defined incorrectly could be recognised and eliminated.

3. THE APPROACH TO A SOLUTION

The method presented is based on using a reference model in the form of an *entity relationship model* [1]. The reference model defines the permissible *relations* between *objects* of two *classes*. The data to be checked are first analysed and then compared with the reference model. As a result, it is possible to obtain an assessment of the data with regard to the validity of their content. For example, in this way, incorrectly positioned sea marks (perhaps buoys situated on land or lights without holders) can be detected.

In the following, the method of putting this solution approach into action will be described. The terminology used here is largely the same as that now universally employed in the literature. However, because the terms utilised in the field of object-oriented modelling are often interpreted in different ways, the expressions used here will now be explained again briefly.

4. TERMINOLOGY

The main feature of object-oriented modelling is the classification of information. A *class* is the definition of an *object*. A class defines all characteristics that an object of this class has or can have, in the form of a class name and one or more *attributes*. Each attribute is of a particular data type with a defined range of values. Each object is an individual *instance* of a particular class, and accordingly has the attributes that are defined in the object's class. Each attribute of an object has a specific value.

Between classes, relations can be defined; these definitions are called *relationships*. Relationships define logical relations between classes. Objects can have *relations* to each other when these relations have been defined between their classes. Between two classes, there can be a maximum of one relationship. An object can have relations to more than one other object.

The reference model described here defines the classes and their attributes, and the relations between different classes. Modelling of the behaviour of objects has been deliberately omitted because, for ECDIS at the present time, such behaviour is of no relevance for checking object-based data.

5. THE DATA STRUCTURE

The data to be checked in the BANET project conform to a data standard (S-57) [2] which is internationally agreed in hydrography, provides for an object-based data structure, and has been provided by the hydrographic services of various countries. Figure 1 gives a schematic overview of the basic object-oriented structure of the data by means of a depth area.

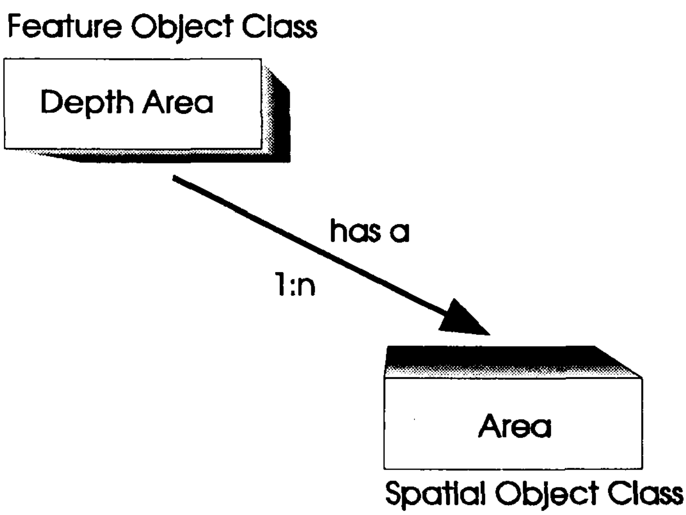

FIG. 1.- S-57 class definition and relationship (aggregation).

Every object conforming to the S-57 standard (called "S-57 object") consists of two partial objects - a *feature object* and one or more *spatial objects*. Each feature object belongs to a class defined in the standard, with corresponding attributes and attribute values, and has one or more spatial objects. A spatial object represents the spatial part of the object, and belongs to one of the spatial classes *point*, *line* or *area*. Figure 2 shows a schematic diagram of this situation.

An ECDIS data set consists of a number of such S-57 objects; the bulk of the data consists of the coordinates contained in the spatial objects. For the consistency check, it is sufficient to consider only the information about the spatial relations between the various spatial objects. In a first step, the spatial extents of the objects are analysed, and the resulting spatial relations between various spatial objects are used for the subsequent check of consistency. As an example, Figure 3 shows two feature objects which are overlaid spatially on each other, together with the associated spatial objects.

