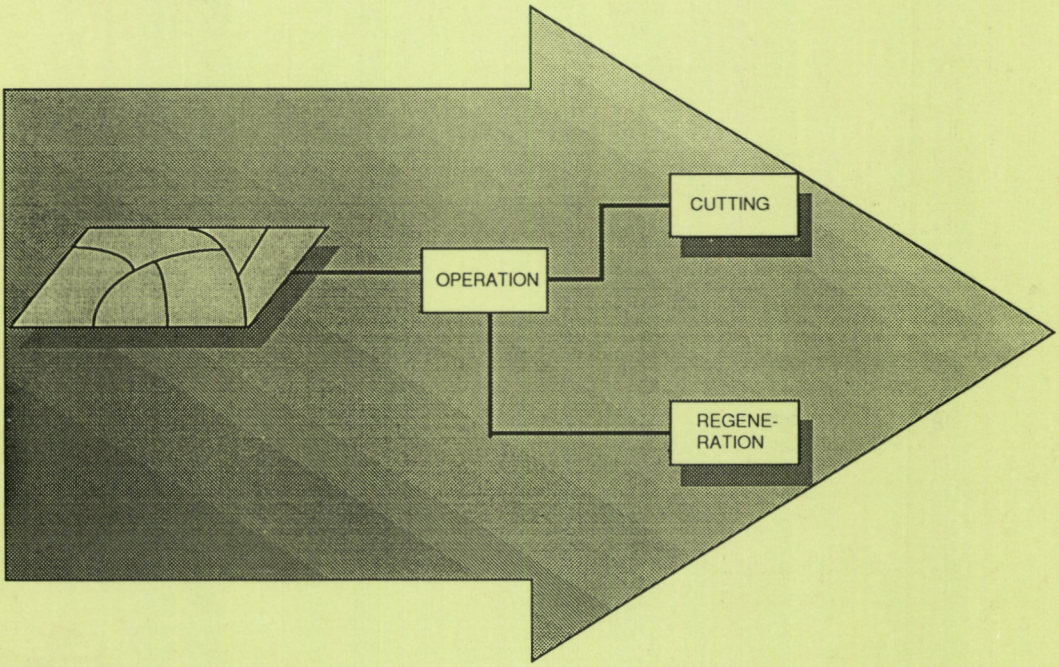
A DATA MODEL FOR A GIS-BASED FOREST INFORMATION SYSTEM

Raito Paananen



Metsäntutkimuslaitoksen tiedonantoja 493
The Finnish Forest Research Institute. Research Papers 493

METSÄNTUTKIMUSLAITOS
Kirjasto

A DATA MODEL FOR A GIS-BASED FOREST INFORMATION SYSTEM

Paikkatietojärjestelmään perustuvan metsätietojärjestelmän tietomalli

Raito Paananen

Finnish Forest Research Institute
Department of Forest Production

Metsäntutkimuslaitoksen tiedonantoja 493
The Finnish Forest Research Institute. Research Papers 493
Vantaa 1994

METSÄNTUTKIMUSLAITOS
Kirjasto

Paananen R. 1994. A data model for a GIS-based forest information system. Seloste: Paikkatietojärjestelmään perustuvan metsätietojärjestelmän tietomalli. Metsäntutkimuslaitoksen tiedonantoja 493. The Finnish Forest Research Institute. Research Papers 493. 54 + 24 p. ISBN 951-40-1357-3. ISSN 0358-4283

This study concerns with conceptual data modelling methods for planning geographic information systems for forest management. The study was made in the Finnish Forest Research Institute as part of a project called The Research Forest Database and Planning System. The project was started in 1991 to develop the information processing and planning of the research forests owned by the institute. The primary objective of the project was to define and develop a GIS-based forest information and planning system. In the system both up-to-date and history information of the forests can be integrated with the experiments and conservation areas for forest planning purposes.

This study was part of the project including the functional and information analysis and modelling for the new system. This report includes the conceptual data model description with an example of its implementation for handling the forest stand history information.

Tämä tutkimus on toteutettu osana TUTGIS-hanketta, jossa kehitetään Metsäntutkimuslaitoksen tutkimusmetsien operatiivista tietojärjestelmää. Hankkeen tavoitteena on tuottaa paikkatietojärjestelmään perustuva tutkimusmetsien tiedonhallinta- ja suunnittelujärjestelmä, jolla hallitaan metsävarojen nykytila ja mahdollistetaan metsien tutkimuskäytön integroitu suunnittelu. Järjestelmän päätoimintoja ovat paikkaan sidotun kuvioittaisen metsävaratiedon hallinta, metsäsuunnittelu ja kokeiden hallinta. Tässä tutkimuksessa tuotettiin järjestelmän perustan muodostavat loogiset mallit tutkimusmetsien toiminnoista (toimintomalli) ja tiedoista (käsitekaavio). Tässä raportissa kuvataan käsitekaavion rakenne ja sisältö sekä esimerkki toteutetusta metsikköhistorian hallintarakenteesta.

Työn tekemisen on tehnyt mahdolliseksi arvokas apu, jota ovat työn eri vaiheissa antaneet TUTGIS-hankkeen tutkijat Jussi Saramäki, Tuula Nuutinen, Kai Blauberg, Markku Juvakka, Aki Nalli, Jorma Nykänen ja Janne Soimasuo sekä Metsäntutkimuslaitoksen tutkimusalueyksikön henkilöstö.

Keywords: forest information system, data modelling, GIS

Publisher: The Finnish Forest Research Institute; Project: 303901-0. Accepted for publication by Professor Jari Parviainen, Research Director, in February 2, 1994.

Distribution: The Finnish Forest Research Institute, Department of Forest Production, P.O. Box 18, FIN-01301 Vantaa, Finland.

ISBN 951-40-1357-3
ISSN 0358-4283

Contents

Page
1 Information System development 4
1.1 Types of information 4
1.2 Types of information systems 4
1.3 IS development life cycle and methodologies 5
1.4 Data Models 6
1.5 Conceptual data modelling 8
1.5.1 Principles 8
1.5.2 ER model 8
1.5.3 Advanced data modelling concepts (the EER model) 9
1.5.4 ER/EER notations 10
1.6 Data modelling for GIS 12
1.7 Temporality in spatial databases 12
1.8 IS development in forestry 13
1.8.1 Data models 13
1.8.2 Temporality in forest information systems 14
1.8.3 Overview of the FFRI project 15
1.9 Aim of the study 16
2 Methods 16
2.1 Information Engineering 16
2.1.1 Overview of the methodology 16
2.1.2 Business Area Analysis 17
2.2 The modelling approach in this study 18
2.3 Application of the model 20
2.3.1 Inventory design 20
2.3.2 Test databases 21
3 Results 21
3.1 Research forest activities 21
3.2 ER schema of the research forests 22
3.2.1 General aspects 22
3.2.2 Spatial entities of basic mapping 23
3.2.3 Basic inventory data 25
3.2.4 Special land use data 27
3.2.5 Forest planning 30
3.2.6 Experiments 31
3.2.7 Forest stand history 32
3.2.7.1 Operation history 34
3.2.7.2 State history 35
3.2.8 Spatial entity types with interfaces to other systems 37
3.3 An application: Stand history data structures 37
3.3.1 Operation history data structures and operations 37
3.3.2 State history data structures 43
4 Discussion 46
4.1 Conceptual modelling 46
4.2 Forest stand history 47
4.3 Forest management 47
References 50
Seloste 53
Appendices

1 Information System development

1.1 Types of information

The two basic concepts in information processing are data and information. Davis & Olson (1985) define data as symbols which represent, describe or record reality. Data symbols are not the same as in reality. Information (datalogical information) is defined by Davis & Olson (1985) as data that has been processed into a form that is meaningful to the recipient and is of real or perceived value in current actions or decisions. Knowledge can be seen as a special type of datalogical information, typically complex and variable information about a specific area of human activities (Virtanen 1989).

Forest resource information is geographical in nature. Geographical data are referenced to locations on the earths surface by using coordinate systems. Geographic features are things that can be recognized in a map, e.g. road, river or lake. They have a location and some common descriptive information. Lehan (1986) defines feature as a physical entity that is recognized in the mans definition of reality.

Spatial objects are digital representations of geographic features representing the location, geometry and topology of the feature. Location is usually defined with some specific coordinate system. Geometry refers to the dimensions and shape of the objects. Topology represents the relationships between connecting or adjacent spatial objects. The primitive spatial objects are one-dimensional points, two-dimensional lines and two-dimensional polygons. All spatial data can be reduced into these three basic primitives. A map is a set of points, lines and polygons that are defined both by their locations in space with a reference to a coordinate system and by their non-spatial descriptive attributes (Burrough 1986).

1.2 Types of information systems

Information system (IS) is a collection of activities that regulate the sharing and distribution of information and the storage of data that are relevant to the organization (Batini et al. 1992). Management information system (MIS) can be defined as an integrated system for providing information to support operations, management and decision-making functions in an organization (Davis & Olson 1985). MIS uses procedural logic to manipulate data. Decision support system (DSS) is a computer-based information system used to support decision making activities in situations where it is not possible or not desirable to have an automated system to perform the entire process (Ginzberg & Stohr 1982). Expert system (ES) is a computerized advisory program that attempts to imitate or substitute the reasoning processes and knowledge of experts in solving specific types of problems (Turban 1990). Usually the term knowledge base system (KB) is used interchangeably with ES.

All types of information systems may be based on databases. A database is any large collection of structured data stored in a computer supporting shared access of many users. Database management system (DBMS) is a collection of software for managing

a database. Database together with its management software comprise a database system (Elmasri & Navathe 1989).

The role of information systems in forestry organizations has been discussed in Kaila & Saarenmaa (1990). An integrated forest information system can be seen as a MIS that brings together the information important to the management of the forest.

Geographic information system (GIS) may be seen as a database system in which the data are spatially indexed, and upon which a set of procedures operates in order to get answer queries about geographic features in the database (Smith et al. 1987). In GIS the spatial information is stored in information layers. Information layer is a digital overlay of uniformly attributed spatial data (Langran 1992b). Coverage is a commonly used synonym for layer. A GIS typically consists of five major components (Burrough 1986):

1. Data input and verification, which contains all the various operations to transform data from existing maps etc. to digital form.
2. Data storage and database management, which concern the way how data is stored, structured and organized and how they are perceived by the users.
3. Data output and presentation concerns the ways how data is displayed and the results of the analysis are reported to the users.
4. Data transformation includes both data manipulations and spatial analyses. The former is concerned e.g. with error corrections or changes to the form of the data while analyses are made in order to get answers to specific questions.
5. Interaction with the user (user interface).

There can be seen three main stages in the development of GIS systems (Sijainninhallinnan... 1991, McLaren 1990). 1. The eldest structure is to store the spatial (locational) information in its own file management system. The attribute data are stored in a separate database system. The two systems are not integrated. 2. In the next stage, the two separate systems are connected by using an interface from the GIS to the database system with logical connections based on common identifiers. The interface can be implemented with subprogram libraries or with SQL-queries. Data can be widely accessed for various needs, but the integrity between spatial and attribute information requires specific mechanisms. 3. In the third stage both the spatial information and its attributes are stored in the same integrated database system.

1.3 IS development life cycle and methodologies

Information systems development is a change process taken with respect to object systems in a set of environments by a development group to achieve or maintain some objectives (Lyytinen 1987). Information systems development life cycle can be defined as the life span beginning with the idea that a system is needed ending in the discard of the system (Connors 1992). The classical life cycle of IS development project is called the waterfall life cycle. It consists of five main phases, requirement specifications, system analysis, system design, implementation (coding, testing) and integration (operation). A typical characteristic is the linear, sequential progression from one phase to the next. This classical IS development life cycle has been

discussed e.g. in Batini et al. (1992), Connors (1992), Loomis (1990) and Yourdon (1989).

Requirement collection and system analysis are concerned with the so-called mission of the system, i.e the application areas and the problems that the system should solve. These phases are carried out in interaction with the users.

Design is concerned with the specification of the structure of the information system. Design can be divided into database design and application design. Database design is a complex process that involves several decisions at different levels. Database design is normally decomposed into conceptual, logical and physical design (Batini et al. 1992). These activities are called data modelling processes.

Implementation includes the programming of the operational version of the system. Recently prototyping tools are being utilized prior to the final implementation to make simplified versions of the system to verify the needs of the users. Validation and testing are made to assure system quality and to verify that the implementation reflects the design specifications. Operation starts with the initial loading of data and ends when the system is replaced.

Information systems development methodology is an organized collection of concepts, beliefs, values and normative principles supported by material resources. The purpose of the methodology is to help a development group successfully change object systems, to perceive, generate, assess, control, and to carry out changing actions in them (Lyytinen 1987).

System development methodologies combine tools and techniques to guide the development process. While the life cycle gives a measure of project control, methodologies provide tools to improve the productivity and quality of the system analysis and design. There are various approaches to systems development. Three main categories currently in use are structured analysis and design (SA), information engineering (IE) and object-oriented analysis and design (OOA,OOD). For a brief comparison of the methodologies, see Fichman & Kemerer (1992). In structured analysis methodologies the emphasis lies on the modelling of processes. For a typical SA presentation, see Yourdon (1989). Information engineering is a comprehensive methodology that extend the data-oriented approach to the entire development life cycle. IE is developed by James Martin (see e.g. Martin 1990). The OOA methodologies also rely on information modelling but they encapsulate data and behaviour: all processes are encapsulated within objects (Fichman & Kemerer 1992).

1.4 Data Models

Data modelling is an activity where a data model is applied to derive a logical organization of data that is documented in a schema (Klein & Hirschheim 1987). A data model is respectively a way of perceiving, organising and describing data. Elmasri & Navathe (1989) define data model as a set of concepts that can be used to describe the structure of a database. Shlaer & Mellor (1988) consider data model (or as they name it information model) to be a thinking tool used to aid in the formalization

of knowledge. Tsichritzis & Lochovsky (1982) state that data models enable us to capture the meaning of data as related to the meaning of the world in an appropriate amount which is adequate for the desired use of the data. Data models define general rules for the specification of the structures of the data and also of the operations that are allowed on the data.

Data models can be categorized according to the level of abstraction used. Usually three levels are distinguished (Elmasri & Navathe 1989, Hull & King 1987):

1. Conceptual models. These are the most high-level models. The term semantic data model is also commonly used. Conceptual models describe the logical structure of the data for a community of users. Concepts that are used are common in the language used in the problem domain. A specific term, Universe of Discourse, is used in this context (Klein & Hirschheim 1987, Wieringa 1989). Universe of Discourse, UoD, is a slice of real world (a mini world) in the problem domain containing a set of entities which are of interest to the relevant people. A conceptual model is an abstract entity which embodies a common understanding among the relevant people of the UoD (Wieringa 1989).

   A conceptual data model usually contains an organization of concepts and a graphical notation suitable for describing and defining the vocabulary and conceptualizing of the problem domain. The most common conceptual data model is the ER (Entity-Relationship) model.

2. Physical models are the most low-level models. These models provide concepts that describe how the data is stored in the computer. Record formats, indexes and access paths are typical structures of these models.

3. Implementation models or logical models are placed between the two former abstraction levels. Implementation models provide concepts that can be understood by the end user but which are not far from the way data is organized in the computer. The three typical implementation models used in databases are relational, hierarchical and network models.

The result of applying a data model to a specific problem domain is a schema, which is a description of a data collection in a chosen abstraction level. The schemas can be defined in three levels according to the three-schema architecture (the ANSI/SPARC architecture, Tsichritzis & Klug 1978):

1. The internal schema represents the physical storage structure of the database. This schema is produced applying the physical data model.
2. The conceptual schema describes the Universe of Discourse in question. This schema can be produced by using the conceptual data model or the implementation data model.
3. The external schema (also called user view) describes the database from the viewpoint of a group of users. This schema can also be produced using the conceptual data model or the implementation data model. An external schema normally contains parts of the conceptual schema.

1.5 Conceptual data modelling

1.5.1 Principles

The aim of conceptual modelling is to specify an explicit conceptual model of the Universe of Discourse. A conceptual model provides a formal basis for common understanding of the UoD. It defines the allowable ways in which information about the UoD can be stored (and manipulated) and provides a basis for interpretation of external and internal syntactical forms which represent information about the UoD (Wieringa 1989). There is a relation between a conceptual schema, the UoD it represents and information systems that implement the schema. Conceptual schema is an abstract mathematical structure expression of the UoD.

Wieringa (1989) presents three roles of conceptual models in the development and use of an information system. Firstly, a conceptual model represents the possible entity types, their possible states, processes and interaction in the UoD. This is the descriptive role of the model. There can be many different databases (occurrences) which correspond to a certain schema. The rules for generating the schema specify properties that must be true for all occurrences of the schema. (Tsichritzis & Lochovsky 1982). In addition to the descriptive role, a conceptual model may also have normative or institutional roles. It may include or create rules to specify what is permitted, forbidden or obliged in certain situations in the UoD (Wieringa 1989).

Conceptual models encapsulate structural aspects of objects (Hull and King 1987). The advantages of conceptual models lie in the support of database design and evolution. They provide a variety of abstraction mechanisms and trough them serve as a buffer between the form of requirements collected from the users and the low-level computer-oriented form of record-oriented physical models.

The abstraction mechanisms of conceptual models include classification, identification, specialization, generalization, aggregation and association (Batini et al. 1992, Elmasri & Navathe 1989). Classification (categorization) involves classifying similar objects into object classes (entity types). Identification is a process, where all abstract concepts and real objects are made uniquely identifiable by means of an identifier. Specialization is a process where class of objects is further divided into subclasses, i.e conceptual refinement. Generalization is inverse to specialization, a conceptual synthesization process where several object classes are combined into a higher level abstract class. Aggregation is used to build composite objects from their component objects. Association is used to associate objects from several independent classes, i.e the definition of relationships among classes.

1.5.2 ER model

The most common conceptual data model is the Entity-Relationship (ER) model. It was first introduced by Chen (1976). After that numerous extensions have been proposed to the ER model. Elmasri & Navathe (1989) incorporate their most

important concepts into the ER model and call the resulting model the enhanced ER or EER model (also extended ER model).

The basic object that the ER model represents is an entity, which is a thing in the real world with an independent existence. The existence can be physical or conceptual. Each entity has particular properties, called attributes, that describe and identify it. Each attribute is associated with a value set (domain), which specifies the set of values that can be assigned to that attribute. A set of entities that have the same attributes define an entity type.

Relationship is an object that connects one or more entities. A relationship type among a number of entity types is a set of associations among entities of these entity types. Associations indicate that the participating entities are related to each other some way in the real word. The degree of a relationship type is the number of participating entity types. A relationship type of degree two is called binary and relationship type of degree three ternary.

Relationship types usually have constraints that limit the possible combinations of entities participating in relationship instances. These constraints are derived from the miniworld situation represented by the relationships. These constraints are called cardinality ratios and participation constraints. Together they make up the structural constraints of a relationship type.

Cardinality ratio constraint specifies the number of relationship instances of one relationship type that an entity can participate in. Common cardinality ratios for binary relationship types are 1:1, 1:N, and M:N.

The participation constraint specifies whether the existence of an entity depends on its relation to another entity via the relationship. There are two types of participation constraints, total and partial. Total participation means that every entity of an entity type must be related to another entity via the relationship. Partial participation means that only some part of entities of an entity type are related to another entity via the relationship. Relationship type with total participation is also called mandatory and relationship type with partial participation respectively optional.

1.5.3 Advanced data modelling concepts (the EER model)

The EER model contains all the modelling concepts of the ER model. In addition to these, it includes the concepts of subclass and superclass.

In some cases an entity type may have numerous additional subgroups of its entities that are meaningful and need to be represented explicitly because of their significance to the database application. For example, an area delineated in a forest map (a compartment) may be further grouped into forest stands, agricultural areas, lakes etc. The set of entities in each subgrouping is a subset of the entities that belong to the compartment entity type, e.g. a forest stand is also a compartment. These subgroupings are called subclasses of the entity type compartment and compartment is called the

superclass for each of these subclasses. The process of defining subclasses for an entity type is called specialization.

The reason for defining subclasses may be that a subclass has specific attributes, or a subclass may participate in specific relationship types. For example, the data collected from of a forest stand may differ from the data of an agricultural area, or sample plots are measured only in a forest stand.

An important concept associated with subclasses is attribute inheritance. An entity that is a member of a subclass inherits all the attributes of the entity as a member of the superclass. It also inherits all relationship instances for relationship types in which the superclass participates.

For superclass/subclass relationship types there also exist constraints of disjointness and completeness. If a superclass entity can be a member of at most one of its subclasses, the subclasses are disjoint, e.g. a compartment may be either forest or lake, but not both. If the subclasses are not disjoint, their set of entities may overlap. The same entity may be a member of more than one subclass of the specialization. Completeness constraint includes two alternatives. Total specialization specifies that every entity in the superclass must be a member of some subclass in the specialization. Partial specialization allows an entity not to belong to any of the subclasses.

1.5.4 ER/EER notations

Three typical ER notations are shown in Fig. 1. All show entity types forest and tree and the relationship between them. In notation a) (the so-called Chen notation see e.g. Elmasri & Navathe 1989)) entity type FOREST STAND includes attributes ID, AREA and SITE. The underlined attribute ID is the key attribute (identifier). Entity type TREE has two attributes, NR (key attribute) and SPECIES. There is a relationship type CONTAINS between the two entity types indicating that in a forest stand there may be trees. The pair of integer numbers in the relationship line associate the cardinality ratios. The numbers (0,n) mean that a FOREST STAND may contain zero or more trees. Zero because it is assumed that clearcuts are also forest stands. From the TREE point of view, numbers (1,1) mean that each tree must be associated exactly to one forest stand. Generally, we can associate a pair of integer numbers (min., max.) with each participation of an entity type in a relationship type, where min ge 0, min le max, and max ge 1. This notation includes both the cardinality ratio and the participation constraint. Min = 0 implies partial participation and min > 0 implies total participation.