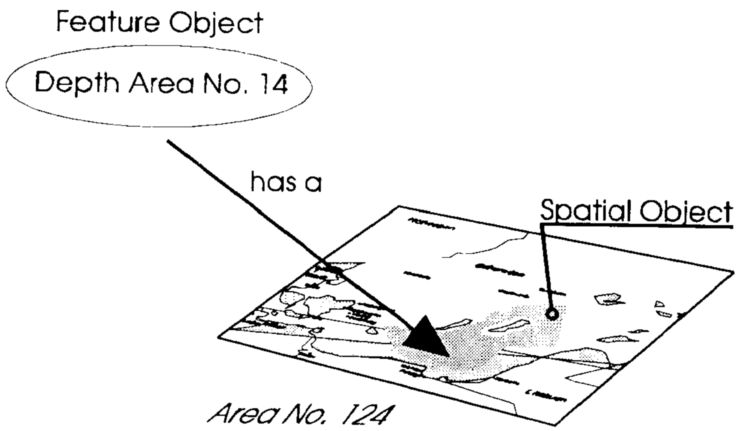

FIG. 2.- Feature object and spatial object.

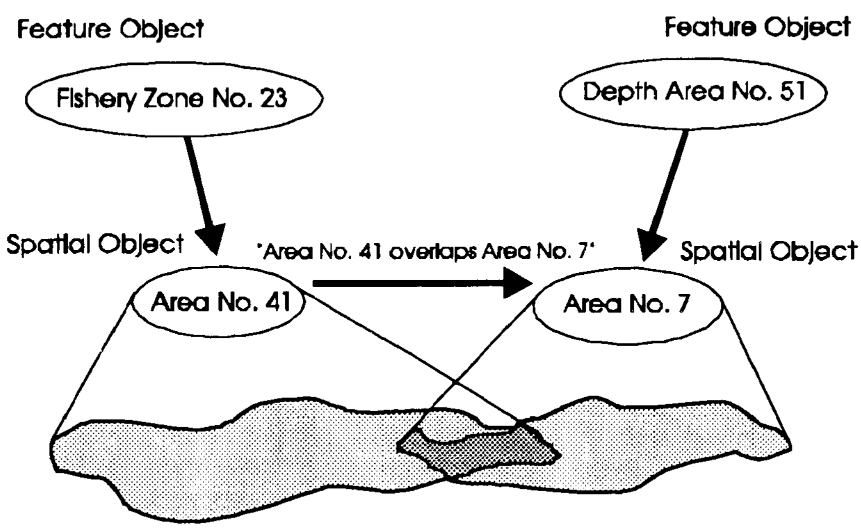

FIG. 3.- Spatial relation between spatial objects.

6. THE REFERENCE MODEL

The reference model defines the permissible classes and their attributes, as well as the relations between various feature objects. In the reference model, the relations which can exist between the various objects of particular classes are defined. In addition, a number of rules are established, containing logical links between several relationships and constituting an extension of the established constraint conditions. These rules make it possible to model more complex situations, such as the specifying of permissible relations depending on the existence of other relations.

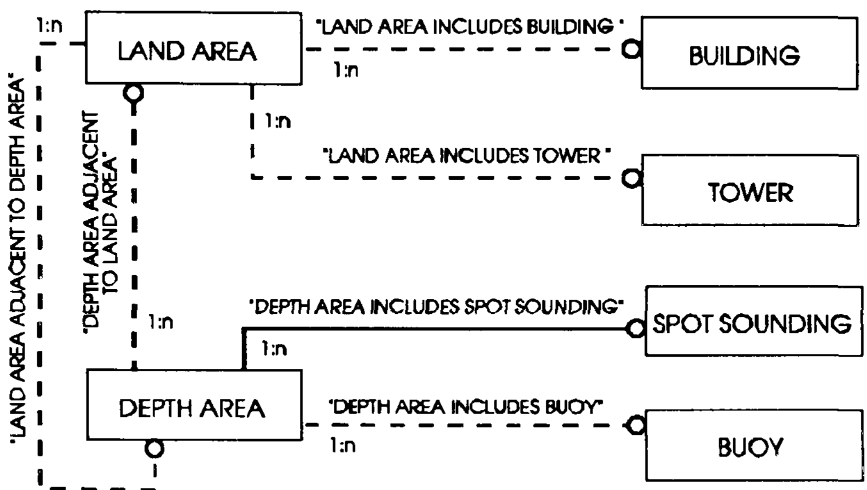

FIG. 4.- The reference model (example).

As an example, Figure 4 shows the structure of such a reference model. It is mainly an extended entity relationship model, such as is known in particular in connection with relational databases. Two classes in each case (represented by the rectangular boxes) can be connected by a relationship (represented by connecting lines). Each relationship has additional constraints, which define the maximum number of relations that an object of a particular class can have to various objects of another class (cf. Section 4).

Furthermore, there are two different types of relationship: *optional* and *mandatory relationships*. In the case of an optional relationship, an appropriate relation between two objects is permissible, but does not necessarily have to exist. A mandatory relationship specifies that each object of the class concerned must have an appropriate relationship to at least one other object. A mandatory or optional relationship from class A to class B means implicitly that there must also be a relationship (at least an optional one) from class B to class A. In Figure 4, mandatory relationships are indicated by continuous lines and optional ones by dashed lines.

The relationships described in the model specify the logical connections between objects of various classes. The data, on the other hand, contain purely spatial relations (e.g. "area A overlaps with area B", "point C lies within area D", etc.) which cannot be directly compared with the definitions contained in the model. To be able to use the reference model for the consistency check, it is therefore necessary to create an additional reference catalog which uniquely assigns the relations defined in the model to the corresponding spatial relations. Via this catalog, spatial relations present in the data can then be compared with those defined in the model.

7. THE CONSISTENCY CHECK

In the consistency check, the objects contained in the data are compared with the classes, attributes and relationships defined in the reference model. In addition, the rules laid down there (see 7.3) are checked for fulfillment. From this, it is then possible to derive an assessment of the data sets to be checked.

The aim of the consistency check is to discover random or systematic errors made by the data producer during data production, insofar as such errors can be detected in the structure of the data.

7.1 Checking the Classes and Attributes

The check of the classes is the first step in the consistency check. Here, the objects are checked for valid class names, and information about faulty objects is output. Because conformity with the class conventions is a basic prerequisite for the subsequent processing, the data set must be excluded from further checking if errors occur.

Next, the attribute values of the individual objects are checked for permissibility and completeness. During this check, it is ensured that each object has all of the attributes mandatorily prescribed in the reference model, and that all attribute values are filled with valid values. Objects with faulty attribute values are detected and are output.

7.2 Checking the Relations

The check of the relations forms the core of the method presented here. The relations defined in the model are compared with the spatial relations present in the data, and any relations deviating from the model are indicated.

7.2.1 Check of Mandatory Relationships

As already mentioned, mandatory relationships define those relations that all objects of a particular class must have to objects of a different class. A check is now performed to see whether every object fulfills all of the conditions stipulated, i.e. whether each object has all relations that are specified as being mandatory. For a relation specified in the model as being mandatory, the corresponding spatial relations for it are first selected in the reference catalog. If there is an object whose spatial object does not have any of these spatial relations to any spatial object of an object of the second class concerned, this situation does not fulfil the special mandatory relationship and is detected as being faulty.

This check makes it possible, for example, to detect lights without associated holders (masts, buoys) or the (incorrect) positioning of spot soundings on land areas.