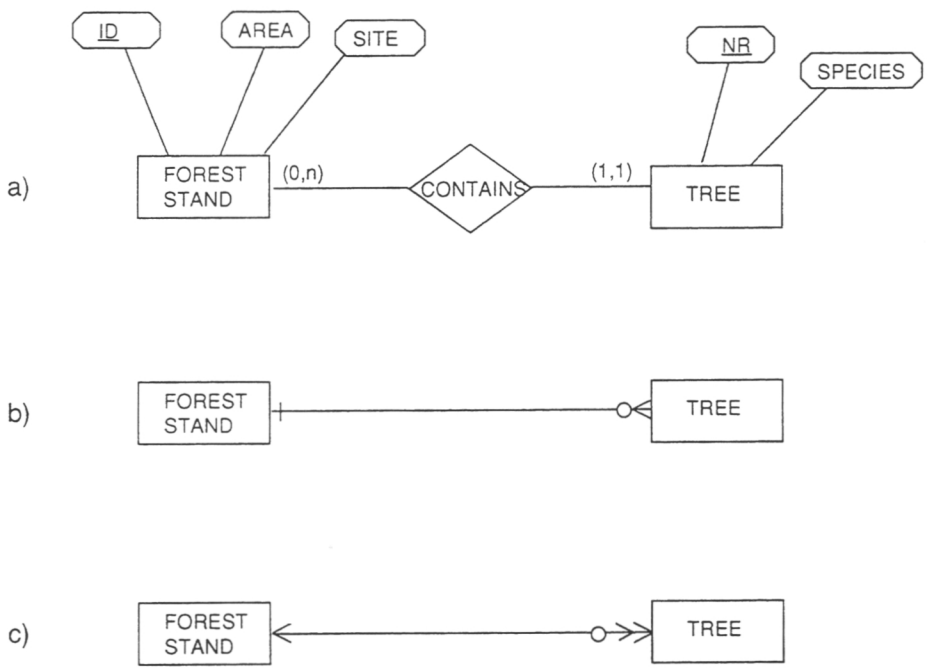


Figure 1. Examples of alternative ER notations: a) Chen , b) IE , c) Bachman.

Notation b) is used e.g. in Information Engineering (A Guide... 1990). In the notation attributes are not displayed. Relationship types are depicted as single lines joining the entity type boxes. The name(s) of the relationship type can be displayed along the line. The cardinality of the relationship type is depicted in the ends of the relationship lines as follows:

- a bar that crosses the line perpendicularly indicates only one (min. = 1, max. = 1)
- a crows foot at the end of the line indicates one or more (min. = 1, max. = n). The maximum number is not displayed in the diagram.

The participation constraint is depicted as a circle on the relationship line next to the cardinality symbol. The circle indicates partial participation. When the circle is omitted, it indicates total participation.

Notation c) is the so-called Bachman notation (Bachman 1969). The notation is close to the IE notation, with the exception of different cardinality symbols. One arrow indicates only one, double arrow one or more. The participation constraint is a respective circle as in IE.

The subclass/superclass relationships can be represented in the notations by defining 1:1 optional relationships between the superclass and its subclasses.

1.6 Data modelling for GIS

The entity-relation modelling is a general tool in business world database design. Laurini & Thompson (1992) note that there exist only a few examples of its use in geographic information systems design. Structures analysis methodology has been used in some projects (e.g. Bulger & Hunt 1991). An ER diagram is normally included in the structured analysis tools. Armstrong (1988) presents the use of entity-category-relationship diagrams in the design of temporal spatial databases.

Laurini & Thompson (1992) state that semantic data modelling could provide appropriate tools to identify the complex data structures of geographic information. The essential part of the modelling is to choose the presentation of spatial objects (point, line or polygon). Effective modelling for spatial databases requires attention to various other elements, such as the cardinalities of associations. Many-to-many relationships are quite common in geographic information and they must be decomposed and treated carefully. For example, a forest stand delineated in the inventory may belong to two or more forest lots and one lot always consists of many stands. The same situation exists between forest lots and owners.

The implementation data modelling of spatial databases is discussed in various papers. For relational model, see e.g. van Roessel (1987). For object-oriented spatial data modelling, see Worboys et al. (1990). For a Finnish review of relevant articles, see Sijainninhallinnan... (1991).

1.7 Temporality in spatial databases

Maps are usually two-dimensional. Normally attributes are considered as the third cartographic dimension and time as fourth dimension. Maps describe geographic entities. Each entity within the modelled system has location, attributes and a lifespan (time when the entity exists). Usually one of the components is fixed, one is controlled and only one can be measured on an interval or ratio scale (Langran 1992b). On a traditional map time is fixed. Attributes are included using different symbols and tones. Only location can be measured.

Langran (1992b) presents the following conceptions of cartographic time:

1. The space-time cube represents one time and two space dimensions in a theoretical three-dimensional cube. The model can be implemented in a CAD-system without topological relations.
2. Sequent snapshots of time slices represent changes as a series of states. It doesn't, however, represent the events that change the state. Also several changes between adjacent snapshots are not detected.
3. A third image of geographic time is a base state with amendments superimposed. Instead of states the model records change with its type and timing.
4. Space-time composite includes accumulated geometric changes in one coverage. It is based on a base state of some starting point representing the geometry and topology of the coverage in a chosen time. Each change (with distinct spatial location and extent) causes the geographic objects to break down into discrete objects with own distinct history. The resulting spatial objects (e.g. polygons)

represent the greatest common spatiotemporal units including distinct temporal attribute sets. Space-time composite reduces three dimensions (location, attributes, time) into two, so space can be treated atemporally and time can be treated aspatially. Temporality is an attribute of spatial objects.

The alternatives of handling the aspatial attribute information in a relational data model include relation-level versioning, tuple-level versioning and attribute-level versioning (Langran 1989, 1992b). Relation-level temporality creates and stores a new snapshot of a table when any of its attributes change. There exists various versions of tuple-level temporality handling. E.g. each tuple is supplied with time stamps denoting its lifespan. New tuples are appended to the relation without deleting existing ones. Attribute-level versioning supplies each attribute with respective time stamps requiring variable-length fields to store lists of attribute versions.

1.8 IS development in forestry

1.8.1 Data models

There exist numerous IS development projects where data modelling and systems analysis methodologies have been applied. The projects presented below are closely related to this work.

Large conceptual models have been produced in Finland for forestry purposes in the Finnish Forest and Park Service (Paikkatietojärjestelmien... 1989), in the Finnish Forest Research Institute (Saarenmaa et al. 1990) and in the Ministry of Agriculture and Forestry (Metsätalouden... 1991).

The Finnish Forest Research Institute (FFRI) has defined an information strategy for the whole institute. The strategy was formulated by studying the existing IS and future requirements using the IE-methodology. During the work two conceptual models, activity model and data model, were produced. The activity model described hierarchically the basic functions and work processes of the organization. The data model was an entity-relationship diagram of the data objects that are used by the activities. IS architecture was derived by forming a matrix of the supplementary activities and subject areas, arranging it by logical dependencies (create, update, delete) and defining information systems from the arranged matrix.

The same methodology as in the FFRI project was applied when an information architecture (containing data and business models) was developed for the whole forestry sector. The architecture is based on distributed processing of geographically referenced information in the forestry organizations.

Nalli (1992) has made a conceptual model to describe geographic information in multiple use forestry. He modelled the different spatial objects that should be considered in multiple use planning into point, line and polygon entity types and also discussed the principles of defining relationships between the spatial entity types.

In the Netherlands information and business models for an average forest enterprise have been produced using the IE methodology to make the knowledge of information modelling and management available for the forest enterprises (Borsboom & Six Dijkstra (1992).

The Ontario Ministry of Natural Resources and ESRI Canada have developed a forest management decision support system based on ARC/INFO GIS-tools (Bulger & Hunt 1991). In the development process, structured analysis and design methodology with a CASE tool have been utilized.

Kaila & Saarenmaa (1990) state that one general descriptive model can be defined for forestry, where all activities and data objects can be described. This definition can be made using formal conceptual modelling procedures. From this description can data flows and applications be derived. Each forest organization can use only those parts of the descriptive model needed for its business.

1.8.2 Temporality in forest information systems

Temporality in forest GIS has been studied by Armstrong (1988), Bulger & Hunt (1991), Langran (1992a) and Kennett (1992). There are two sorts of forest GIS temporality: events that may cause change and changes to the forest state itself (Langran 1992a). The temporal aspects of forest management can be summarized as follows:

- Compartmentwise forest inventories present sequent snapshots of the state of the forest. The state is described with spatial objects (forest stand polygons) and respective descriptive attributes
- Change caused by the continuous processes (growth, mortality) can be assumed to be aspatial. It can be predicted and calculated using growth models.
- Discrete events that cause change include silvicultural activities, fires, insect infestation, and unusual weather, such as windstorms. They have spatial and temporal location and extent (Langran 1992a). Temporal location concerns the state of the event, whether it is completed, underway or planned. Temporal extent spans the period when an event is performed. Spatial location and extent are defined by the area in which the event occurs. These events may cause that for one geographic entity 1. only spatial attributes change (e.g. a part of a stand is detached to build a new entity), 2. only feature attributes change (no spatial change), 3. both spatial and feature attributes vary, 4.a new geographic entity is established.

Langran (1992a) introduces alternative ways of handling forest temporal information. 1. Simplest method is to include narrative descriptions of activities in the database records. 2. Snapshots of the forest states are taken by rasterizing the vector data. 3. Third is the space-time composite method discussed in Chapter 1.7. 4. Feature histories are described by treating the attributes that describe different versions of the feature as separate database records. Geometric change is treated by referencing the correct spatial objects to each version of the feature as it changes over time.

There are some problems associated with the implementation of silvicultural operation history. 1. There are many types of activities and they all have different sets of

descriptive attributes. For example, cutting is described with method and outturn while regeneration attributes may include tree species, plant type, and planting density. These attributes can not be stored in a common database table. Furthermore the different types of activities must be examined together to get the information needs fulfilled. So it is not sufficient to create separate information layers for each activity type. 2. The same type of activity may occur on the same place many times, e.g. planting and repair planting. 3. Third problem is associated with the update of the inventory layer. An area may be regenerated using various methods (seeding, planting) with a homogenous result from the inventory perspective in some ten years. Some kind of generalization procedure must be applied to combine the regeneration polygons to basic inventory polygons.

Kennett (1992) introduces a practical silvicultural operation history model. Spatial data control is based on the ideas of space-time composite being a result of spatial overlay operations. The attribute data is stored in the relational database tables. A specific master table and individual activity data tables are separated. The master table forms the linkage between the spatial database and the activity attribute tables.

Bulger & Hunt (1991) present a temporal forest GIS-solution based on two principles, time stamped layers and database transaction history. Time stamping includes a mechanism that copies all the old (retired) data into a history layer. Transaction processing provides transaction tables for the recording of all events which change the database (both thematic and spatial data). An accumulated transaction history allows the reconstruction of the database as it existed at any point of time in the past.

1.8.3 Overview of the FFRI project

This study was made in the Finnish Forest Research Institute as part of a project called The Research Forest Database and Planning System. The project (named TUTGIS) was started in 1991 to develop the information processing and planning of the research forests owned by the institute. The primary objective of the project was to define and develop a GIS-based forest information and planning system. In the system both up-to-date and history information of the forests can be integrated with the experiments and conservation areas for forest planning purposes. For an overview of the project and its background, see Nuutinen (1991) and Paananen & Nuutinen (1993). This study is part of the project including the functional and information analysis and modelling for the new system.

The Finnish Forest Research Institute (FFRI) has some 150 000 hectares of state forest as its disposal. The forest area is used mainly for experimental purposes. The areas that are not in experimental use are managed according to the law concerning state forests. The new strategy for the forests is presented in Metsäntutkimuslaitoksen... 1993.

In the forests there are 2 300 experiments. Some 65 000 hectares are set aside for conservation purposes. The conserved areas include national parks, nature reserves etc. The forests are divided into research areas for management purposes.

When the project started the forests were managed according to the management plans. Management planning was made for each research area approximately every tenth year. Planning was based on a compartmentwise forest inventory using aerial photographs and field surveys. During the inventory 1:10 000 forest maps were produced using a vector based mapping system. The stand descriptions were stored in sequential attribute files used for statistical calculations. The attribute files and mapping systems were loosely integrated to produce thematic maps. For the experiment data there existed a register that contains general information about the purpose and approximate location of the experiments.

The system had some shortcomings. The inventory data were not updated, so the information about the stands may be over 10 years old. There was no accurate information about the location of the experiments, or the information was spread over numerous organizational units. There was no automated system for collecting silvicultural operation history with the exception of regeneration data that has been collected into manual files.

1.9 Aim of the study

The aim of this study was to define and test a logical descriptive data model of the primary data objects for the inventory, planning and monitoring of research forests.

The problem can be decomposed into 3 subproblems:
1. To analyse the requirements of research forests activities for information systems.
2. To present the structure and contents of the FFRI forest management information system in a conceptual schema.
3. To design and test a GIS-based database according to the conceptual data model with special respect to a) basic inventory data, b) stand history data and c) integration of experiment and stand data.

There are three underlying assumptions of the model and modelling method: 1. The model is independent of the implementation. There may be many different databases (occurrences) which correspond to the descriptive model (Wieringa 1989). 2. The resulting schema contains only basic data elements of the activities. Derived or aggregated data is excluded. 3. The descriptive model can successfully be utilized in the implementation of operational forest information systems.

2 Methods

2.1 Information Engineering

2.1.1 Overview of the methodology

Information Engineering methodology (IE) is a pragmatic, business-oriented methodology that considers the entire enterprise (Martin 1990). IS is seen as a support to achieve the strategic goals of the organization. IE has seven stages, five of which address various levels of information system development (A Guide... 1990). The

stages are Information Strategy Planning (ISP), Business Area Analysis (BAA), Business System Design (BSD), Technical Design (TD), Construction, Transition and Production.

During ISP a broad view of the information requirements of the business is established (for an example of the use of ISP tools, see Saarenmaa et al. 1990). In BAA stage a more detailed analysis on a particular segment of a business (business area) is performed and in BSD an application system is described supporting a segment of a particular business area in detail disregarding the target computing environment. During technical design the results of business system design are tailored to a specific target computing environment. The characteristics of the hardware environment, operating system and DBMS are considered. In construction stage all of the executable components of a system are created, e.g. programs, databases and screen formats. Transition refers to the installation of the system in a production environment, possibly replacing existing systems or parts of them.

IEF is an automated implementation of the IE methodology. It is a set of tools to capture the information needs of high abstraction levels to transform them into executable application system (A Guide... 1990). The IEF supports currently five first of the above listed stages. In this study, a business area analysis was made to analyse the requirements of the research forest system. Corresponding IEF tools were applied on OS/2 operating environment.

2.1.2 Business Area Analysis

Business area analysis involves the definition and refinement of the activities a business performs (called business functions and business processes), the things with which it deals (entities) and the interaction between the two. It is a refinement of a subset of the information architecture developed in the ISP stage. BAA is used to identify and define the business activities that make up business functions, data required for each business activity, the sequence of business activities and how business activities affect the data (A Guide... 1990).

The main tasks performed to achieve the objectives are data analysis, activity analysis and interaction analysis. In data analysis, the data used to represent the relevant things to the business (entities) and their interrelationships are defined resulting in a conceptual data model. In activity analysis, business functions are examined to determine the processes they comprise. The result of the task is an activity model. In interaction analysis the effect of activities on data are analysed and presented in an interaction model.

IEF provides tools to perform the above mentioned tasks. Data is modelled by building an Entity Relationship Diagram (ERD). Business activities are modelled hierarchically on an Activity Hierarchy Diagram (AHD). Interaction between activities and data can be modelled using data/activity matrices and action diagrams.

Data modelling in IEF includes some specific concepts not mentioned in Chapter 1.5.3. In this study subject areas and partitionings were applied. Subject area is defined

as an area of interest to the enterprise centred on a major resource, product or activity (A Guide... 1990). It consists of a set of entity types closely related to each other. Subject areas illustrate the essential structure of the schema.

An alternative way to represent subclass specializations in IEF is the use of partitionings. Partitioning is defined as a basis for subdividing entities of one type into sublasses (A Guide... 1990). The classification of a particular entity along a partitioning is based on the value of a specific attribute of the entity being partitioned. The attribute is called classifying attribute. For example, the partitioning of forest map compartments into forest stands, lakes etc. could be based on classifying attribute land cover class. Partitionings can be presented in IEFs ER diagrams by using boxes where the subclass boxes are placed inside.

2.2 The modelling approach in this study

In this study three ER notations are applied. Full IE notation with subject areas and partitionings is used in Appendix 1 where the whole model is presented. In Figs. 2-9 Bachman notation is used. A modified Chen notation is applied in Fig. 10.

The requirements of the system were collected by analysing current systems, interviewing researchers and those responsible for managing the research forests. There exist numerous reports and instructions concerning the use of research forests (Metsäntutkimuslaitoksen... 1985, Tutkimusaluetyöryhmän... 1989a, Metsäntutkimuslaitoksen... 1989b, Metsäntutkimuslaitoksen... 1993).

The conceptual schema of the FFRI forest management UoD represents it at a certain level of abstraction. Parts of the UoD are not modelled at all, and of the chosen parts some were modelled to a deeper level of detail. The objectives of the project affected the selection. The following selections were made:

1. The detailed integrity constraints were excluded of the schema. Constraints between attributes (e.g. the allowed values of trees height-diameter value combinations) were not defined in the schema.
2. Attributes were defined only for part of the schema. Those attributes that could be obtained from the FFRI forest management UoD or those that were necessary and useful in the test database design were included in the schema. For the included attributes the following properties were defined: 1. type (numeric or text variable), 2. description of the attribute, 3. length of the attribute values in bytes, 4. optionality of the attribute (are null values allowed), 5. possible default value, 6. attribute domain (permitted values).

From a data modelling perspective, geographic features are entities that have in addition to the descriptive attributes three specific (spatial) attributes, location, geometry and topology. Location is an identifying attribute.

In this study, certain geographic entity types were specialized into two entity types, spatial and feature entity types. The conceptual division is discussed e.g. in Langran (1992b). Spatial entity type (spatial objects) includes geometric and topological

attributes of the geographic entity and the feature entity type includes the corresponding thematic attributes.

Primitive spatial entity types (geometric primitives) are point, line and polygon. All other spatial entity types can be seen as subclasses of them. For example, the subclasses of the entity type polygon represent the thematically distributed geographic objects that are implemented in distinct information layers. The entities of one spatial entity type subclass have same feature attributes (relationships to corresponding feature entities) and they belong to the same logical group of geographic objects. The subclasses are assumed to be disjoint.

The approach was based on the assumption of the implementation. The most common approach in geographic information system construction is to segregate the aspatial (feature) data from the spatial data keeping the last in special structures while storing the feature attribute information in relational database tables (see Chapter 1.2).

In this study, those feature attributes that were suggested to be stored in relational database tables, were specialized and placed in feature entity types. The relationships between spatial and feature entities were defined using normal ER relationship types. So the model is not purely conceptual, as implementation aspects are taken into consideration. The principle is illustrated in Fig. 2.

Spatial information is stored in information layers. Some of these layers may have a connection to some database tables containing the feature attributes. The aim of this study was not to describe the technical details how the geographic information is exactly stored in these layers. The approach aimed at finding the logical groupings of geographic features and their relations.

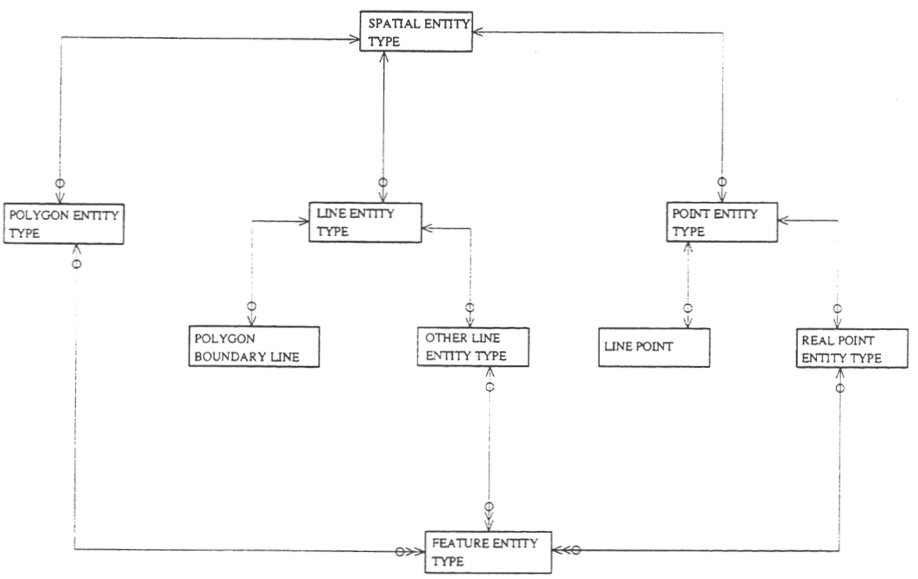


Figure 2. Basic entity types of the geographical information schema.

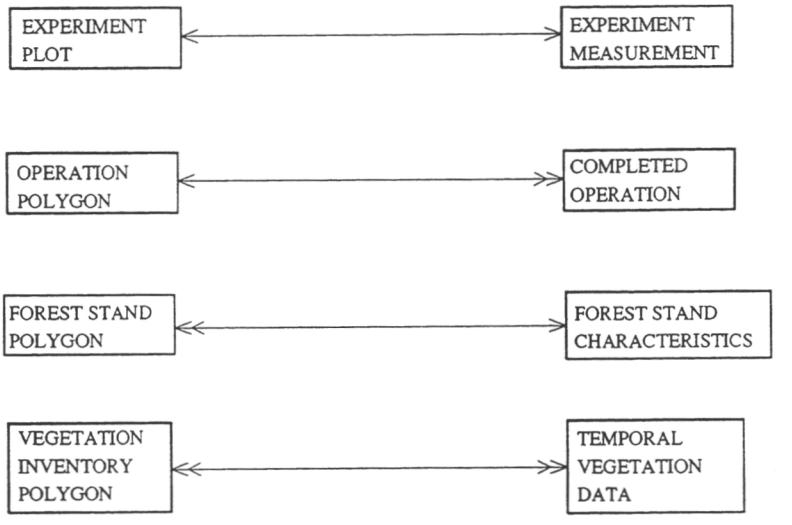


Figure 3. Relationship types of spatial and feature entity types.

The relationship types between spatial and feature entity types may have different structural constraints. This is illustrated in Fig. 3. In 1:1 relationship type for each spatial entity there is one corresponding feature entity type. E.g. an experiment plot may have one measurement data set (the latest measurement). 1:N represents the case where a spatial entity type has several attribute sets, e.g. an operation history polygon includes data of several completed operations. N:1 represents the case where one geographic feature comprises of 2 or more spatial objects. E.g. a forest stand is represented by 2 spatial polygons because a road divides the stand into two parts. M:N may be like N:1, in addition to it there may be feature attributes from several points of time.

2.3 Application of the model

2.3.1 Inventory design

For the part of the ER schema that concerns basic inventory data a detailed analysis was made to define the classifications of the stand characteristics. The development of the classification was based on the current inventory contents, information needs and existing stand characteristic classifications of the National Forest Inventory (Valtakunnan metsien... 1986) and Finnish Forest and Park Service (PATI-maastotyöohje... 1991). After the detailed analysis an inventory method with field guides and data collecting forms were designed. The new inventory design was tested in the Kivalo research forest. A total of 2000 hectares was inventoried during summer 1992. The inventory design is presented in Juvakka (1993).

2.3.2 Test databases

The FFRI has purchased Ingres relational database management system (RDBMS) and ARC/INFO GIS-system. The research forest information system was built upon these two commercial software systems. ARC/INFO is a file oriented geographic database system. It contains modules for both raster and vector data storage and handling. Spatial data storage is based on hierarchical data model and the information layer principle. In ARC/INFO spatial and attribute data are stored separately and are connected by using an interface from the GIS to the database system. The attribute data can be stored either in ARC/INFOs own tabular database (INFO) or in an external relational database. Connections between spatial features and an external database tables are established using the so-called database integrator. Database integrator allows ARC/INFO applications to view and use various external relational databases. External attribute tables can be stored in the DBMS and related to the spatial objects. A relate makes a connection between a record in the so-called spatial feature attribute table and a corresponding row in the related attribute table. Spatial feature attribute tables include internally attributes concerning location, geometry and topology. Users may add their own attributes, eg. the logical identifiers. An item (relate item in ARC/INFO terminology) in one spatial feature attribute table is used as a relate key (common attribute) to a corresponding column in the related table. The relate may connect two ARC/INFO tables, ARC/INFO and external database table, or two external database tables (so-called stacked relate). (Managing tabular... 1991.)

From the Kivalo inventory material a geographical database was designed and implemented. The database consists of tables in Ingres relational database system and information layers in ARC/INFO. The relational database tables were based on a part of he IEF data model (subject area basic inventory data). An IEF transformation module was utilized to generate the data description language (DDL) statements of the database tables from the model. The primary statements had to be modified to get suitable statements for Ingres RDBMS.

The ARC/INFO-layers were defined during the digitizing process using the data model (subject area spatial data). The background information was digitized from basic maps (1:10 000) and stand boundaries were digitized from field transparencies laid over the basic maps. Each element (e.g. roads, administrative lines) was stored in its own information layer according to the data model. A total of 27 layers were created during the data input. For the forest history structures a small test database was designed and implemented using materials of Kivalo area. The test database and applications designed in the project are documented in future TUTGIS publications.

3 Results

3.1 Research forest activities

The main activities of research forests are: 1. goal definition, 2. basic information processing, 3. forest management planning, 4. forestry operative control, 5. forest operations, 6. conservation area planning, 7. conservation area operative control, 8. conservation area operations, 9. estate management, 10. experiment planning, and 11.

experiment control and management. These activities were further detailed into functions and processes. The most important was the modelling of basic information processing, an activity that serves all other main activities.

3.2 ER schema of the research forests

3.2.1 General aspects

The whole ER schema is included in Appendix 1, and in Figs 4-9 parts of the schema are extracted and rearranged to illustrate basic inventory data, special land use data, planning, experiments and stand history schemas.

In the schema of Appendix 1, subject areas are used to aggregate spatial entities and feature entities into logical groups. All spatial entity types are included in subject area spatial data. The entity types not included in subject area spatial data represent the feature entity types data that are connected to the spatial entities to build complete geographic objects. Some of them refer to various databases or systems that were defined in the FFRI information system architecture (Saarenmaa et al. 1990).

In Appendix 1, the feature entities of basic forest inventories are grouped in subject area basic inventory data. The land ownership system (estates) are included in subject area kihti referring to a respective separate system. Special land use spatial entity types are grouped into two subject areas, administrative land use layers and functional land use layers. The corresponding feature entity types for both spatial subject areas are included in subject area land use data. The planning system has an interface to the TOTTI system. The information strategy defines TOTTI to be a system of planning and budgeting.

The interrelationships of points, lines and polygons are not illustrated in Fig. 2 and in Appendix 1 except the entity type line point, which is a specific subclass of point that comprises lines. The subclasses of polygon, point and line entity types inherit the attributes concerning location, geometry and topology. Because these attributes are internal structures of the GIS database, they were not included in the conceptual schema.

Some spatial entity types are subclasses of both line and point entity types. This means that some geographic features may be presented either by lines or points and may be included in one information layer.

For each point, line and polygon subclass, the most important attribute is the identification attribute (named ID). With polygons it refers to the user-defined identifier of the polygon (e.g. forest stand number). With lines and points it refers to the type of entity (e.g. type of road). The ID coding system (three-digit numbers) was derived from the Nalle mapping system (Nalle-metsäkarttaohjelmisto 1984).

Raster data was modelled to be a specific subclass of point named cell having special attributes row and column (for the identification of location) and standard size and shape (i.e. cells cover constant areas).

3.2.2 Spatial entities of basic mapping

Basic mapping schema consists of entities that are used to store the background map information collected by the National Board of Survey. The modelling of them was based on the new conceptual model of basic map information (Maastotietojen... 1993). The types of basic map entities were adjusted to the model of the basic map.

The various spatial objects of the basic maps were modelled into separate entities. A total of 23 entity types were defined from that specialization. The textual information of basic maps (e.g. names) is assumed to be connected in each entity type it concerns (in ARC/INFO it can be implemented using the so called annotation data types). The basic map entity schema is presented in Fig. 4.

Entity types ADMINISTRATIVE LINE (forest lot boundaries, commune boundaries etc.), ESTATE POLYGON (forest lots owned by the FFRI) and BOUNDARY MARK (points) refer to the land ownership system. Some entity types can be either points or lines. BUILDING may be presented either by its outer edge (line) or by digitizing the central point of the building. CONSTRUCTED STRUCTURE refers to other man-made structures than buildings, eg. fenced electric power stations. Those constructed structures that are presented by lines can be built to polygons. CONTOUR ELEMENT includes contour lines and altitude points. STREAM OBJECT contains under 5 meters wide streamlines and respective point objects (e.g. springs). TERRAIN OBJECT refers to remarkable point objects (e.g. stones) and edges of rocks. The latter can be built to polygons. Point entity type TREE refers to protected or otherwise remarkable separate trees, and trees selected for breeding.

Roads are presented by centre lines. They can be built to polygons (ROAD POLYGON) using buffer-operations. Those transmission lines that have an area (e.g. electric transmission lines) can respectively be buffered to polygons (TRANSMISSION POLYGON).

Line entity types that represent polygon boundaries include AGRICULTURAL LAND BOUNDARY, SHORELINE (shorelines of lakes and over 5 meter wide streams) and BUILT-UP AREA BOUNDARY. These boundary lines, together with constructed structure lines and terrain lines may delineate BASIC MAPPING POLYGONs. The relationship types between those entity types illustrate this function. Basic mapping polygon is an information layer where the land cover is filled exhaustively with water areas, constructed and urban areas, agricultural areas and forests. Forest area is not divided into stands, but there may be refined classification inside the other areas. The classification is made using values of an identifying attribute (ID). Polygon entity type PEATLAND POLYGON refers to peatland boundaries and polygons.

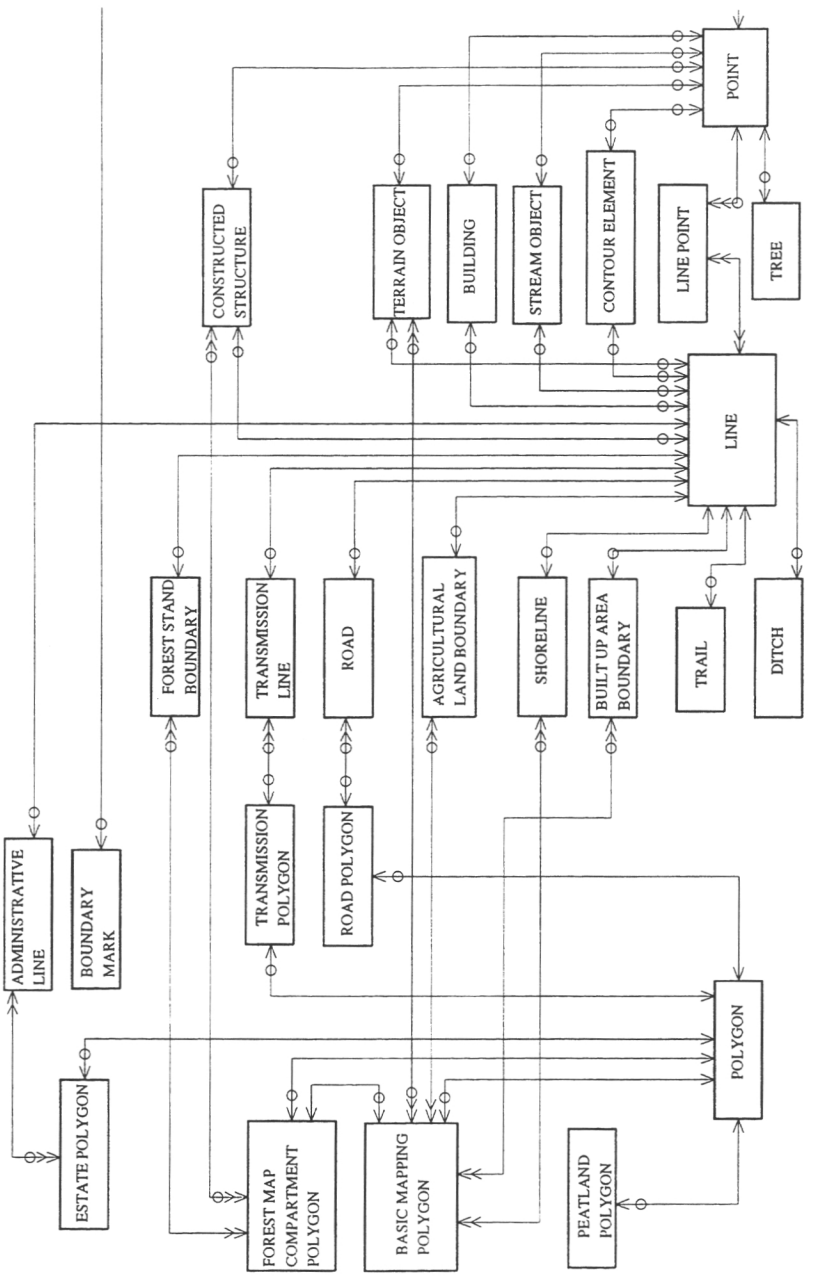


Figure 4. ER schema of the basic map entities.

Entity types FOREST MAP COMPARTMENT POLYGON and FOREST STAND BOUNDARY are also included in Fig. 4, although they are not basic mapping entities. The area owned by the organization is filled exhaustively with forest map compartment polygons. Some forest map compartment polygons are copied BASIC MAPPING POLYGONs, ROAD POLYGONs or TRANSMISSION POLYGONs. Most compartment polygons are comprised of FOREST STAND BOUNDARY lines.

3.2.3 Basic inventory data

The ER schema of compartmentwise forest description is presented in Fig. 5 (see also Fig. 4). The spatial entities include entity types FOREST STAND BOUNDARY, FOREST MAP COMPARTMENT POLYGON and INVENTORY SAMPLE PLOT POINT. The last mentioned entity contains locational information of the inventory sample plots. Forest stand boundaries are acquired in the compartmentwise inventory and updated by the organization. For a detailed description of the attributes and domains of basic inventory data, see Appendix 2 and Juvakka (1993).

FOREST MAP COMPARTMENT is the central feature entity type of the inventory data. The respective spatial entity is FOREST MAP COMPARTMENT POLYGON. The relationship is one-to-many, so a forest map compartment may consist of one or more separate spatial objects (polygons). E.g. a homogenous forest stand may be represented with 2 polygons because a road divides the stand into two parts.

Forest map compartment is further classified into five subclasses according to the attribute land cover class (for definitions of land cover classes, see Juvakka 1993). 1. FOREST STAND is the most important subclass of forest map compartment. It is a forest area homogenous in respect of soil, site and growing stock. 2. AGRICULTURAL COMPARTMENT refers to fields, pastures and meadows that are used for agricultural purposes. 3. BUILT-UP COMPARTMENT refers to housing and industrial areas, gravel pits etc. 4. ROAD COMPARTMENTs include all types of communication routes, transmission lines etc. that are considered as areal entities. 5. WATER COMPARTMENT contains water areas that are over 5 meters wide. Subclasses 3 - 5 could also be one subclass of FOREST MAP COMPARTMENT because in this study no specific attributes or relationship types have been defined for them. The full specialization was maintained in the schema to illustrate the land cover classification.

The total forest area is divided into research forests and blocks for management purposes (entity types RESEARCH FOREST and BLOCK). There is no corresponding spatial entities, and research forest and block numbers were modelled to part of the identifier of forest map compartments.

The forest area is also divided into several lots according to the land ownership system of Finland. This is represented by a relationship to entity type ESTATE. The spatial entities of estates were discussed earlier. Each compartment always belongs to one lot.

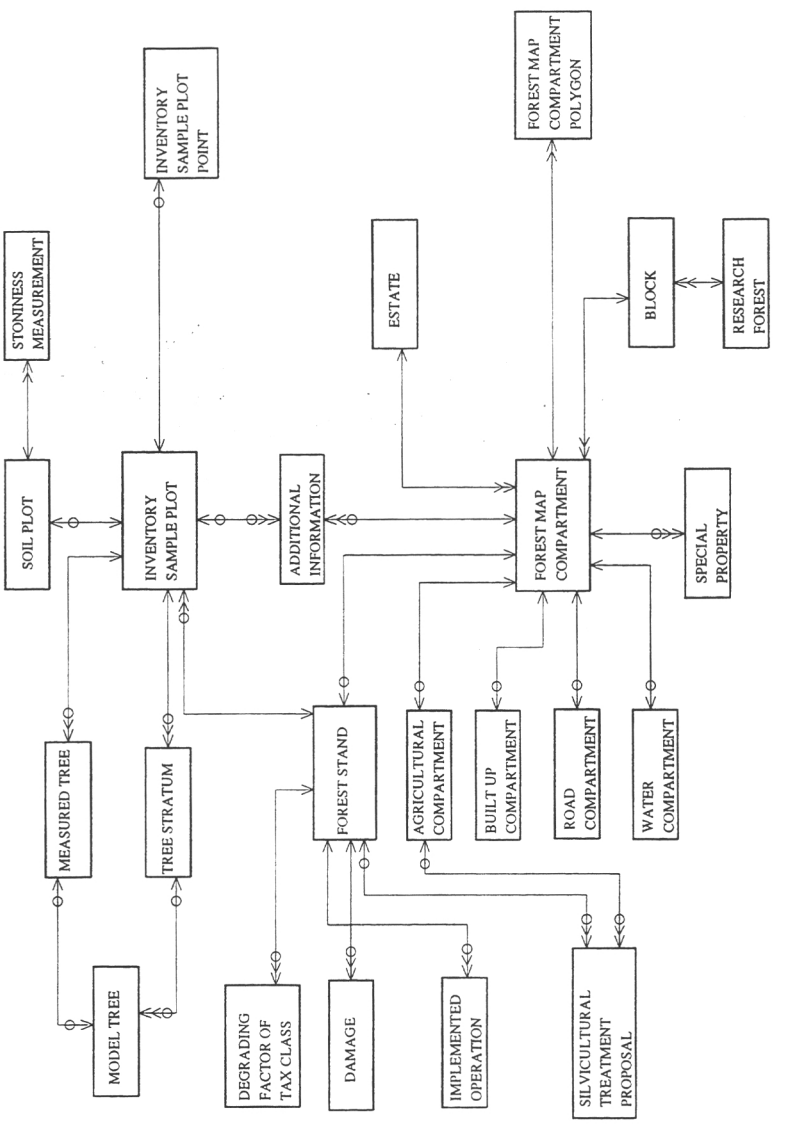


Figure 5. ER schema of compartmentwise forest description.

Basic inventory entities are already normalized (first normal form) for implementation. As a result of the normalization there exist entity types that could also be multivalued attributes of either forest map compartment or forest stand. These entity types include DEGRADING FACTOR OF TAX CLASS and SPECIAL PROPERTY. The former is an additional attribute used in forest taxation and the latter a special observation made in a compartment indicating needs for additional special inventory and planning. ADDITIONAL INFORMATION contains textual descriptive information of the forest map compartment. It can also be connected to a sample plot measured in a forest stand.

FOREST STAND is related to a set of specific entities that describe it in detail. For measurements of the growing stock INVENTORY SAMPLE PLOTs are placed in stands. Plots may be either variable size plots (relascope plots) or fixed size plots. For each plot, the growing stock can be measured either treewise (MEASURED TREE) or by tree strata (TREE STRATUM). Tree strata include weighted mean values of the growing stock stratified according to tree species and canopy layer. MODEL TREEs are calculated after the inventory from stratumwise mean estimates or measure trees to add growing stock volumes and growth to the measured growing stock values. An inventory sample plot may also be a unit for soil measurements. The entity types SOIL PLOT and STONINESS MEASUREMENT refer to soil type, thickness of the horizons and stoniness measurements (soil measurements were made in the test inventory).

For forest stands there may also be data concerning damages (DAMAGE), completed operations (IMPLEMENTED OPERATION) and suggested silvicultural operations (SILVICULTURAL TREATMENT PROPOSAL). Proposals may also be made for agricultural compartments (reforestation). There may also be some data collected about completed operations (IMPLEMENTED OPERATION). This data can be moved into the detailed history data structures.

3.2.4 Special land use data

The ER schema of special land use data is presented in Fig. 6 and the attribute lists of the entities is listed in Appendix 3. Special land use spatial entity types were divided into two logical groups, administrative land use layers and functional land use layers.

Administrative land use layers refer to areas that are based on an administrative decision (e.g. act, resolution of a government office). The boundaries of the areas are stable and the use of the areas are regulated by the decisions. Functional land use layers contain areas that are delineated in special inventories and planning. The instructions for the use of the areas can vary depending on the type and location of the specific area.

The problem of overlapping land use types is solved by using the information layer concept. By defining information layers the distinct areal divisions can be maintained properly and by using GIS overlay operations they can be integrated for specific needs.