7.2.2 Check of Optional Relationships

Optional relationships can exist between two objects, but do not necessarily have to. For each spatial relation between two spatial objects that is contained in the data, a check is performed to see whether, for that spatial relation, there is a corresponding relationship between the two classes of the associated feature objects in the reference model. If such a relationship does not exist (optional or mandatory), there is an impermissible relation between two feature objects, and this is detected as a data error. At the same time, this check can detect missing relationships in the reference model.

7.3 Checking the Rules

The rules permit extended mandatory relationships depending on the existence of particular relations in the data. By means of these rules, several optional relationships can be linked together to form a mandatory condition. For example, one such rule might specify that "each light is situated on a tower, a beacon or a buoy," i.e. that every object of the "light" class has a particular relation to another object of the class "tower", "beacon" or "buoy". In addition, rules can also depend on the attribute values of particular objects. Thus, for example, neighbouring depth-areas must have adjacent depth values.

8. CHECKING THE REFERENCE MODEL

For the actual consistency check of the data, the consistency of the reference model is an essential prerequisite. Therefore, various basic principles were established, which are used to check the model and are intended to ensure the model's suitability for the comparison:

i) The names used in the relationships must correspond to the reference catalog.

ii) If a relation is defined from class A to class B, there must also be a relationship from class B to class A.

iii) Rules may only be dependent on "optional relationships".

iv) All relationships referenced by rules must also exist in the model.

9. IMPLEMENTATION

9.1 General Approach

The method described was implemented as a software solution in the course of the R&D project "BANET" and was used to check object-based data. In the following, there is an overview of the development results and their basic functional principles.

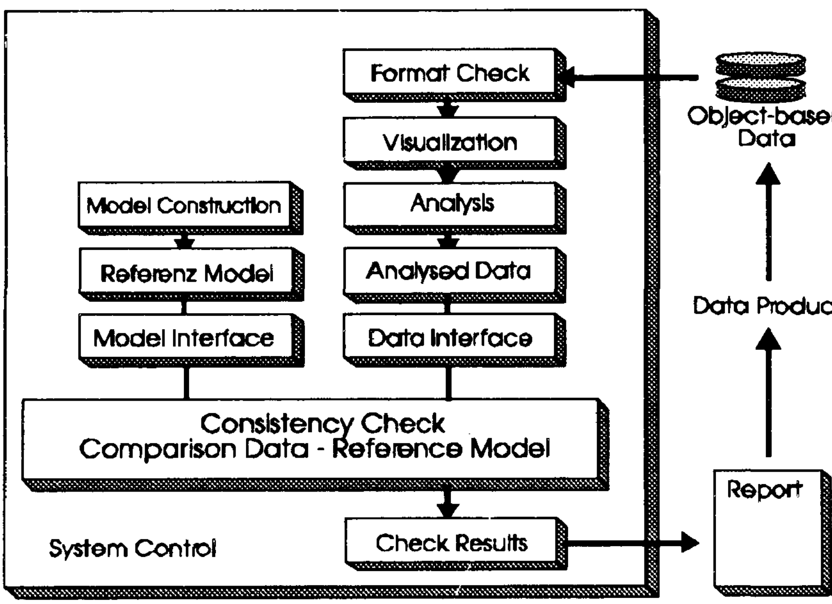

FIG. 5.- Implementation of the method.

Figure 5 shows a schematic diagram of the structure of the tool, with its components and interfaces. The main parts of the system are the model creation procedure, the analysis of the data, the actual check of consistency, and the evaluation of the check results. The data analysis is additionally preceded by a format check and by visualisation of the data; this visualisation makes it possible to check formal conditions (format, file structure). The consistency check accesses the reference model and the data via the model interface and the data interface, and passes the results on for evaluation.

9.2 Data Preprocessing

Before the comparison with the reference model, the data must be analysed to ascertain their spatial relations. In addition, the objects and attribute values contained in the data are needed. In a first step, the data provided in S-57 [2] are checked for correct format and are visualised, as a result of which it is already possible to perform a spot check. Furthermore, the information contained in the data sets is prepared for further processing. The subsequent data analysis compares the spatial objects with one another, and for the consistency check it provides the relations between them, together with the object names and attribute values.