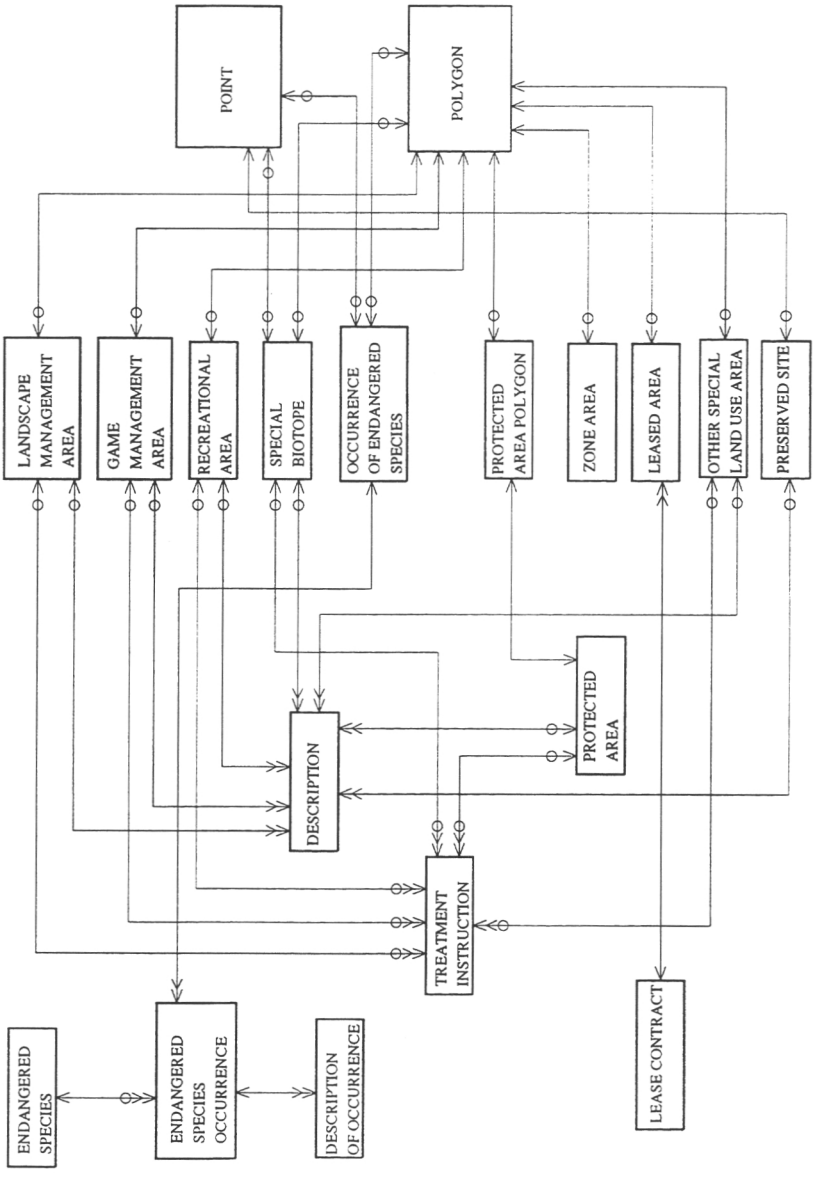


Figure 6. ER schema of special land use data.

Polygon entity types PROTECTED AREA POLYGON, ZONE AREA, LEASED ESTATE AREA, OTHER SPECIAL LAND USE AREA and point entity type PRESERVED SITE were grouped to administrative land use areas. Protected areas contain nature reserves, national parks and special areas that are established under the Nature Protection Act. Besides those areas there are also the areas that are part of protection programmes. The layer contains not only outer borders of the areas, but also inner subdivision based on combinations of land use classification, treatment instruction classification and access regulations classification. E.g. a part of a protected area (one polygon in the information layer) may be set aside of public access and it may be preserved untouched. The classifications are commonly used with protected areas (Luonnonsuojelualueiden... 1982). The basic feature attributes of protected areas can be obtained from external nature preserve area database (LSA-database). It is maintained by the Environment Data Centre. Attribute LSA IDENTITY CODE is an identifier to the LSA database.

Zone areas are results of zoning activity. There are four types of zone plans (attribute ZONE PLAN TYPE) and in each type there can be various land use classes (attribute LAND USE CLASS). The most detailed zone plan is stored in the layer. The zone area attributes are in this schema included in the spatial entity and no feature entity types are needed.

Leased areas have been leased out for a fixed time. There are various types of leased areas (attribute AREA TYPE). Each lease contract is identified with a contract number. The specification of the lease contracts data contents was not included in the study.

Other special land use areas are established by other administrative decisions than protected areas, typically resolutions of the FFRI. They include e.g. protected peatlands, nature management forests, wilderness areas and various areas for forest tree breeding purposes. Preserved sites are specific nature objects (located to a point) that are protected under an act or by resolution of the FFRI. These places may include e.g. relics and historic places.

Most of the spatial entity types are related to feature entity types DESCRIPTION and TREATMENT INSTRUCTION. These entity types were not modelled to attribute level, and it may be difficult to formalize the heterogeneous descriptions and treatment instructions to exact data types to be utilized in forest management.

Functional land use areas include polygon entity types LANDSCAPE MANAGEMENT AREA, GAME MANAGEMENT AREA and RECREATIONAL AREA. Entity types SPECIAL BIOTOPE and OCCURRENCE OF ENDANGERED SPECIES may be polygons or points.

Landscape management areas are treated with special attention to landscape maintenance and development. Game management areas are respectively areas where the protection of game is considered to be an important practice (e.g. areas that are protected at nesting-time or areas where hunting is forbidden). Recreational areas are used for recreation, outdoor activities, picking of berries etc., and they are treated

according to the specific requirements. Special biotopes are small scale sites that are not included in administratively protected areas. They may have protectional values or they need special attention in forest management planning. Attributes were not defined in detail.

Occurrences of endangered species include both plant and animal species. The areas may also be buffer zones created around and endangered species occurrence. The location, species and occurrence descriptions and treatment instructions (entity types ENDANGERED SPECIES, ENDANGERED SPECIES OCCURRENCE and DESCRIPTION OF OCCURRENCE) are suggested to be obtained from the UHEX-system. UHEX is a special database for endangered species maintained by the Environment Data Centre. The attribute definitions in this schema are consistent with the UHEX definitions. E.g. for each endangered species UHEX contains data about the current state of the species (STAGE CODE), main reason for being endangered and the biotopes where the species may occur.

3.2.5 Forest planning

Forest planning system utilizes the data of the basic inventory and land use. For the planning system, there are special spatial entity types defined for the storage of the plans. The planning system and its data contents are described in detail by Nuutinen (1994a, 1994b). In this schema (Fig. 7) only the main entity types and their relationships are presented.

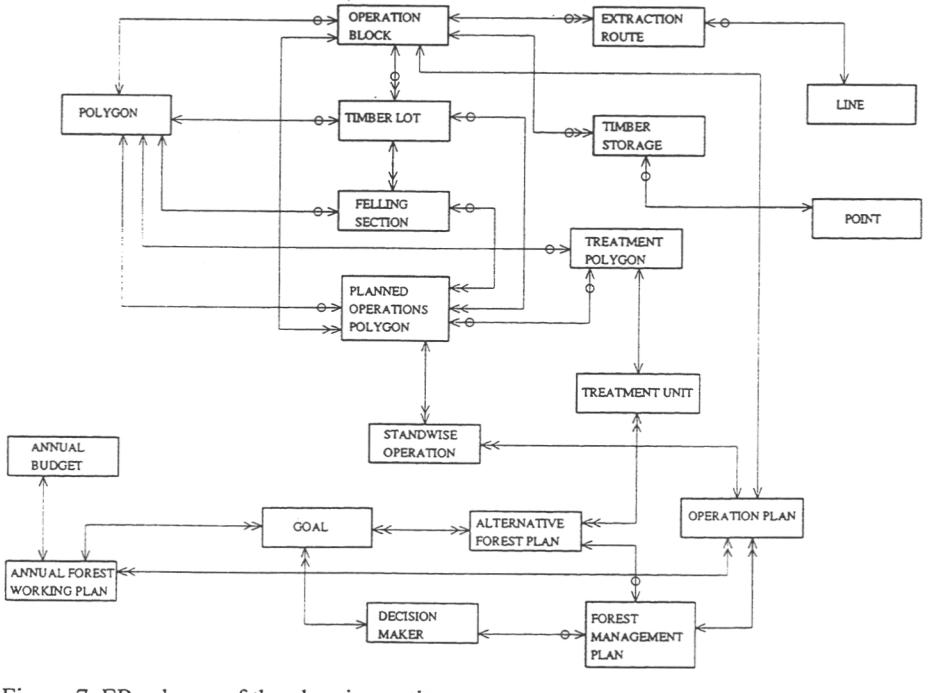


Figure 7. ER schema of the planning entity types.

PLANNED OPERATIONS POLYGON denotes a short-term planning unit. The delineations of operative planning normally follow the stand boundaries, but they may sometimes be delineated regardless of the stand boundaries. OPERATION BLOCK is a set of planned operations polygons, e.g. a logging unit that contains several stand polygons. Polygon entity types TIMBER LOT and FELLING SECTION, line entity type EXTRACTION ROUTE and point entity type TIMBER STORAGE are associated to short-term planning and management of harvesting.

One planned operations polygon (e.g. a cutting unit) is always treated uniformly with one or more STANDWISE OPERATIONs. Standwise operations include the descriptive data of any planned silvicultural operations, but the detailed data content is not included in this study. An OPERATION PLAN always consists of many standwise operations. An operation plan may be cutting plan or other silvicultural operation plan.

TREATMENT POLYGON is a variable content spatial entity type for various planning situations. Treatment polygon layer may include unions (overlays) of information layers needed in the planning process (Nuutinen 1994b). The corresponding feature entity type, TREATMENT UNIT includes the description of the forest (originating from the basic inventory data) and keys to the goals, restrictions and treatment instructions of the polygons included in the analysis process. Treatment polygons may be copied to PLANNED OPERATIONS POLYGONs, if the planning process ends in an operational plan that is later implemented.

ALTERNATIVE FOREST PLANs are based on simulation process where the stand data is updated and future development with treatment alternatives are calculated for each treatment unit. Entity type GOAL includes the management objectives that guide the planning process. DECISION MAKER is the instance that sets the goals and selects the acceptable FOREST MANAGEMENT PLAN among the alternative plans. The process with detailed data definitions is described by Nuutinen (1994b).

The system system for the planning and budgeting is not defined in this context, but two entity types, ANNUAL BUDGET and ANNUAL FOREST WORKING PLAN have been included in the schema. They concern a research forest. Annual forest working plan is based on the forest operation plans and guided by planning goals. Annual budget is respectively based on the annual working plan.

3.2.6 Experiments

The ER schema for the experiments is presented in Fig. 8. Two spatial entity types were defined, EXPERIMENT STAND POLYGON and EXPERIMENT UNIT. Experiment unit is the actual area of the experiment (usually permanent sample plot). It may sometimes be a single point. Experiment stand polygon denotes the effective area around the experiment unit that is treated according to the researchers' instructions. It may be a strip of fixed width around the plot or a larger area, e.g. a whole stand.

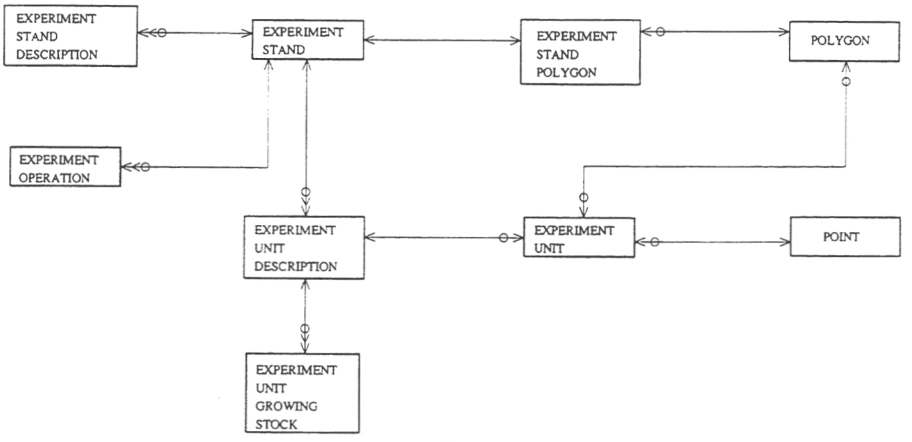


Figure 8. ER schema of the experiments.

The feature entity types of experiments are stored in an existing relational database system called experiment database (KOEREKISTERI). The content of the register is not presented completely in this context. Only those entity types and attributes that are meaningful for the forest management are presented. For full description, see Lehto and Isomäki (1993). Attributes of EXPERIMENT STAND and EXPERIMENT STAND DESCRIPTION refer to experiment stand polygon. EXPERIMENT UNIT DESCRIPTION and EXPERIMENT UNIT GROWING STOCK correspond to experiment unit. For each experiment stand, there may be descriptions from several points of time. Respectively, there may be growing stock estimates from several points of time for one experiment unit. EXPERIMENT OPERATIONs include planned, partly implemented and completed operations of experiment stands. They are not recorded accurately in the experiment unit level, but the presented principle is enough for forest management purposes to get an indication about the location and schedules of planned experiment operations.

3.2.7 Forest stand history

The ER schema for the history is presented in Fig. 9 and the attribute lists of the entities is listed in Appendix 4. The forest stand history solution is based on the space-time composite principle presented by Langran (1992b). Two spatial entity types, OPERATION HISTORY POLYGON and STATE HISTORY POLYGON, were defined.

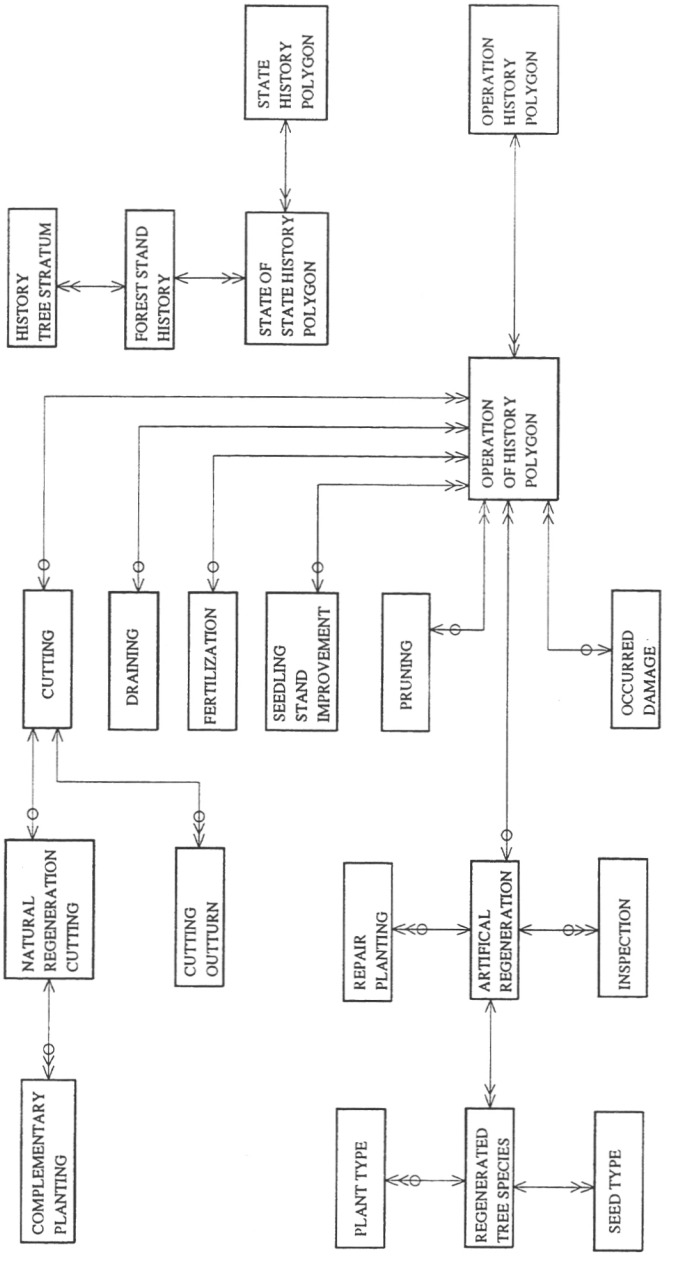


Figure 9. ER schema of the forest stand history.

3.2.7.1 Operation history

Information about implemented operations is essential in experimental forests when searching for suitable locations for different types of experiments. For example, when placing growth and yield experiments it is a necessity to know whether the stand has been fertilized in the past or not.

OPERATION HISTORY POLYGON refers to the greatest common spatiotemporal units in respect of events. Events refer here to silvicultural operations and different kinds of damages. Together they are called operations. An operation history polygon has an unique history of completed operations distinct from the neighbour polygons.

For each operation history polygon, there are several completed operations (OPERATION OF HISTORY POLYGON). Each operation may be one of several types (CUTTING, DRAINING, FERTILIZATION, SEEDLING STAND IMPROVEMENT, PRUNING, ARTIFICIAL REGENERATION or OCCURRED DAMAGE). Each operation is identified (in addition to time and location) using specific attributes, e.g. cuttings is identified using the cutting method and fertilizations using the type and amount of fertilizer. Because of the space-time-composite principle, each operation having spatial location and extent different from the operation history layer's existing geometry and topology causes the changed portion of the layer to break from its parent spatial object to become a distinct object with its own operational history. This principle is denoted with the 1:N mandatory relationship between all types of operations and entity type OPERATION OF HISTORY POLYGON.

E.g. when a thinning is made, it may be stored as one operation history polygon and its attributes are stored in CUTTING. When the subsequent operation, e.g. pruning, is made, it may be applied only to a part of the cutting area. In the operation history coverage, the original polygon is divided into two polygons, the other including both cutting and pruning, the other one only cutting. The original entity of type CUTTING can be maintained, but it is now related to two polygons (two entities of type OPERATION OF HISTORY POLYGON). An example of ARC/INFO and relational database implementation of the schema is presented in Chapter 3.3.

Entity type CUTTING has one subclass, NATURAL REGENERATION CUTTING. Natural regeneration is promoted using soil preparation and clearing. Sometimes natural regeneration needs also one or more COMPLEMENTARY PLANTINGs. The normalization of entity type CUTTING resulted in entity type CUTTING OUTTURN that includes the outturn of the cutting divided to timber assortments.

The definition entity types related to entity type ARTIFICIAL REGENERATION was based on the analysis of manually maintained archives of artificial regenerations. In the schema, clearing, soil preparation and prescribed burning activities are enclosed as attributes to ARTIFICIAL REGENERATION. Each regeneration includes one or more tree species (REGENERATED TREE SPECIES). Tree species information always contains information about the seeds (SEED TYPE) and when planting is used, also information about the plants used (PLANT TYPE). Sometimes the regeneration fails and REPAIR PLANTING is needed. The regenerated area is monitored using INSPECTIONs.

3.2.7.2 State history

The state history spatial object (STATE HISTORY POLYGON) includes the greatest common spatiotemporal units in respect of state of the forest. It follows the space-time-composite concept. For each state history polygon, there may be many states of the forest (STATE OF STATE HISTORY POLYGON). The state is identified using start and finish dates and an identifier (the number of the stand during the time span). Each state is related to the description of the forest that was effective during the time span (feature entity types FOREST STAND HISTORY and HISTORY TREE STRATUM). The attribute entity type structure is based on the tuple-level versioning principle discussed in Chapter 1.7.

The principles of updating the state history can be stated as follows:

1. As the stand is invented for the first time, its spatial and feature data is copied to the state history provided with starting dates (starting date = inventory date). This inventory coverage provides the so-called base state to the space-time-composite (Langran 1992b)
2. As the stand is invented again, the state history entity of first period is closed by updating the closing dates.
3. The spatial and feature data of the new inventory is copied to the state history provided with starting dates (starting date = inventory date). If the stand boundaries have changed, the coverage is fragmented using the space-time-composite approach.

The update principle denotes that state history also contains the effective state of the forest (stands with missing closing dates). An inventory is in this context understood not only as an extensive region-wide inventory, but also a complete revision of the delineation and description of one stand. It normally includes new measurements of the growing stock. Such a small-scale inventory is needed e.g. after thinnings to measure the growing stock. On the other hand, corrections to existing descriptions (e.g. correction of site type) does not cause the updating of the state history. Such changes that are of less importance can be handled either by using the so-called transaction logs or by updating them directly to the state history entities (without changing the date attributes) in addition to the update of the basic inventory data.

The data content of the state history (feature entity types FOREST STAND HISTORY and HISTORY TREE STRATUM) may nearly be the same as in basic inventory data. Only time stamps (closing dates) should be added. The plotwise growing stock measurements of the basic inventory may also be summed to the history to store the stratumwise mean values of the whole stand. The model trees are not copied to the history.

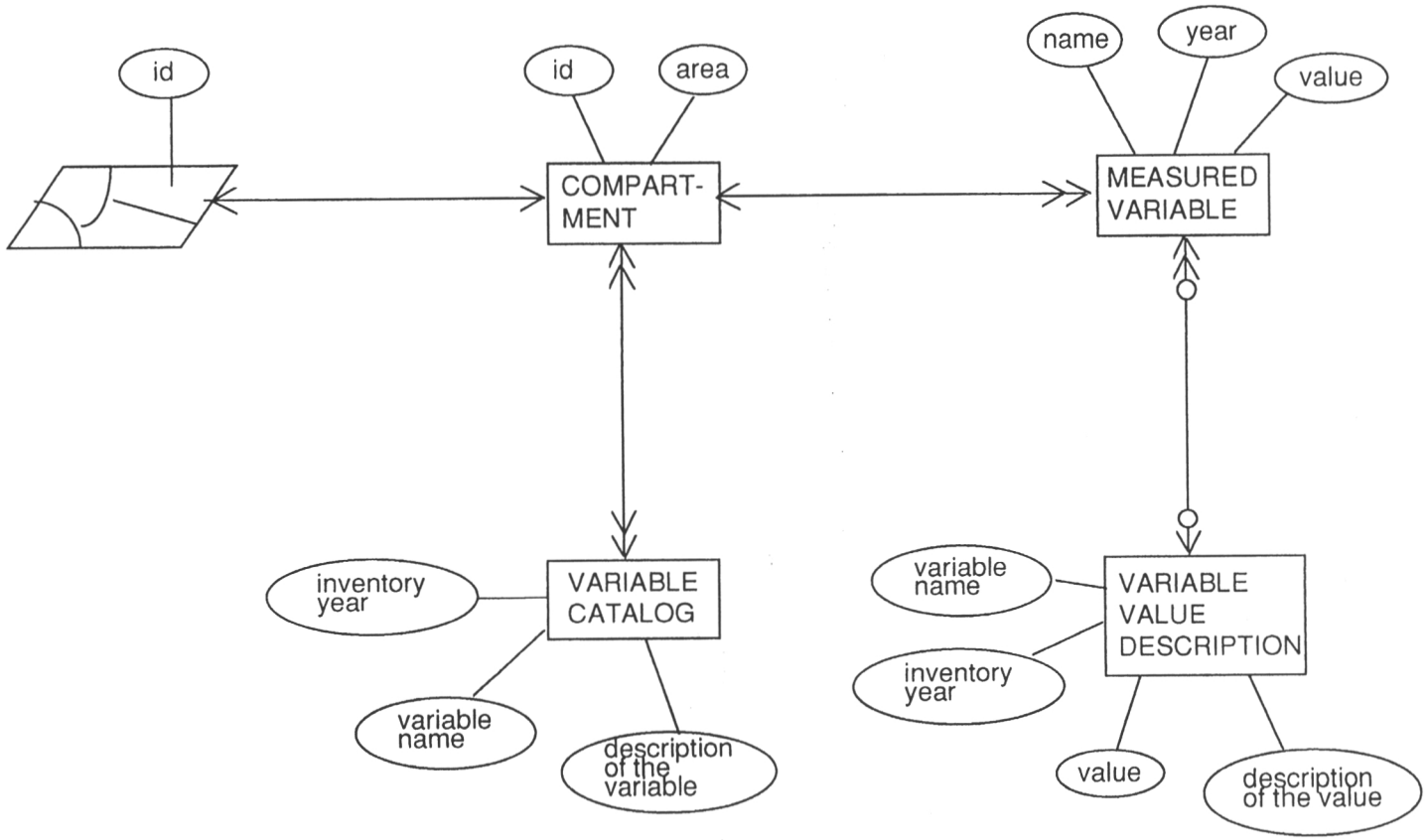


Figure 10. A general conceptual schema of a multitemporal compartmentwise forest inventory information.