For the preprocessing of the data, it was possible to use commercial software that already existed [3], so that objects, relations and attribute values that have already been completely analysed and prepared are made available to the system.

9.3 The Reference Model

The reference model was constructed in machine-readable form, and has about 200 class definitions with the associated attributes, and several thousand relationships. In addition, there is a comprehensive catalog of rules. To ensure that the model can also be used for other platforms and tools, the relationships were modelled in EXPRESS. EXPRESS is an internationally standardised descriptive language for information, which permits both description of program functionality and the modelling of object-based structures [4]. For the development of the reference model, a tool obtainable on the market [5] was used, which - by means of a graphic man-machine interface - permits convenient and transparent creation of relationships and classes. Figure 6 shows an example of the graphic presentation of a relationship; in test box 1, the associated code from an EXPRESS file is shown.

9.4 Consistency Check of the Data

The consistency check comprises the method described in the preceding sections. The reference model and the analysed data are available to the test tool via the model interface and data interface. With the aid of the reference catalog, the data are interpreted and are compared with the definitions contained in the model. The test result is output in readable form and is available to the data producer for the purpose of checking.

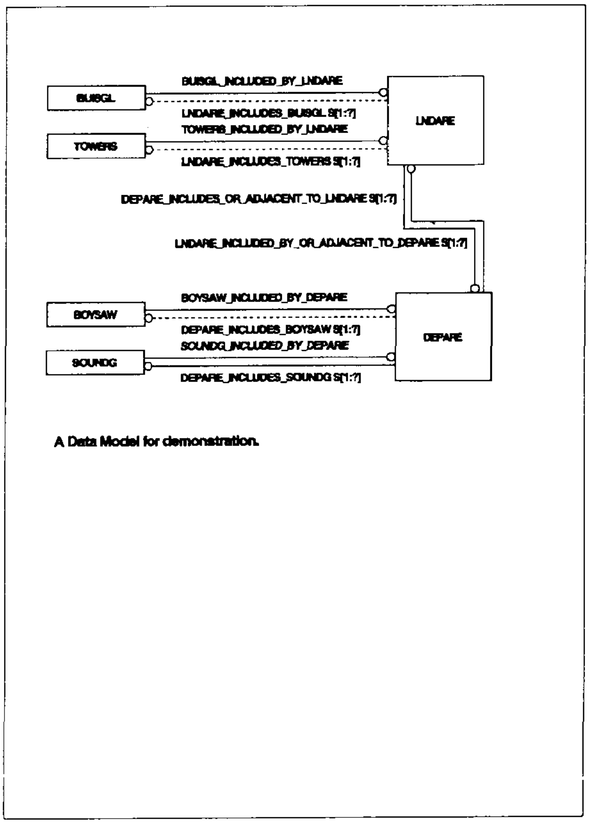

FIG. 6.- Entity-relationship model in EXPRESS.