A general illustration of a multitemporal compartmentwise forest inventory information model is presented in Fig. 10. The spatial data is organised according to the space-time-composite principle, so each polygon has a unique series of inventory attributes. The stand variables used and their values may both vary as function of both space and time. Entity type COMPARTMENT denotes the basic feature entity. The definitions of stand variables used in each inventory (and compartment) are stored in VARIABLE CATALOG. The actual stand attributes for each compartment and inventory are stored in MEASURED VARIABLE and the descriptions of the attribute values in VARIABLE VALUE DESCRIPTION.

3.2.8 Spatial entity types with interfaces to other systems

There are also some spatial entity types included in the model that are not specified in detail. Their relationship types to feature entities present interfaces to special systems and databases (see Appendix 1). These entity types and interfaces include:

1. Spatial entity type TREE is related to JALTA systems entity type SELECTED TREE FOR BREEDING. JALTA is a system for forest tree breeding data.

2. Spatial subject area VEGETATION LAYERS and feature subject area VEGETATION INVENTORY DATA refer to vegetation inventories. The inventory data may include VEGETATION POLYGONs and VEGETATION SAMPLE PLOT POINTs. For an example of a vegetation inventory database made in Pallas-Ounastunturi national park, see Eeronheimo (1993).

3. Spatial subject area SOIL LAYERS is defined for soil surface information contents. It is a presentation similar to the digital maps of quaternary deposits made by the Geological Survey of Finland. Soil type polygons are presented with SOIL POLYGONs, and the SOIL MAP POINT denotes indefinite parts of a soil polygon where the surface soil type (0-30 cm) differs from the soil type below it. SOIL SAMPLE POINTs are points where specific measurements with drillings are made. For details of geological surveys and maps, see Haavisto (1983).

4. For the archiving of photographs, the data of PHOTOGRAPH REGISTER is located to PHOTOGRAPH POINTs. Photograph register was not included in the IS architecture.

5. SITE MAPPING POLYGON is defined for site type inventories and forest taxation purposes.

3.3 An application: Stand history data structures

3.3.1 Operation history data structures and operations

The operation history data structures include one ARC/INFO coverage named PL_OPHIST. It contains all the polygons of the implemented operations accumulated

using the space-time-composite. No base state is needed for the coverage, so it is not exhausting. The areas where no operations have been completed are null areas (polygon identifiers are set to zeroes). The polygon attribute table (PAT) definition of the coverage is presented in Table 1.

In a PAT-table, first three attributes (area, perimeter. pl_ophist#) are ARC/INFO-supported. The fourth (pl_ophist-id) is a user-defined identifier of the polygons. In this case, the only requirement for the identifier is that it should be unique over the coverage. For practical purposes, research forest number is included in the table to enable reasonable numbering of the polygons. Arckey is a technical attribute, it is a converted character composite of pl_ophist-id and research_forest_nr. It is used in the relates to build a single unique key to the relational database tables. In ARC/INFO the linking attribute between polygon attribute tables and external database tables must be a single attribute. A sample listing of the PAT-table is presented in Table 2.

Table 1. Polygon attribute table definition of the operation history coverage.

item name	width	output width	type	alternate name
area	4	12	F	
perimeter	4	12	F	
pl_ophist#	4	5	B	
pl_ophist-id	4	5	B	polygon-id
research_forest	2	3	B	
arckey	8	8	C	

Item name = name of the attribute, width = storage size (in bytes), output width =output width(in characters), type = attribute type (F=real number, B=binary integer, I=integer(1 byte/digit), C=character string), alternate name = optional alternative name of the attribute.

Table 2. A sample listing of the PL_OPHIST.PAT- table.

attribute	polygon 1	polygon 2
area	9,861.688	959.93 (square meters)
perimeter	781.691	160.963 (meters)
pl_ophist#	2	3
pl_ophist-id	1	2
research_forest	112	112
arckey	11200001	11200002

Table 3. Relational database table definitions of the operation history.

column name	type	length	nulls	defaults
Table Operation:				
research_forest	integer	2	no	no
compartment_nr	integer	2	no	no
arckey	char	8	no	no
area	float	4	yes	no
operation_type	char	3	no	no
operation_code	char	7	no	no
Table Cutting:				
operation_type	char	3	no	no
research_forest	integer	2	no	no
cutting_nr	char	7	no	no
method	integer	2	no	no
additional_text	char	50	yes	no
Table Regeneration:				
operation_type	char	3	no	no
research_forest	integer	2	no	no
regeneration_nr	char	7	no	no
regeneration_date	date		no	no
height_above_sea	integer	2	yes	no
site_type	char	12	yes	no
site_class	integer	1	yes	no
soil_type	integer	1	yes	no

Column name = name of the attribute, type = attribute type, length = storage size (in bytes), nulls = can the attribute have missing values, defaults = does the attribute have default value (0 in integer types, blanks in character types).

An example of relational database table definitions of the operation history is presented in Table 3. The tables correspond to entity types OPERATION OF HISTORY POLYGON, CUTTING and ARTIFICAL REGENERATION. The latter two were entity types not mapped completely, only a few attributes were defined in the table to illustrate the principle. The tables were named operation, cutting and regeneration. Their relations are also illustrated in Fig. 11.

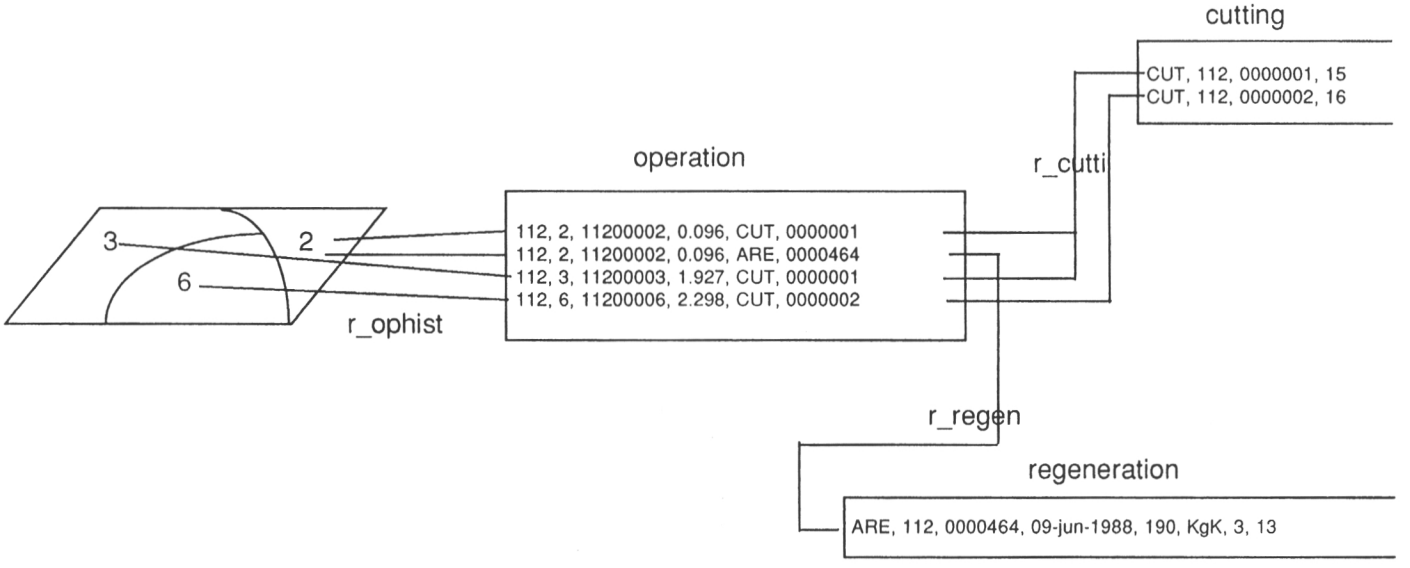


Figure 11. An illustration of operation history relational database tables.

The attribute operation_type of table OPERATION refers to the specific operation type and table. It is here defined as a character string, e.g. CUT for cutting and ARE for artificial regeneration. Operation code denotes a unique identifier of an operation. E.g. cutting may be numbered in a research forest area beginning from 1 and ending at 9999999. The field is a character string allowing alphanumeric identifiers, because they are used in the manual regeneration archives of FFRI.

Several relates were defined to connect ARC/INFO PAT-table and corresponding Ingres-table. The definition listings of the relates are presented in Table 4. Relate type denotes the type of connection made and relate access denotes the RDBMS access mode (RW = read and write, RO = read only).

Table 4. Relate definitions of the operation history.

relate name	table	database	item	column	relate type	relate access
R_OPHIST	operation	luotihist	arckey	arckey	first	rw
R_REGEN	regeneration	luotihist	operation_nr	regen_nr	first	rw
R_CUTTI	cutting	luotihist	operation_nr	cutting_nr	first	rw
RV_REGE	op_regen	luotihist	arckey	arckey	first	rw

Relate R_OPHIST of Table 4 presents the basic relate between operation history coverage PAT-table and corresponding Ingres-table OPERATION. The attribute arckey presents the composite identifier of the polygons in both tables. LUOTIHIST is the name of the test history database.

The relation is 1:N, because for each operation history polygon, there may be several completed operations. The relate type FIRST denotes that for each PAT record, the first relational table row matching the relate key is returned. For browsing through all the related rows, a special approach called cursor has to be used. With cursors, an application can access the related data one row at a time. The values of columns in the current row are available for update and listing (Managing tabular... 1991).

Relates R_REGEN and R_CUTTI present stacked relates between the table OPERATION and the tables that contain the specific attributes for different operations (in the illustration artificial regeneration or cutting). The relate attribute is operation_nr, which in case of the regenerations denotes the regeneration number and in case of cuttings denotes the identifier of a cutting.

To find and relate only the relevant operation history polygons (those having one ore several desired operations, e.g. implemented artificial regenerations) a three stage procedure has to be completed:

1. A View of the table OPERATION is created containing only rows with desired operations. The result is a subset of the rows of the original table. For example, he view for regenerations can be created using the following ARC/INFO commands:
   Arc: dbmsexecute luotihist
   >: create view op_regen as select * from operation where operation_type = ARE
   >: end

2. A relate to the view is defined. It is presented in Table 4 (relate RV_REGE).

3. A selection operation of in the polygon coverage is made to select only those polygons that have implemented artificial regenerations (or cuttings):
   Arc: reselect pl_ophist pl_regen poly
   Reselecting POLYGON features from PL_OPHIST to create PL_REGEN
   Enter a logical expression. (Enter a blank line when finished)
   >: reselect rv_rege//operation_type = ARE

A new coverage (pl_regen) is created containing those polygons that have a corresponding row in the view of regenerations.

Respectively, a stacked relate relating the view and the detail table of regenerations can be defined and that can be used e.g. to produce thematic maps. In the example below, a polygon shading could be made using soil_type as the controlling attribute:
Arcplot: polygonshades pl_regen rv_rege//r_regen//soil_type

This kind of structures can also be used to produce thematic maps having only relevant polygon boundaries. In fact, a part of the space-time-composite coverage is extracted to yield a map having only the desired operation histories. According to the space-time-composite principle, an operation history polygon has an unique history of completed operations that is distinct from the neighbour polygons. Because some polygons may be distinct in respect of cuttings, but not in respect of regeneration, they can be joined when regarding only the regeneration history. From the above coverages (e.g. PL_REGEN) the boundary lines that are not relevant in respect of the regeneration history can be removed. This is made in ARC/INFO using DISSOLVE operation.

A typical query from operation history could be to get all operations of certain type made in certain area over a certain period of time. This kind of queries can be solved with views, relates and reselect operations as follows:
1. a view is created, for example for regenerations:
   create view example1 as select * from operation, regeneration where operation.operation_type = ARE and (operation.operation_nr = regeneration.regen_nr) and (regeneration.regeneration_date between date1 and date2)
2. a relate is defined for this view
3. respective polygons are reselected from the coverage.

A possible query could also be to get all the operations made in specific compartment over time. For this query, a view must also be is created containing the desired

compartments. Using this view and relates it is possible to get the detailed operation information for the compartment.

3.3.2 State history data structures

The state history test database was created for a small area from two successive inventories (1981 and 1992 inventories). The space-time-composite layer of the two states was named SSINV. It was created using the 1981 inventory coverage (named FCOMP81) as base state. The 1992 inventory (coverage PL_FCOMP) was accumulated to 1981 inventory using a union-operation.

The union operation included the original attributes of the both 1981 and 1992 inventory coverages in the resulting polygon attribute table (Table 5). SSINV-ID is the identifier given by the union operation. Attributes PL_FCOMP#...ARCKEY originate from coverage PL_FCOMP (inventory 1992) and attributes FCOMP81#...COMPARTMENT_ID originate from FCOMP81. COMPARTMENT-ID is the compartment number of the 1981 inventory.

Table 5. Polygon attribute table definition of the state history coverage.

item name	width	output	type	alternate name
AREA	4	12	F	
PERIMETER	4	12	F	
SSINV#	4	5	B	
SSINV-ID	4	5	B	
PL_FCOMP#	4	5	B	
PL_FCOMP-ID	4	5	B	
RESEARCH_FOREST	2	3	B	
BLOCK_NR	2	2	B	
COMPARTMENT_NR	2	4	B	
SUBCOMPARTM_NR	2	2	B	
ARCKEY	11	11	C	
COMPARTMENT-ID	5	5	C	

Item name = name of the attribute, width = storage size (in bytes), output width =output width(in characters), type = attribute type (F=real number, B=binary integer, I=integer(1 byte/digit), C=character string), alternate name = optional alternative name of the attribute.

The principle is also illustrated in Fig. 12. The base table contents correspond to the definition in Table 6. From the table it can be noted that e.g. compartment 641 in 1981 inventory was in 1992 inventory divided into parts of two compartments, numbers 107 and 105. In the space-time-composite this is represented by two polygons, numbers 439 and 453.

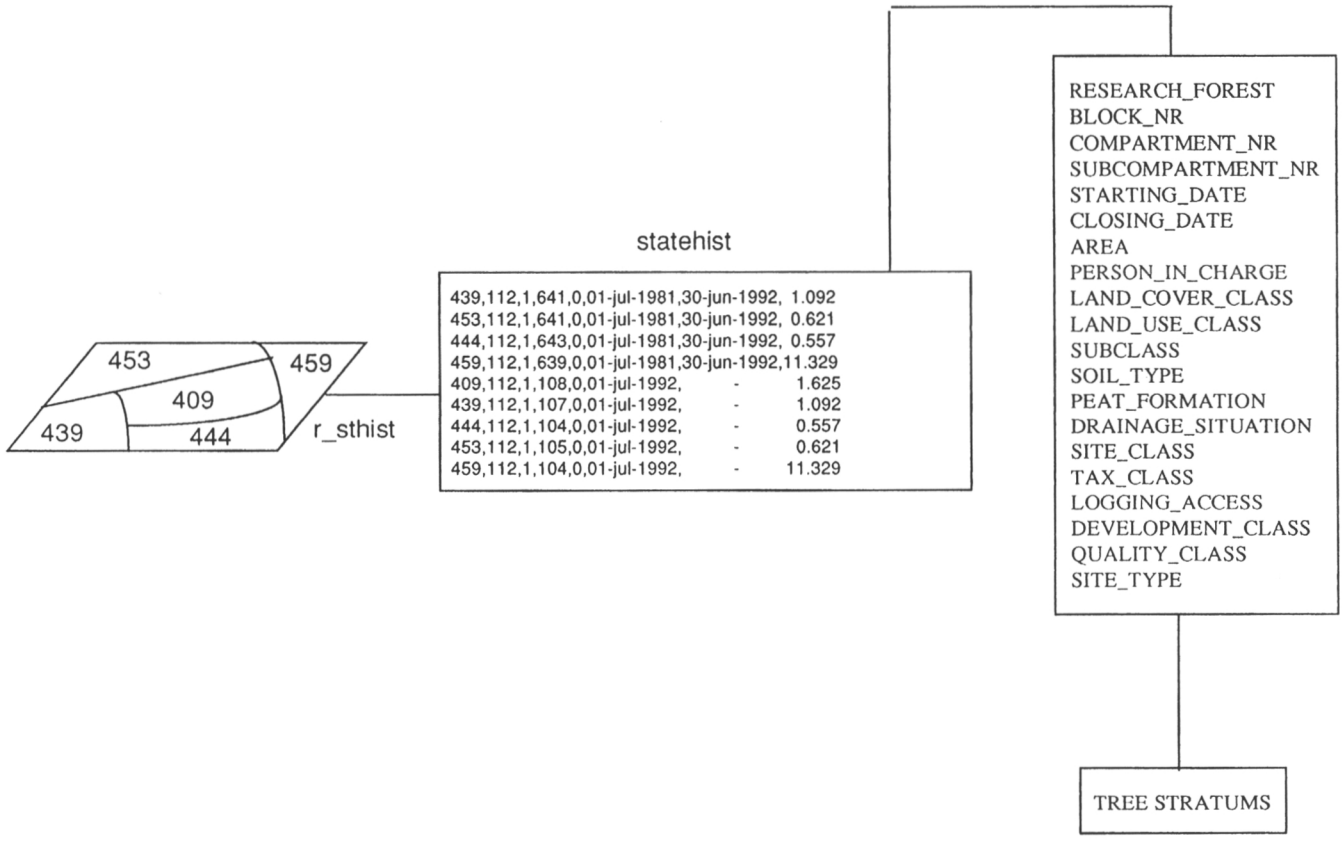


Figure 12. An illustration of state history relational database tables.

Table 6. Relational database table definition of the state history base table.

column name	type	length	nulls	defaults
ssinv_nr	integer	4	no	no
research_forest	integer	2	no	no
block_nr	integer	2	no	no
compartment_nr	char	11	no	no
subcompartm_nr	integer	2	no	no
starting_date	date	no	no	
closing_date	date	yes	no	
area	float	4	yes	no

Column name = name of the attribute, type = attribute type, length = storage size (in bytes), nulls = can the attribute have missing values, defaults = does the attribute have default value (0 in integer types, blanks in character types).

Table 7. Relate definitions of the state history.

relate name	table	database	item	column	relate type	relate access
R_STHIST	statehist	luotihist	ssinv-id	ssinv_nr	first	rw
R_INV85	inv85 (view)	luotihist	ssinv-id	ssinv_nr	first	rw

The definitions of the state history relates are presented in Table 7. R_STHIST is the basic relate between the state history coverage PAT-table and the state history base table STATEHIST. The relate attribute is the polygon identifier, which can be a system defined attribute. It only has to match in the relate parts. As in the case of operation history, the relation is one-to-many, because for each state history polygon, there may be several states of the forest.

Relate R_INV85 in Table 7 presents a similar case to the operation history, where a specific operation was extracted by using views and reselect operations. Here, the state of the forest in a specific point of time has been restored by using a similar procedure listed in the following:

1. A view of the table STATEHIST is created containing only those rows that were valid forest states at the beginning of 1985.
   Arc: dbmsexecute luotihist
   >: create view inv85 as select * from statehist where (starting_date < 01-jan-1985)
   >: and (closing_date > 01-jan-1985)
   >: end
2. A relate to the view defined. It is presented in Table 7 (relate R_INV85).

Using the relate it is possible to restore the coverage geometry and topology to the beginning of 1985 (In this example it presents the inventory 1981, because no changes

have been made to the test data between the two inventories). The restoring is done using a dissolving operation similar to the operation history. Coverage state85 represents now the forest state in 1.1.1985. This example lacks the forest description tables of the states. Each state should be related to the description of the forest that was effective during the time span (the data model feature entity types FOREST STAND HISTORY and HISTORY TREE STRATUM should be mapped to relational database tables). This is discussed in Chapter 3.2.7.2.

4 Discussion

4.1 Conceptual modelling

Batini et al. (1992) list the benefits of using conceptual modelling in database design. The two most important are that the schema is independent of the target DBMS and that it can be used after the design process to ease the understanding of the data schemas and applications that use them during the operation and maintenance of the database.