9.5 Result of the Check

Text 2 shows an excerpt from a test record. The first message occurred during the check of the mandatory relationships. In the data, an anchorage area (ANCHARE) was found which does not overlap with any depth area (DEPARE) or is enclosed by a depth area. The second output indicates an infringement of a constraint of a relationship. For the spatial relation contained in the data between the objects FE__00001 of the "depth area" (DEPARE) class and FE__00025 of the "railway" (RAILWY) class, no corresponding relationship could be found in the reference model (3rd message). For the same depth area, a value for the mandatory attribute DRVAL2 is missing (4th message). The 5th message is a result of the check of rules. An object of the "lights" (LIGHTS) class was found which does not have any spatial relation to any object of the classes mentioned, and thus does not have a holder.

```
SCHEMA Schema_a;

 ENTITY LNDARE;
  LNDARE_INCLUDES_BUISGL : OPTIONAL SET [1:?] OF BUISGL;
  LNDARE_INCLUDES_TOWERS : OPTIONAL SET [1:?] OF TOWERS;
  LNDARE_INCLUDED_BY_OR_ADJACENT_TO_DEPARE : SET [1:?]
OF DEPARE;
 END_ENTITY;

 ENTITY DEPARE;
  DEPARE_INCLUDES_BOYSAW : OPTIONAL SET [1:?] OF BOYSAW;
  DEPARE_INCLUDES_SOUNDG : SET [1:?] OF SOUNDG;
  DEPARE_INCLUDES_OR_ADJACENT_TO_LNDARE : SET [1:?] OF
LNDARE;
 END_ENTITY;

 ENTITY BUISGL;
  BUISGL_INCLUDED_BY_LNDARE : LNDARE;
 END_ENTITY;

 ENTITY TOWERS;
  TOWERS_INCLUDED_BY_LNDARE : LNDARE;
 END_ENTITY;

 ENTITY BOYSAW;
  BOYSAW_INCLUDED_BY_DEPARE : DEPARE;
 END_ENTITY;

 ENTITY SOUNDG;
  SOUNDG_INCLUDED_BY_DEPARE : DEPARE;
 END_ENTITY;

END_SCHEMA;
```

Text 1 EXPRESS-code (example).

10. SUMMARY

The method described for the consistency check of object-based data structures has proven itself to be a useful aid in the checking and assessment of hydrographic data. As a result of the object-orientation, new quality-testing possibilities are obtained which are superior to conventional data structures and which, especially in safety-relevant applications such as in shiphandling, constitute

```
Feature object FE_0200003 does not meet mandatory relationship
ACHARE_INCLUDED_BY_OR_OVERLAPS_DEPARE.

For the relationship named SOUNDG_INCLUDED_BY_DEPARE at most
1 relations are allowed,
but 3 have been found for the feature object named FE_0200004.

No model relshp meets spat. rel. a_includes_l from FE___00001 (class
DEPARE)
to FE___00025 (class RAILWY).

Mand. Attr. DRVAL2 of feature obj. FE___00001 (class DEPARE)
missing.

Feature obj. FE___00211 (class LIGHTS) has no spatial relation to any
feature obj.
of the following classes: BOYSAW, TOWERS, MSTCOM, BCNSAW
```

Text 2 Check report (example).

an additional advantage. Due to the universal approach, application of the method in other branches of cartography too is certainly conceivable. Furthermore, use of an object-oriented reference model for consistency checking is definitely appropriate whenever large quantities of object-based data have to be checked.

The principles worked out in the course of the BANET project are being used by the industrial partners after completion of the project, and so it can be expected that, in the future, the method developed will be further refined and new checks will be added to it. This might include, for example, the introduction of spatially oriented relations or the checking of persistent objects (i.e. objects generated during the runtime of a program) in a system. Use as an on-line check during the production process is also conceivable.

One particular aspect in future will certainly be that of run-time optimisation, because in practical use at present the test procedures described can still take several hours per data set. By the utilisation of suitable, fast search processes in conjunction with higher computing power, a time saving can still be achieved here.

11. References

[1] International Hydrographic Organization, "IHO Transfer Standard for Digital Hydrographic Data (S-57)", Version 2.0, International Hydrographic Bureau, Monaco, 1993.

[2] EMBLEY, David W., "Object-oriented system analysis: a model-driven approach", W. EMBLEY, Barry D. KURTZ, Scott N. WOODFIELD, New Jersey, USA, 1992.

[3] BÜTTGENBACH, Gert, "Technische Dokumentation für PROOF S-57", SevenCs GmbH, Hamburg, Germany, 1994.

[4] ISO TC184/SC4/WG1, "EXPRESS Language Reference Manual", St. Louis, USA, 1990.

[5] GIDA GmbH, "Anwendungsbeschreibung zu EXEP/EGE", Version 2.0, Berlin, Germany, 1994.