In this study the first benefit can not fully be achieved because the implementation aspects affected the abstraction level of the conceptual schema. On the other hand geographic database implementation is usually based on segregated spatial and attribute databases and the schema of this study supports this kind of implementation better. It can also be more effectively utilized after the design process.

Conceptual models represent a static view of data. The dynamic nature of geographical information can not fully be included in an ER-model. For example, the relationships between lines and polygons (e.g. which lines may be copied to polygon boundaries, which not) can not be properly modelled. The relationships between polygon subclasses were not modelled at all. There may be no benefit in drawing many-to-may-relationships between them, because the relationships can be properly built only by using GIS overlay operations.

There exist two principles for geographic data organization. In the information the layer principle the world is seen as a set of locations with each location having some attribute value. In the opposite, the object principle the world is seen basically empty but filled with a set of geographic objects (points, lines & polygons) (Sijainninhallinnan... 1991). The presented data model contains features of both but its is based on the layer principle with thematic distribution. In the most polygon layers each location has a value of a certain attribute, e.g. land cover class in forest map compartment-layer. On the other hand, there also exists features of the object principle. E.g. terrain object-layer is otherwise empty but there exists objects, e.g. stones, in certain locations.

The thematic distribution of geographic objects into layers assumes that for map presentation purposes the different layers are united into desired combination. Distinct layers can be used as a base for the storage of the geographic objects into other layers. The implementation for the model is the most effective when a layer-oriented GIS-system is used.

Some of the feature attributes are included in the spatial entity (e.g. the attributes of line entity type road). They could also be modelled to separate feature entities. The decision was based on the assumptions of effective database implementation and on the use of the attributes in non-GIS-operations. For example, the growing stock calculations are not spatial operations (only area is needed) and it is more effective to implement them outside the spatial database. On the other hand, the advantage of detaching road type etc. from the spatial entity to separate relational tables is questionable.

4.2 Forest stand history

Space-time composite places heavy demands on attribute-processing procedures and on the linkages between spatial and attribute data. Its side-effects are a) the representation decomposes into progressively smaller units (fragmentation), b) identifiers of changed objects must be altered retroactively (Langran 1992b). In addition to that there must be smooth procedures to eliminate the so-called sliver polygons. Sliver polygons are small polygons that are built in overlay operations when two parallel lines match inexactly.

Different activities of a regeneration chain are seldom carried out on exactly the same area. E.g. soil preparation may only be made for a part of the regeneration area. These activities (e.g. clearing, soil preparation) could be modelled into distinct entity types within the presented framework. For simplification, it was assumed in this study that the series of operations in a regeneration activity are applied to the same regeneration area. This comes partly from the practical fact that the existing FFRI manual regeneration archives contain no information about the spatial diversity of regeneration activities.

In the operation history the time stamp could also be in the base table to get the areas where operations have been made in a certain period of time. The presented structure however enables the use of different types of time stamps with different operations. E.g. with regenerations a more accurrate time stamp may be needed than with cuttings.

The implementation of an application based on the presented forest stand history requires that the GIS-system supports one-to-many relationships between the spatial and attribute data structures or the GIS-system is based on an integrated database with transparent linkages between spatial and attribute data.

4.3 Forest management

Activities of a forest information system are data collection, data input, data storage, updating, manipulation and presentation. A forest information system includes data about forest (and natural) resources, human resources, infrastructures, industries and markets. This study concentrated on the natural resources data contents, flows and operations.

Information about the forest resources are collected in a forest inventory. The aim of forest inventory is to get a sufficient description of the forest for different purposes. This description is then connected to other data sources (basic maps, other relevant data) for planning and operational purposes. After the operational activities the description should be updated for future needs. Monitoring information accumulates from implemented operations and successive inventories.

The results for a forest inventory are presented for spatial units which may also be called description units. Description units can be homogenous in size (a grid of raster cells) or vary in different parts of the invented area (forest stands). The data contents for the description units are obtained from measurement units. Measurement unit may differ in size from the description unit, e.g. for growing stock measurement of a forest stand the unit is normally a relascope plot.

The traditional description unit for operational purposes is a forest stand. It is defined as a delineated forest area homogenous in respect of soil, site and growing stock. Forest stand is however seldom homogenous in respect of all the mentioned factors. Traditionally the most dominant factor is the growing stock and its future treatment needs. So forest stand delineation is often more based on management than biological basis and stands are regarded as treatment units. This partly comes from the fact that forest inventory and planning are mixed up and planning decisions (treatment proposals) can be undertaken by inventory staff.

By using advanced GIS techniques it is theoretically possible to unpack the homogenous forest stands to separate independent soil, site and growing stock description units. These units can be invented separately or the data can be obtained from existing external sources. This can be assumed to result in more detailed and accurate forest description. By storing the different description unit types in separate information layers and by using overlay operations these unit types can be manipulated and new layers can be derived for various purposes. GIS technology allows this kind of functionality but in practical field work it is difficult to make more than one delineation at a time and it is presumed that either a) more than one field inventories (one for each information layer) or b) external data sources where e.g. soil or site maps can be acquired with acceptable resolution and accuracy.

In forest planning numerous data sources are needed to consider all the factors that affect the planning situation. These data elements (conservation areas, special biotopes, roads, wildlife habitats etc.) can be effectively united to the description units by using spatial data processing (GIS) techniques.

One of the most essential requirement for a GIS-based forest management information system is to be able to integrate and manipulate several information layers at the same time. The information may be obtained from various data sources, it may be of varying resolution and may include varying amount of attribute data. The forest stand (description unit) is only one among the layers. The system should be able to process both raster and vector data using overlays, reclassifications, generations and statistical operations in addition to the standard mapping functions. The system development must be based on a proper logical data requirements analysis and modelling. The result of this study may serve as a framework for such modelling.

REFERENCES

Armstrong, M.P. 1988. Temporality in spatial databases. Proceedings of the GIS/LIS88 Conference. San Antonio, 30.11- 2.12.1988. (2): 880-889.

Bachman, C. 1969. Data Structure Diagrams. Data Base 1(2).

Batini, C., Ceri, S & Navathe, S. 1992. Conceptual database design - An Entity-Relationship approach. Benjamin/Cummings. 470 p.

Borsboom, N. & Six Dijkstra, H. 1992. Information management in Dutch forestry and nature conservation. In: Seminar on the use of information systems in forest management. Joint ECE/FAO/ILO Committee on forest technology, management and training, Garpenberg, Sweden, 14-18.9.1992. 10 p.

Bulger, D. & Hunt, H. 1991. The forest management decision support system project. The Forestry Chronicle 67(6): 622-628.

Burrough, P.A. 1986. Principles of geographical information systems for land resources assessment. Clarendon Press, Oxford. 194 p.

Chen, P. 1976. The entity relationship model - toward a unified view of data. ACM transactions on database systems 1(1).

Connors, D.T. 1992. Software development methodologies and traditional and modern information systems. ACM Sigsoft, Software engineering notes 17(2):43-49.

Davis, G.B. and Olson, M.H. 1985. Management information systems. Conceptual foundations, structure and development. McGraw-Hill. 693 p.

Eeronheimo, H. 1993. Paikkatietojärjestelmien käyttö luonnonsuojelualueiden perusselvityksissä - esimerkkinä Pallas-Ounastunturin kansallispuiston kasvillisuuskartoitus. Tutkimuspäivä Rovaniemellä 1993. Metsäntutkimuslaitoksen tiedonantoja 479: 42-51.

Elmasri, R & Navathe, S. 1989. Fundamentals of database systems. Benjamin/Cummings. 802 p.

Fichman, R.G. & Kemerer, C.F. 1992. Object-oriented and conventional analysis and design methodologies. Comparison and critique. Computer, October 1992:22-39.

Ginzberg, M.J. & Stohr, E.A. 1982. Decision support systems: Issues and perspectives. In Ginzberg, M.J., Reitman W & Stohr, E.A. (Eds.): Decision support systems. North-Holland.

A Guide to Information Engineering using the IEF. Computer aided planning, analysis and design. 1990. Texas Instruments. 449 p.

Haavisto, M. (ed.) 1983. Maaperäkartan käyttöopas. Summary: Basic mapping of quaternary deposits in Finland. The Geological Survey of Finland. Guide 10. 80 p + append.

Hull, R. & King, R. 1987. Semantic database modeling: Survey, Applications, and research issues. ACM computing surveys 19(3):201-260.

Juvakka, M. 1993. Metsäntutkimuslaitoksen tutkimusalueiden kuvioittaisen perusinventoinnin maastotyöohje. Metsäntutkimuslaitos, metsänkasvatuksen tutkimusosasto. 15.5.1993. Mimeograph, 31 p + append.

Kaila, E. & Saarenmaa, H. 1990. Tietokoneavusteinen päätöksenteko metsätaloudessa. Summary: Computer-aided decision making in forestry. Folia Forestalia 757. 34 p.

Kennett, R. 1992. Integration of silvicultural record-keeping and geographic information systems. Proceedings of the GIS92 symposium. Vancouver, Canada, 10.-13.2.1992. 6 p.

Klein, H.K. & Hirschheim, R.A. 1987. A comparative framework of data modelling paradigms and approaches. The Computer Journal 30(1): 8-15.

Langran, G. 1989. A review of temporal database research and its use in GIS applications. International Journal of Geographic Information Systems 3(3): 215-232.

- 1992a. Forest monitoring using temporal GIS tecniques. Proceedings of the GIS92 symposium. Vancouver, Canada, 10.-13.2.1992. 6 p.
- 1992b. Time in geographic information systems. Taylor & Francis. 181 p.

Laurini, R. & Thompson, D. 1992. Fundamentals of spatial information systems. The A.P.I.C. Series, Number 37. Academic Press.

Lehan, T.J. 1986. The influence of spatial information on data queries. Proceedings of the ASPRS-ACSM annual convention, Falls Church, Virginia. p. 250-257.

Lehto, K. & Isomäki, A. 1993. Koerekisteri, käyttöopas. Ohjelmaversio 1.1. Metsäntutkimuslaitos 15.3.1993. Mimeograph. 50 p + append.

Loomis, M.E. 1990. The database book. Macmillan publishing company. NewYork. 465 p.

Luonnonsuojelualueiden hoitotoimikunnan mietintö. Komiteanmietintö 1982:26. Helsinki.

Lyytinen, K.1987. A taxonomic perspective of information systems development: Theoretical constructs and recommendations. In: Boland, J & Hirschheim, R.A. (eds.): Critical issues in information systems research. John Wiley & Sons, Chichester, pp. 3-41

Maastotietojen kohdemalli. 1993. Maanmittauslaitoksen julkaisuja 71. 76 p.

Managing tabular data. 1991. ARC/INFO users guide. Environmental Research Institute Inc.

Martin, J. 1990. Information Engineering II. Planning and analysis. Prentice Hall, Englewood Cliffs.

McLaren, R.1990. Establishing a corporate GIS from component datasets - the database issues. Mapping Awareness 4(2): 52-58.

Metsäntutkimuslaitoksen alueiden tutkimuskäyttöä koskeva ohjeisto. 1989b. Metsäntutkimuslaitos. Mimeograph, 6 p + append.

Metsäntutkimuslaitoksen hallinnon tietojenkäsittelyn kehittämissuunnitelma vuosille 1986-90. 1985. Metsäntutkimuslaitoksen tiedonantoja 203, 45 p.

Metsäntutkimuslaitoksen maa-alueohjelma 1990-luvulla. 1993. Metsäntutkimuslaitos. Mimeograph, 19 s. + append.

Metsätalouden tietohallintosuunnitelma. 1991. Metsätalouden tietohallintotyöryhmän muistio. Maa- ja metsätalousministeriö. Työryhmämuistio MMM 1991:2. 28 p. + append.

Nalle-metsäkarttaohjelmisto, v. 3.0. 1984. Metsäntutkimuslaitos & metsähallitus. Mimeograph.

Nalli, A. 1992. Monikäytön suunnittelumenetelmä. Metsätalouden suunnittelun syventävien opintojen tutkielma. Joensuun yliopisto, metsätieteellinen tiedekunta. 65 p.

Nuutinen, T. 1991. Utilizing relational database management system in timber sale planning. In: Saarenmaa, H. (ed.). Current advances in the use of computers in forest research. Workshop of the IUFRO Working Party S4.11-03, Joensuu, Finland, 14.2.1991. Metsäntutkimuslaitoksen tiedonantoja 395: 42-49.

- 1994a. GIS in operational planning of forestry. Dissertation. Manuscript.

- 1994b. GISMELA - paikkatietojärjestelmä metsäsuunnittelussa. In: Niemeläinen, P., Kangas, J. & Päivinen, R. (eds.). Integroidun metsäsuunnittelun menetelmiä ja välineitä. Joensuun yliopisto, metsätieteellinen tiedekunta, tiedonantoja 15: 18-33.

Paananen, R. & Nuutinen, T. 1993. Metsäntutkimuslaitoksen tutkimusmetsien tiedonhallinnan ja suunnittelun kehittäminen. Tutkimuspäivä Rovaniemellä 1993. Metsäntutkimuslaitoksen tiedonantoja 479: 8-20.

Paikkatietojärjestelmien kehittämisprojektin loppuraportti. 1989. Metsähallitus. Mimeograph, 16 p + append.

PATI-maastotyöohje 15.8.1991. 1991. Metsähallitus. Mimeograph, 26 p + append.

Saarenmaa, H., Lehto, K., Kaila, E., Salminen, H. & Pöntinen, J. (eds.) 1990. Metsäntutkimuslaitoksen tietojärjestelmästrategia. Työryhmän mietintö. Metsäntutkimuslaitoksen tiedonantoja 350, 82 p.

Shlaer, S. & Mellor, S. 1988. Object-Oriented Systems Analysis. Yourdon Press, Prentice-Hall. Englewood Cliffs. 108 p + append.

Sijainninhallinnan tietojärjestelmän määrittelyprojekti (SIHA). 1991. Maanmittaushallitus. Loppuraportti. 17.5.1991. 54 p + liitt.

Smith, T.R, Menon, S., Star, J.L. & Estes, J.E. 1987. Requirements and principles for the implementation and construction of large-scale geographic information systems. International Journal of Geographic Information Systems 1(1): 19-35.

Tsichritzis, D. & Klug, A. (eds.) 1978. The ANSI/X3/SPARC DBMS Framework. American Federation of Information Processing Societies Press.

- & Lochovsky, F. 1982. Data models. Prentice Hall. Englewood Cliffs, New Jersey. 381 p.

Turban, E.1990. Decision support and expert systems: Management support systems. Macmillan publishing company. NewYork.

Tutkimusaluetyöryhmän mietintö. 1989a. Metsäntutkimuslaitos. Mimeograph, 17 p + append.

Valtakunnan metsien inventoinnin kenttätyön ohjeet. 1986. Metsäntutkimuslaitos, metsänarvioimisen tutkimusosasto. Mimeograph, 86 p + append.

Van Roessel, J.W. 1987. Design of spatial data structure using the relational normal forms. International Journal of Geographical Information Systems 1(1): 33-50.

Wieringa, R. 1989. Three roles of conceptual models in infomation system design and use. In: Falkenberg, E.D. and Lindgreen, P. (eds.): Information system concepts: An in-depth analysis. Proceedings of the IFIP TC 8 / WG 8.1 working conference on information system concepts, Namur, Belgium, 18-20.10.1989, pp. 31-51.

Virtanen, T. 1989. Informaation lajit ja tietohallinto. Hallinnon tutkimus 3: 180-212.

Worboys, M., Hearnshaw, H., Maguire, D. 1990. Object-oriented data modelling for spatial databases. International Journal of Geographical Information Systems 4(4): 369-383.

Yourdon, E 1989. Modern structured analysis. Yourdon Press, Prentice-Hall. Englewood Cliffs. 672 p.

Seloste

Paikkatietojärjestelmään perustuvan metsätietojärjestelmän tietomalli

Tämä tutkimus on toteutettu osana hanketta, jossa kehitetään Metsäntutkimuslaitoksen tutkimusmetsien operatiivista tietojärjestelmää. Tavoitteena on tuottaa paikkatietojärjestelmään perustuva tutkimusmetsien tiedonhallinta- ja suunnittelujärjestelmä, jolla hallitaan metsävarojen nykytila ja mahdollistetaan metsien tutkimuskäytön integroitu suunnittelu. Järjestelmän päätoimintoja ovat paikkaan sidotun kuvioittaisen metsävaratiedon hallinta, metsäsuunnittelu ja kokeiden hallinta.

Tässä tutkimuksessa tuotettiin järjestelmän perustan muodostavat loogiset mallit tutkimusmetsien toiminnoista (toimintomalli) ja tiedoista (käsitekaavio). Työssä käytettiin IE-systeemityömenetelmää. Laitoksen tietojärjestelmästrategian rajausten pohjalta analysoitiin tutkimusmetsiin kohdistuvat toiminnot ja niiden kehittämistarpeet. Tutkimusmetsistä tarvittavien tietojen rakenne ja sisältö kartoitettiin haastattelemalla tutkijoita ja tutkimusaluehallintoa. Erityistä huomiota kiinnitettiin eri tutkimusalojen tietotarpeisiin. Selvitystyössä käytettiin olemassa olevaa kirjallista materiaalia ja analysoitiin nykyjärjestelmien rakenne ja toiminnot.

Toimintomalli laadittiin IE-menetelmän mukaisena toimintohierarkiana. Käsitekaavio laadittiin menetelmään sovitettua ER-mallia käyttäen. ER-malli on käsitemalli, jolla voidaan kuvata tarkasteltavan alueen (tutkimusmetsien) pienoismaailma joukolla korkean tason loogisia käsitteitä, jotka ovat riippumattomia tietokannan toteutustavasta. Käsitekaavioon määriteltiin järjestelmään sisällytettävät tietokohteet ja niiden väliset yhteydet. Osalla kohteita kuvausta tarkennettiin määrittelemällä myös ominaisuudet ja niiden arvoalueet sekä avainkentät.

Käsitekaaviossa kuvattiin geometriset kohteet (paikkatietoydin), niihin liittyvät ominaisuustietokohteet sekä niiden yhteydet. Paikkatietoydin muodostaa oman kohderyhmänsä, jonka ydinkohteita ovat piste, viiva ja alue. Määritellyt sijaintitiedon kohteet kuvattiin näiden kohteiden alakohdeluokkina. Alue-kohteen alakohteiden avulla on määritelty tarvittavat aluejaot. Vastaavasti viivat ja pisteet ryhmitettiin loogisesti yhteenkuuluviksi tasoiksi.

Sijainti- ja ominaisuustiedon hallinnan toteutus eri järjestelmissä (GIS, RDBMS) on huomioitu käsitekaaviossa. Paikkatietoytimen ulkopuolella olevat kohteet kuvaavat niitä ominaisuustietoja, jotka liittyvät eri sijaintitiedon kohteisiin. Käsitekaaviosta on suoraan pääteltävissä eri kohteiden toteutustapa. Paikkatietoydin-kohderyhmä toteutetaan GIS-tietorakenteina ja muut (ominaisuustiedot) relaatiotietokannan tauluina. Yhteys sijainti- ja ominaisuustietojen välille muodostetaan yksikäsitteisillä avaimilla (esim. kuviotunnus).

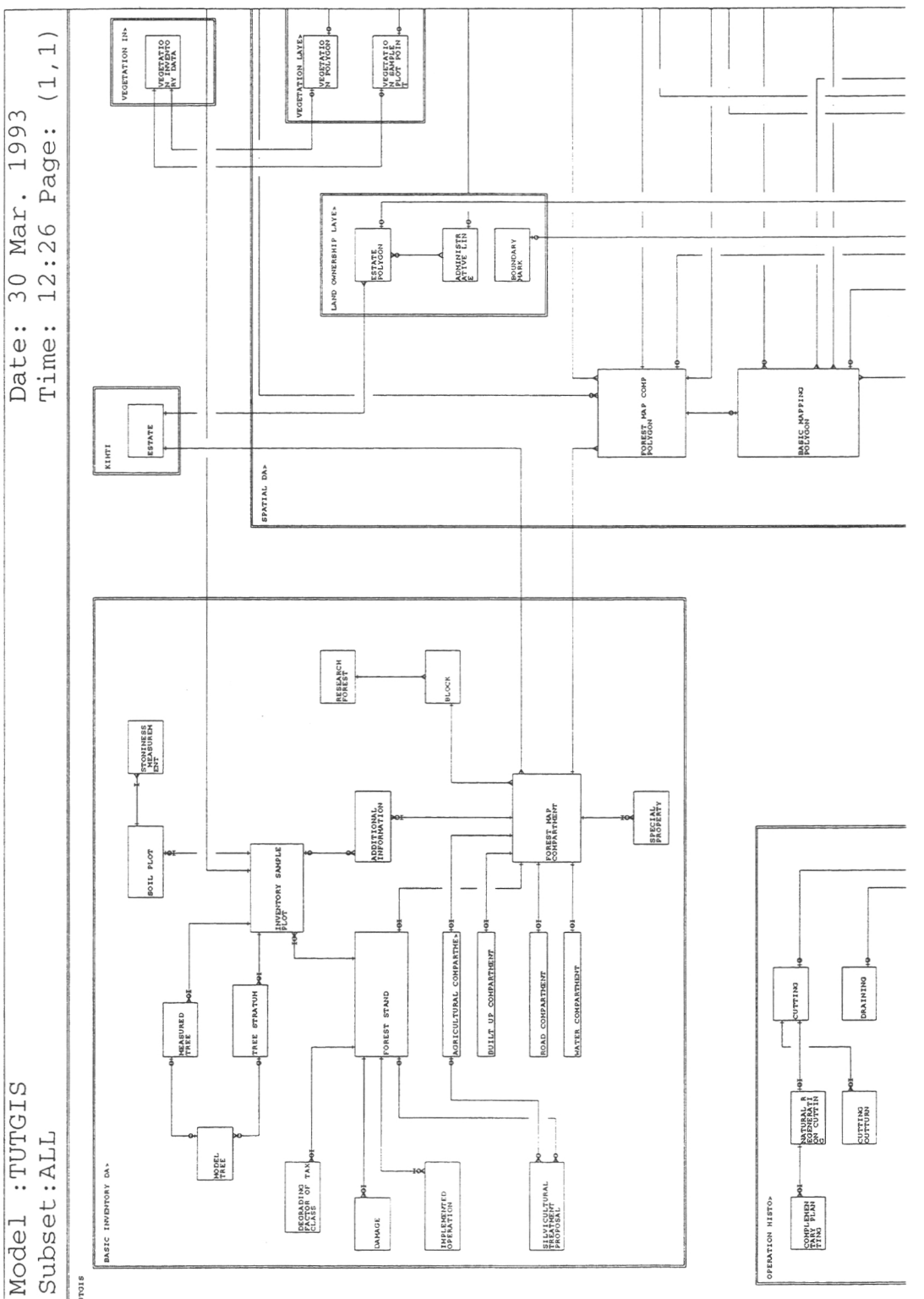
Käsitekaavioon määriteltiin myös ne erilaiset hallinnolliset ja toiminnalliset erityisaluejaot, joita tarvitaan mm. metsäsuunnittelussa. Näitä ovat mm. erilaiset suojelualueet, kaava-alueet, vuokra-alueet, maisemanhoito-, riistanhoito- ja

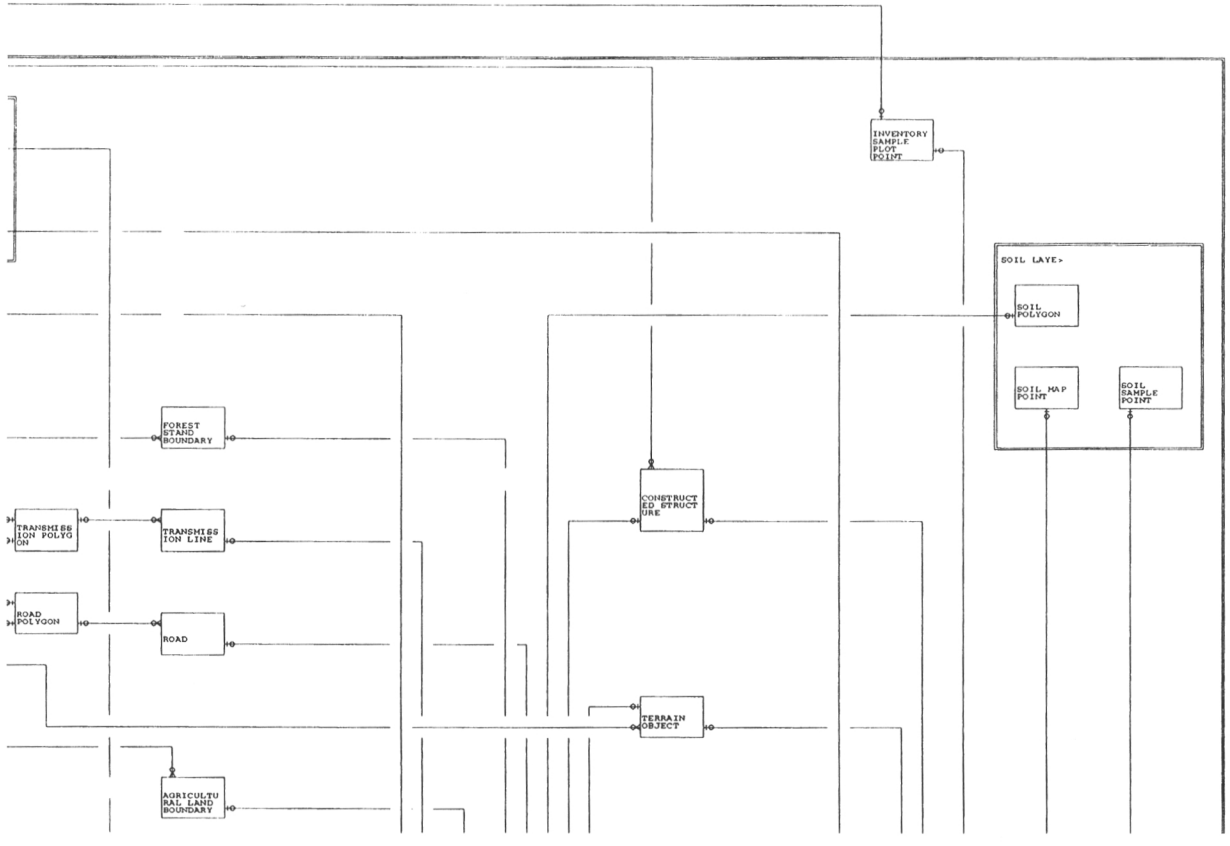
virkistysalueet sekä erityiset biotoopit ja uhanalaisten eliölajien esiintymiskohteet. Erityyppisiä aluejakoja määriteltiin yhteensä noin 20 kpl.

Käsitekaavion pohjalta muodostettiin testitietokanta perusinventoinnin sijainti- ja ominaisuustiedoista. Tietokanta, joka käsittää kuvioihin liittyvät sijaintitiedon kohteet ARC/INFO:n tietokerroksina sekä kuvion ominaisuustiedot Ingres-relaatiotietokannan tauluina, muodostettiin tehtyjen määrittelyjen ja Kivalon tutkimusalueesta kerätyn koeaineiston perusteella. Karttakohteiden digitoinnin yhteydessä voitiin laadittua kohdemallia verifioida ja tehdä siihen tarvittavia muutoksia. Kivalon koeinventointiaineiston ja toimenpidetietojen avulla suunniteltiin ja testattiin myös menetelmää metsiköiden toimenpide- ja tilahistorian hallitsemiseksi.

Tässä raportissa esitellään käsitekaavion rakenne ja sisältö sekä esimerkki metsikön toimenpide- ja tilahistorian hallinnasta. Testitietokannan ja tiedonhallintasovellusten kuvaus julkaistaan TUTGIS-hankkeen tulevissa raporteissa.

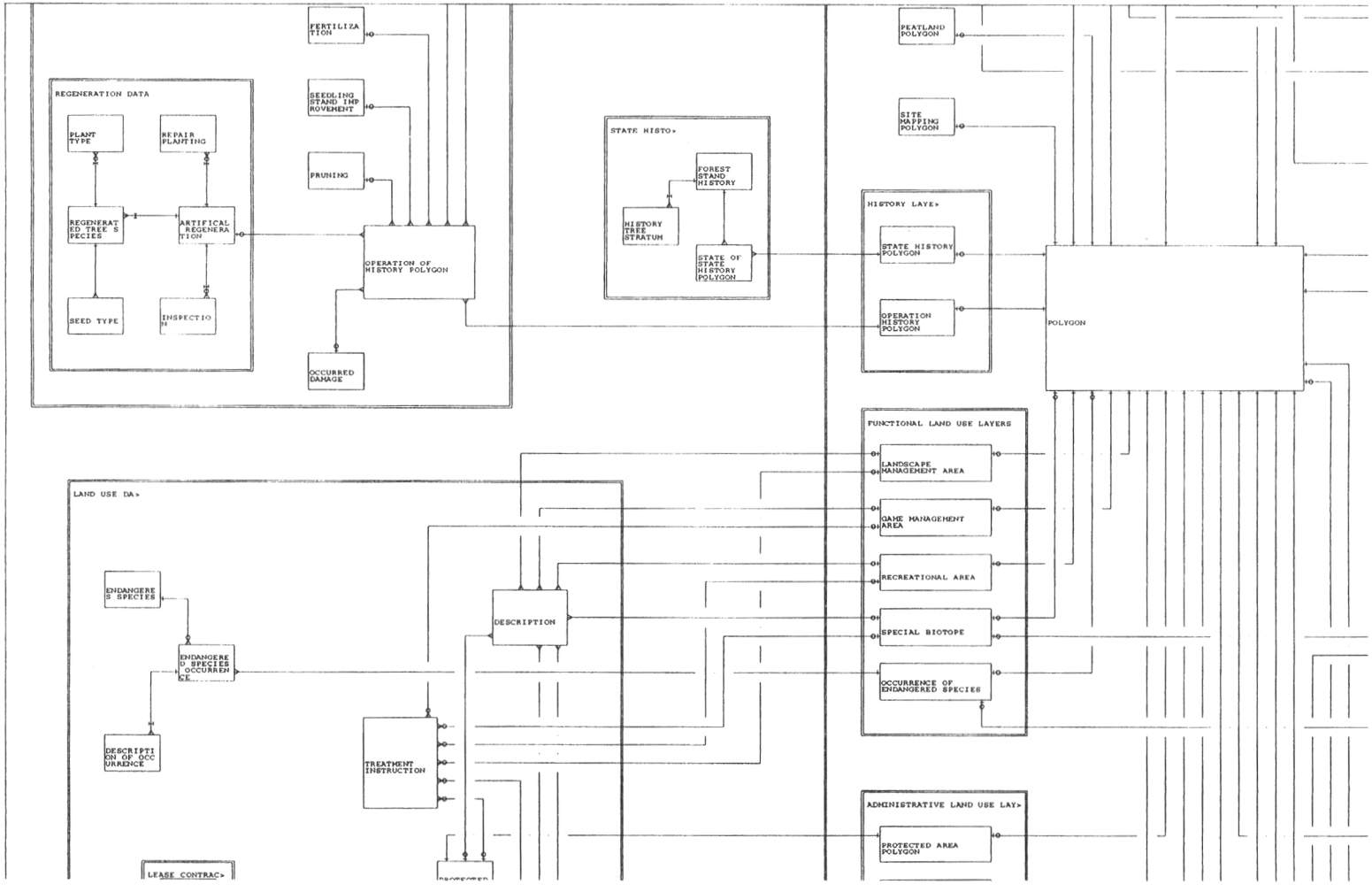
Appendix 1. Conceptual schema of the research forest information system. For constructing the whole schema, see page numbers.




INVENTORY SAMPLE PLOT POINT
SOIL LAYER
SOIL POLYGON
SOIL MAP POINT
SOIL SAMPLE POINT
FOREST STAND BOUNDARY
TRANSMISSION POLYGON
TRANSMISSION LINE
CONSTRUCTED STRUCTURE
ROAD POLYGON
ROAD
TERRAIN OBJECT
AGRICULTURAL LAND BOUNDARY

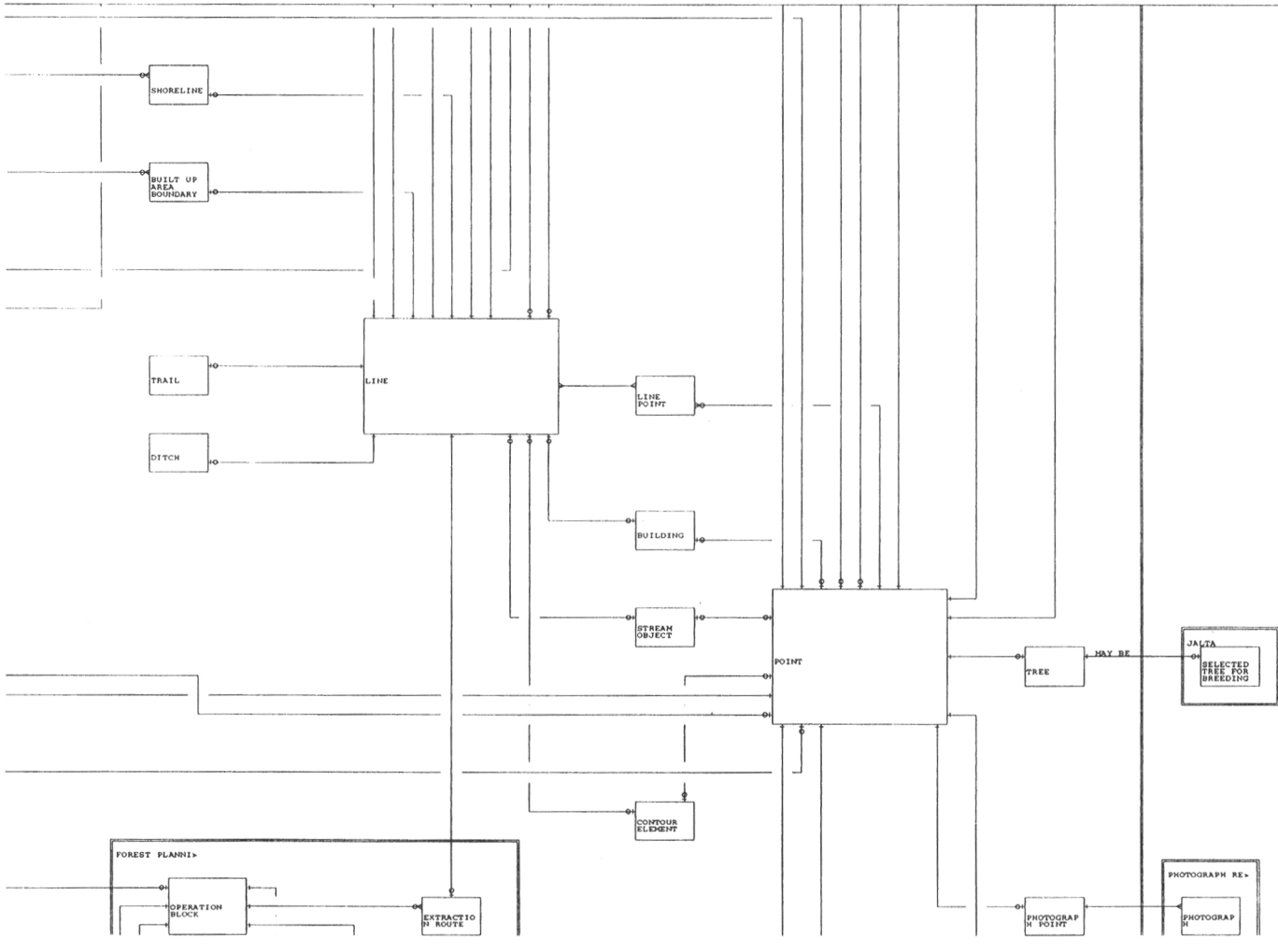
Model :TUTGIS
Subset:ALL

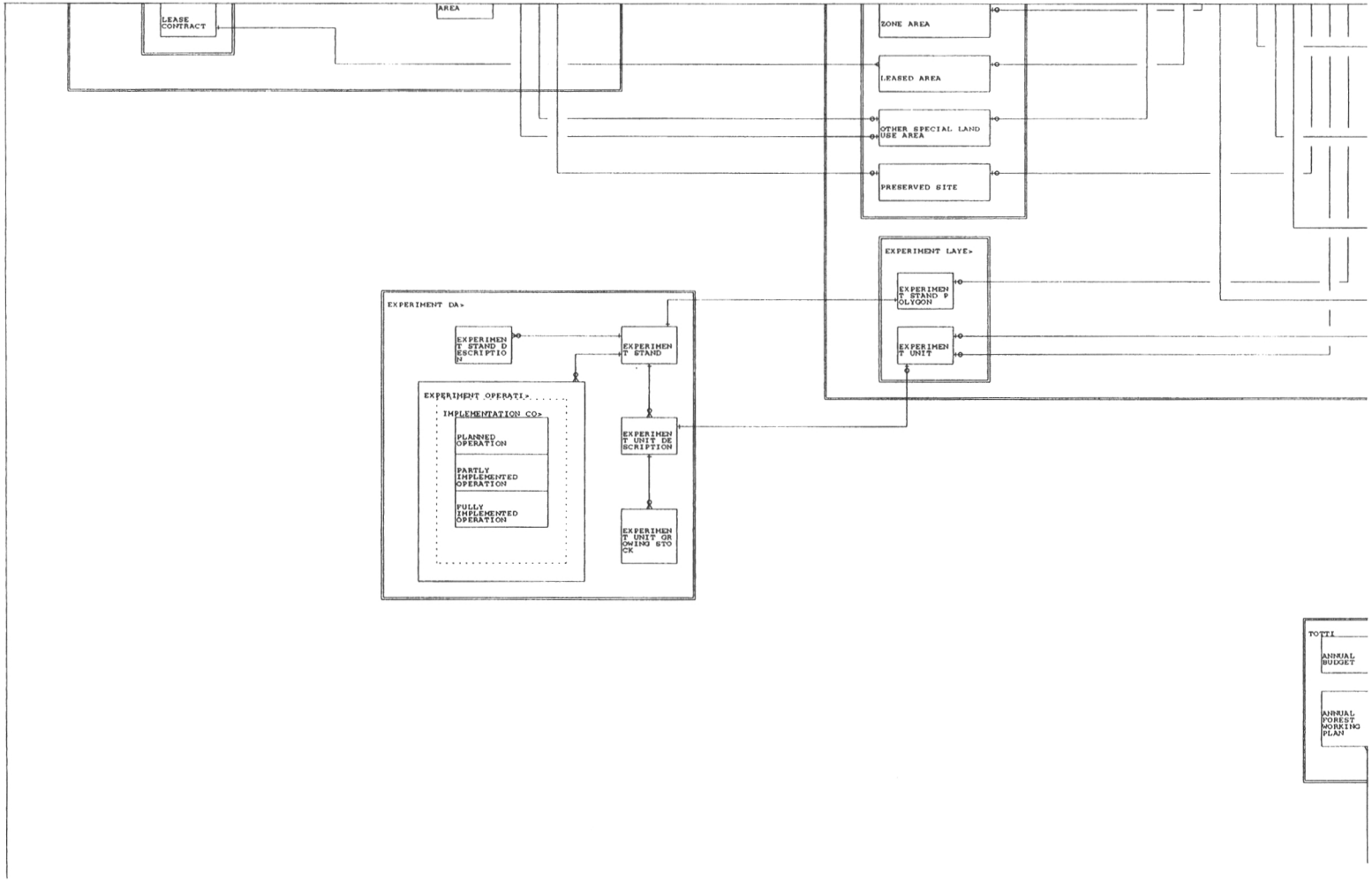
Date: 30 Mar. 1993
Time: 12:26 Page: (2,1)

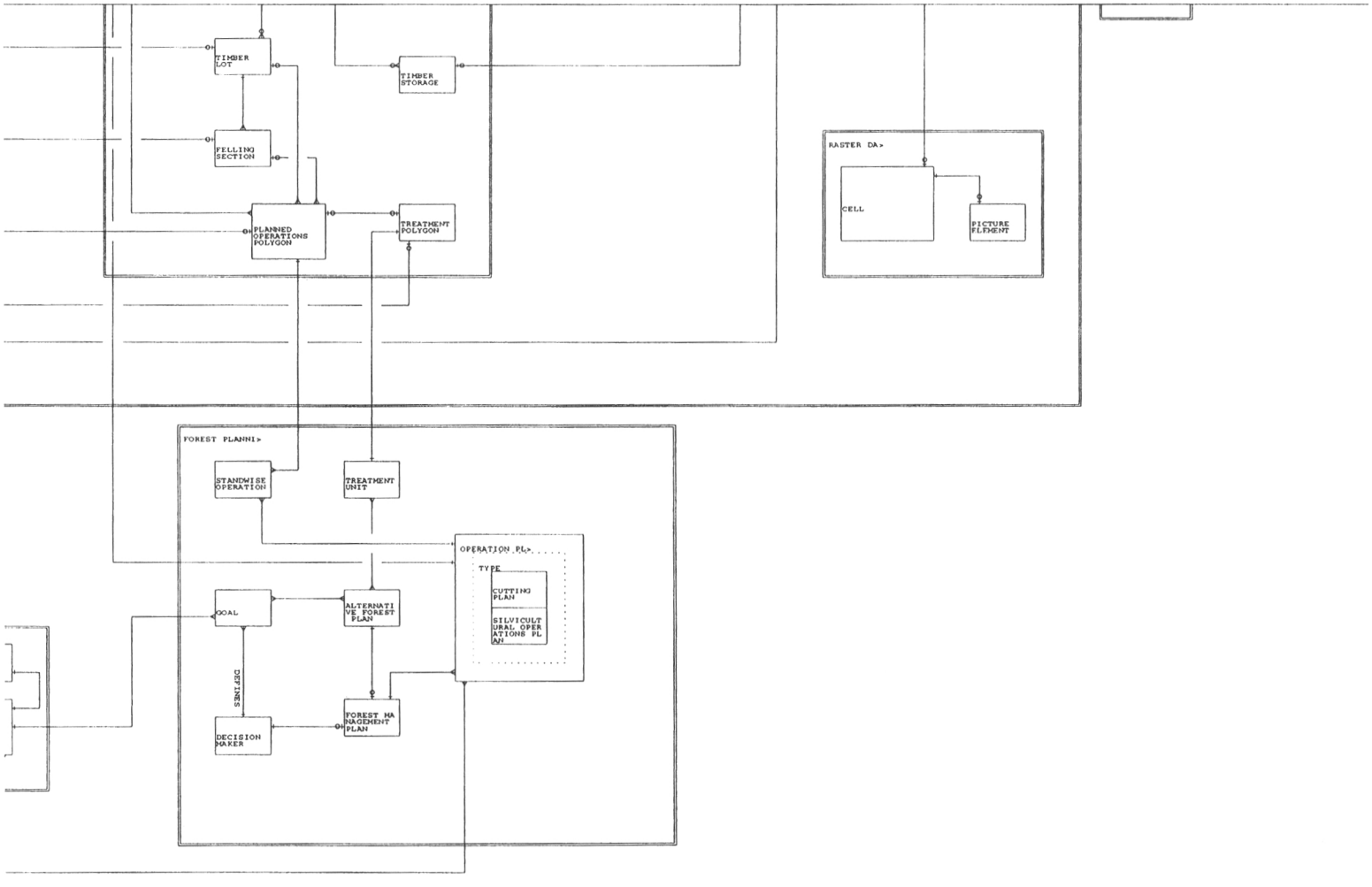


Model :TUTGIS
Subset:ALL

Date: 30 Mar. 1993
Time: 12:26 Page: (2,2)




LEASE CONTRACT
AREA
ZONE AREA
LEASED AREA
OTHER SPECIAL LAND USE AREA
PRESERVED SITE
EXPERIMENT LAYE>
EXPERIMENT STAND POLYGON
EXPERIMENT UNIT
EXPERIMENT DA>
EXPERIMENT STAND DESCRIPTION
EXPERIMENT STAND
EXPERIMENT OPERATI>
IMPLEMENTATION CO>
PLANNED OPERATION
PARTLY IMPLEMENTED OPERATION
FULLY IMPLEMENTED OPERATION
EXPERIMENT UNIT DESCRIPTION
EXPERIMENT UNIT GROWING STOCK
TOTTI
ANNUAL BUDGET
ANNUAL FOREST WORKING PLAN


TIMBER LOT
TIMBER STORAGE
FELLING SECTION
PLANNED OPERATIONS POLYGON
TREATMENT POLYGON
RASTER DA>
CELL
PICTURE ELEMENT
FOREST PLANNI>
STANDWISE OPERATION
TREATMENT UNIT
OPERATION PL>
TYPE
CUTTING PLAN
SILVICULTURAL OPERATIONS PLAN
GOAL
ALTERNATIVE FOREST PLAN
DEFINES
DECISION MAKER
FOREST MANAGEMENT PLAN

Appendix 2. The attributes and relationships of basic inventory data entity types (in alphabetical order).

Entity: ADDITIONAL_INFORMATION
Attributes:
PLOT_NUMBER
TREE_SPECIES
ADDITIONAL_INFORMATION
Relationships:
Sometimes EXPLAINS one INVENTORY_SAMPLE_PLOT
Always EXPLAINS one FOREST_MAP_COMPARTMENT

Entity: AGRICULTURAL_COMPARTMENT
Attributes:
SOIL_TYPE
Relationships:
Sometimes MAY_HAVE many SILVICULTURAL_
TREATMENT_PROPOSAL
Always IS one FOREST_MAP_COMPARTMENT

Entity: BLOCK
Attributes:
NUMBER
NAME
Relationships:
Always BELONGS_TO one RESEARCH_FOREST
Always CONSISTS_OF many FOREST_MAP_COMPARTMENT

Entity: BUILT_UP_COMPARTMENT
Relationships:
Always IS one FOREST_MAP_COMPARTMENT

Entity: DAMAGE
Attributes:
TREE_SPECIES
SOURCE
PHENOMENON_FORM
TIME
GRADE
Relationships:
Always OCCURS_IN one FOREST_STAND

Entity: DEGRADING_FACTOR_OF_TAX_CLASS
Attributes:
DEGRADE_CODE
Relationships:
Always IS_MEASURED_IN one FOREST_STAND

Entity: FOREST_MAP_COMPARTMENT
Attributes:
- RESEARCH_FOREST
- BLOCK_NR
- COMPARTMENT_NR
- SUBCOMPARTM_NR
- ARCKEY
- AREA
- INVENTORY_DATE
- PERSON_IN_CHARGE
- TEMPERATURE_SUM
- HEIGHT_ABOVE_SEA
- LAND_COVER_CLASS
- LAND_USE_CLASS
- COMMUNE_NR
- VILLAGE_NUMBER
- ESTATE_NR_1
- ESTATE_NR_2
- PARCEL_NR

Relationships:
- Sometimes MAY_HAVE many SPECIAL_PROPERTY
- Sometimes MAY_BE one WATER_COMPARTMENT
- Sometimes MAY_BE one ROAD_COMPARTMENT
- Sometimes MAY_BE one BUILT_UP_COMPARTMENT
- Sometimes MAY_BE one AGRICULTURAL_COMPARTMENT
- Sometimes MAY_BE one FOREST_STAND
- Always BELONGS_TO one ESTATE
- Always BELONGS_TO one BLOCK
- Always CONSISTS_OF many
  FOREST_MAP_COMPARTMENT_POLYGON
- Sometimes MAY_BE_DESCRIBED_BY many
  ADDITIONAL_INFORMATION

Identifiers:
- RESEARCH_FOREST
- BLOCK_NR
- COMPARTMENT_NR
- SUBCOMPARTM_NR

Entity: FOREST_MAP_COMPARTMENT_POLYGON
Attributes:
- RESEARCH_FOREST
- BLOCK_NR
- COMPARTMENT_NR
- SUBCOMPARTM_NR
- ARCKEY

Relationships:
- Sometimes MAY_BE_DERIVED_FROM one ROAD_POLYGON

Sometimes MAY_BE_COMPRISED_OF many
CONSTRUCTED_STRUCTURE
Sometimes MAY_BE_COMPRISED_OF many
FOREST_STAND_BOUNDARY
Sometimes MAY_BE_DERIVED_FROM one
TRANSMISSION_POLYGON
Sometimes MAY_BE_DERIVED_FROM one
BASIC_MAPPING_POLYGON
Always IS one POLYGON
Always DELINEATES one FOREST_MAP_COMPARTMENT

Identifiers:
ARCKEY

Entity: FOREST_STAND
Attributes:
SUBCLASS
SOIL_TYPE
HUMUS_LAYER
PEAT_FORMATION
PEAT_DECOMP_GR
DRAINAGE_SIT
SITE_CLASS
TAX_CLASS
LOGGING_ACCESS
DEVELOPMENT_CL
QUALITY_CLASS
REGENERATION_NR
SITE_TYPE
Relationships:
Sometimes MAY_HAVE many
SILVICULTURAL_TREATMENT_PROPOSAL
Always IS one FOREST_MAP_COMPARTMENT
Sometimes MAY_HAVE many
DEGRADING_FACTOR_OF_TAX_CLASS
Sometimes MAY_SUFFER_FROM many DAMAGE
Sometimes MAY_HAVE many IMPLEMENTED_OPERATION
Sometimes CONSISTS_OF many INVENTORY_SAMPLE_PLOT

Entity: IMPLEMENTED_OPERATION
Attributes:
OPERATION_CODE
OPERATION_YEAR
TREE_SPECIES
Relationships:
Always IS_MADE_IN one FOREST_STAND
Identifiers:
OPERATION_YEAR

OPERATION_CODE
IS_MADE_IN FOREST_STAND

Entity: ESTATE
Attributes:
COMMUNE_NR
VILLAGE_NUMBER
ESTATE_NR_1
ESTATE_NR_2
PARCEL_NR
ESTATE_ID
NAME
FOREST_LAND
SCRUB_LAND
WASTE_LAND
AGRICULT_LAND
BUILT_UP_LAND
ROAD
WATER
Relationships:
Always CONSISTS_OF many FOREST_MAP_COMPARTMENT
Always IS_DELINEATED_IN many ESTATE_POLYGON
Identifiers:
PARCEL_NR
ESTATE_NR_2
ESTATE_NR_1
VILLAGE_NUMBER
COMMUNE_NR

Entity: INVENTORY_SAMPLE_PLOT
Attributes:
PLOT_NR
ARCKEY
MEASUREMENT_DATE
TYPE
BASAL_AREA_FACTOR
AREA
COEFFICIENT
SITE_TYPE
DOMINANT_HEIGHT
Relationships:
Sometimes MAY_BE_LOCATED_IN one
INVENTORY_SAMPLE_PLOT_POINT
Sometimes MAY_BE_DESCRIBED_BY many
ADDITIONAL_INFORMATION
Sometimes MAY_CONSIST_OF many MEASURED_TREE
Sometimes MAY_ALSO_BE one SOIL_PLOT
Always IS_MEASURED_IN one FOREST_STAND
Sometimes CONSISTS_OF many TREE_STRATUM

Identifiers:
PLOT_NR
IS_MEASURED_IN FOREST_STAND

Entity: INVENTORY_SAMPLE_PLOT_POINT
Attributes:
PLOT_NR
RESEARCH_FOREST
BLOCK_NR
COMPARTMENT_NR
SUBCOMPARTM_NR
ARCKEY
Relationships:
Always REFERS_TO one INVENTORY_SAMPLE_PLOT
Always MAY_BE one POINT

Entity: MEASURED_TREE
Attributes:
TREE_NR
STRATUM_NR
CANOPY_LAYER
TREE_SPECIES
DIAMETER
HEIGHT
AGE_AT_BH
TOTAL_AGE
DRY_BRANCH_HGHT
Relationships:
Always IS_MEASURED_FROM one INVENTORY_SAMPLE_PLOT
Sometimes IS_MODELLED_INTO one MODEL_TREE
Identifiers:
ID (Primary)
TREE_NR
IS_MEASURED_FROM INVENTORY_SAMPLE_PLOT

Entity: MODEL_TREE
Attributes:
STRATUM_NUMBER
TREE_NUMBER
CALCULATION_DATE
TYPE_OF_MODELING
NUMBER_OF_STEMS
DIAMETER
HEIGHT
AGE_AT_BREAST_HEIGHT
TOTAL_AGE
TREE_VOLUME
TIMBER_VOLUME

PULPWOOD_VOLUME
GROWTH
Relationships:
Sometimes MAY_BE_DERIVED_FROM one MEASURED_TREE
Sometimes MAY_BE_DERIVED_FROM one TREE_STRATUM
Identifiers:
STRATUM_NUMBER
TREE_NUMBER

Entity: RESEARCH_FOREST
Attributes:
NUMBER
NAME
Relationships:
Always IS_DIVIDED_INTO many BLOCK

Entity: ROAD_COMPARTMENT
Relationships:
Always IS one FOREST_MAP_COMPARTMENT

Entity: SILVICULTURAL_TREATMENT_PROPOSAL
Attributes:
DATE_OF_PROPOSAL
PERSON_IN_CHARGE
TREATMENT_CODE
TREE_SPECIES
PROPOSED_YEAR
LAST_YEAR
OPTIONALITY
Relationships:
Sometimes MAY_CONCERN one AGRICULTURAL_COMPARTMENT
Sometimes MAY_CONCERN one FOREST_STAND
Identifiers:
TREATMENT_CODE

Entity: SOIL_PLOT
Attributes:
MEASUREMENT_DATE
ARCKEY
SOIL_TYPE
HUMUS_LAYER
A_HORIZON
B_HORIZON
C_HORIZON
AVG_STONINESS
Relationships:
Always CONTAINS many STONINESS_MEASUREMENT
Always IS_MEASURED_IN one INVENTORY_SAMPLE_PLOT
Identifiers:

IS_MEASURED_IN INVENTORY_SAMPLE_PLOT

Entity: SPECIAL_PROPERTY
Attributes:
PROPERTY_CODE
Relationships:
Always BELONGS_TO one FOREST_MAP_COMPARTMENT
Identifiers:
PROPERTY_CODE
BELONGS_TO FOREST_MAP_COMPARTMENT

Entity: STONINESS_MEASUREMENT
Attributes:
MEASUREMENT_NR
STONINESS
Relationships:
Always IS_MADE_IN one SOIL_PLOT
Identifiers:
MEASUREMENT_NR
IS_MADE_IN SOIL_PLOT

Entity: TREE_STRATUM

Attributes:
STRATUM_NR
TREE_SPECIES
CANOPY_LAYER
REG_METHOD
NUMBER_OF_STEMS
BASAL_AREA
MEAN_DIAMETER
MIN_DIAMETER
MAX_DIAMETER
MEAN_HEIGHT
MEAN_AGE_AT_BH
MEAN_TOTAL_AGE
MEAN_BRANCH_HGHT
Relationships:
Sometimes MAY_BE_MODELLED_INTO many MODEL_TREE
Always IS_MEASURED_IN one INVENTORY_SAMPLE_PLOT
Identifiers:
STRATUM_NR
IS_MEASURED_IN INVENTORY_SAMPLE_PLOT

Entity: WATER_COMPARTMENT
Relationships:
Always IS one FOREST_MAP_COMPARTMENT

Appendix 3. The attributes and relationships of special land use data entity types.

Entity: LEASED_AREA
Attributes:
- COMMUNE_NR
- VILLAGE_NUMBER
- ESTATE_NR_1
- ESTATE_NR_2
- AREA_NUMBER
- AREA_TYPE
- CONTRACT_NUMBER

Relationships:
- Always IS one POLYGON
- Always REFERS_TO one LEASE_CONTRACT

Identifiers:
- AREA_TYPE
- CONTRACT_NUMBER

Entity: OTHER_SPECIAL_LAND_USE_AREA
Attributes:
- RESEARCH_FOREST
- AREA_NR
- SPECIAL_AREA_CLASS

Relationships:
- Always IS_DESCRIBED_BY many DESCRIPTION
- Sometimes MAY_HAVE many TREATMENT_INSTRUCTION
- Always IS one POLYGON

Entity: PRESERVED_SITE
Attributes:
- ID
- NAME

Relationships:
- Always IS_DESCRIBED_BY many DESCRIPTION
- Always IS one POINT

Entity: PROTECTED_AREA_POLYGON
Attributes:
- LSA_IDENTITY_CODE
- SUBAREA_CODE

Relationships:
- Always IS one POLYGON
- Always CONTAINS one PROTECTED_AREA

Entity: ZONE_AREA
Attributes:
- COMMUNE_NR

ZONE_PLAN_ID
LANGUAGE
ZONE_PLAN_TYPE
ZONE_PLAN_STAGE
LAND_USE_CLASS
Relationships:
Always IS one POLYGON

Entity: GAME_MANAGEMENT_AREA
Attributes:
RESEARCH_FOREST
AREA_NR
Relationships:
Sometimes MAY_HAVE many TREATMENT_INSTRUCTION
Always IS_DESCRIBED_IN many DESCRIPTION
Always IS one POLYGON

Entity: LANDSCAPE_MANAGEMENT_AREA
Attributes:
RESEARCH_FOREST
AREA_NR
Relationships:
Sometimes MAY_HAVE many TREATMENT_INSTRUCTION
Always IS_DESCRIBED_IN many DESCRIPTION
Always IS one POLYGON

Entity: OCCURRENCE_OF_ENDANGERED_SPECIES
Attributes:
RESEARCH_FOREST
OCCURRENCE_NR
Relationships:
Sometimes MAY_BE one POINT
Always HAS many ENDANGERED_SPECIES_OCCURRENCE
Sometimes MAY_BE one POLYGON

Entity: RECREATIONAL_AREA
Attributes:
RESEARCH_FOREST
AREA_NR
Relationships:
Always IS_DESCRIBED_BY many DESCRIPTION
Always IS one POLYGON
Sometimes MAY_HAVE many TREATMENT_INSTRUCTION

Entity: SPECIAL_BIOTOPE
Attributes:
RESEARCH_FOREST
BIOTOPE_NR

Relationships:
- Sometimes MAY_BE one POINT
- Sometimes MAY_BE one POLYGON
- Sometimes MAY_HAVE many TREATMENT_INSTRUCTION
- Always IS_DESCRIBED_BY many DESCRIPTION

Entity: DESCRIPTION
Relationships:
- Sometimes MAY_DESCRIBE one PRESERVED_SITE
- Sometimes MAY_DESCRIBE one RECREATIONAL_AREA
- Sometimes MAY_DESCRIBE one OTHER_SPECIAL_LAND_USE_AREA
- Sometimes MAY_DESCRIBE one PROTECTED_AREA
- Sometimes MAY_DESCRIBE one SPECIAL_BIOTOPE
- Sometimes MAY_DESCRIBE one GAME_MANAGEMENT_AREA
- Sometimes MAY_DESCRIBE one LANDSCAPE_MANAGEMENT_AREA

Entity: DESCRIPTION_OF_OCCURRENCE
Attributes:
- TYPE_OF_DESCRIPTION
- ROW_NUMBER
- DESCRIPTION_TEXT

Relationships:
- Always REFERS_TO one ENDANGERED_SPECIES_OCCURRENCE

Identifiers:
- ROW_NUMBER
- TYPE_OF_DESCRIPTION
- REFERS_TO ENDANGERED_SPECIES_OCCURRENCE

Entity: ENDANGERED_SPECIES_OCCURRENCE
Attributes:
- RESEARCH_FOREST
- OCCURRENCE_NR
- SPECIES_NR
- OCCURRENCE_ID
- TREATMENT_URGENCY_CODE
- PRESERVATION_URGENCY_CODE

Relationships:
- Always HAS many DESCRIPTION_OF_OCCURRENCE
- Always HAS one ENDANGERES_SPECIES
- Always OCCURS_AT one OCCURRENCE_OF_ENDANGERED_SPECIES

Identifiers:
- OCCURRENCE_ID
- SPECIES_NR

Entity: ENDANGERES_SPECIES

Attributes:

SPECIES_NR
RUBIN_CODE
RUBIN_CODE_LIST_NUMBER
LATIN_NAME
FINNISH_NAME
DECISION_CODE
STAGE_CODE
REASON_FOR_BEING_ENDANGERED
BIOTOPE

Relationships:

Sometimes MAY_HAVE many
ENDANGERED_SPECIES_OCCURRENCE

Identifiers:

SPECIES_NR

Entity: PROTECTED_AREA

Attributes:

LSA_IDENTITY_CODE
SUBAREA_CODE
NAME
TYPE
SUBTYPE
LAND_USE_CLASS
OPERATION_CLASS
ACCESSIBILITY_CLASS

Relationships:

Always IS_DESCRIBED_BY many DESCRIPTION
Sometimes MAY_HAVE many TREATMENT_INSTRUCTION
Always REFERS_TO one PROTECTED_AREA_POLYGON

Entity: TREATMENT_INSTRUCTION

Relationships:

Sometimes MAY_BE_APPLIED_TO one
OTHER_SPECIAL_LAND_USE_AREA
Sometimes MAY_BE_APPLIED_TO one PROTECTED_AREA
Sometimes MAY_BE_APPLIED_TO one SPECIAL_BIOTOPE
Sometimes MAY_BE_APPLIED_TO one RECREATIONAL_AREA
Sometimes MAY_BE_APPLIED_TO one GAME
_MANAGEMENT_AREA
Sometimes MAY_BE_APPLIED_TO one
LANDSCAPE_MANAGEMENT_AREA

Appendix 4. The attributes and relationships of forest stand history entity types.

Entity: OPERATION_HISTORY_POLYGON
Attributes:
ID
Relationships:
Always IS one POLYGON
Always HAS many OPERATION_OF_HISTORY_POLYGON

Entity: STATE_HISTORY_POLYGON
Attributes:
ID
Relationships:
Always HAS many STATE_OF_STATE_HISTORY_POLYGON
Always IS one POLYGON

Entity: ARTIFICIAL_REGENERATION
Attributes:
REGENERATION_NR
REGENERATION_DATE
AREA
HEIGHT_ABOVE_SEA
ASPECT
SLOPE
SITE_TYPE
SITE_CLASS
SOIL_TYPE
EXPERIMENT_NR
FOREST_DESCRIPTION
CLEARING
YEAR_OF_CLEARING
PRESCRIBED_BURNING
YEAR_OF_BURNING
SOIL_PREPARATION_METHOD
DATE_OF_SOIL_PREPARATION
REGENERATION_METHOD
PLANTING_DENSITY
AMOUNT_OF_MATERIAL
WEATHER
ADDITIONAL_INFORMATION
Relationships:
Sometimes MAY_BE_CHECKED_IN many INSPECTION
Sometimes MAY_NEED many REPAIR_PLANTING
Always USES many REGENERATED_TREE_SPECIES
Always IS many OPERATION_OF_HISTORY_POLYGON

Identifiers:
REGENERATION_NR

Entity: COMPLEMENTARY_PLANTING
Attributes:
DATE
SEED_ID
ORIGIN_OF_SEED
REGENERATION_METHOD
AMOUNT_OF_MATERIAL
Relationships:
Always IS_MADE_IN one NATURAL
_REGENERATION_CUTTING
Identifiers:
DATE
IS_MADE_IN NATURAL_REGENERATION_CUTTING

Entity: CUTTING
Attributes:
OPERATION_CODE
AREA
DATE
CUTTING_METHOD
ADDITIONAL_INFORMATION
Relationships:
Sometimes MAY_BE one NATURAL_REGENERATION_CUTTING
Sometimes HAS many CUTTING_OUTTURN
Always IS many OPERATION_OF_HISTORY_POLYGON
Identifiers:
OPERATION_CODE

Entity: CUTTING_OUTTURN
Attributes:
TIMBER_ASSORTMENT
OUTTURN
Relationships:
Always IS_CUT_IN one CUTTING
Identifiers:
TIMBER_ASSORTMENT
IS_CUT_IN CUTTING

Entity: DRAINING
Attributes:
OPERATION_CODE
DRAINING_METHOD
AREA
DATE
EXPERIMENT_NR

Relationships:
Always IS many OPERATION_OF_HISTORY_POLYGON
Identifiers:
OPERATION_CODE

Entity: FERTILIZATION
Attributes:
OPERATION_CODE
DATE
AREA
METHOD
TYPE_OF_FERTILIZER
AMOUNT_OF_FERTILIZER
Relationships:
Always IS many OPERATION_OF_HISTORY_POLYGON
Identifiers:
OPERATION_CODE

Entity: INSPECTION
Attributes:
INSPECTION_NR
INSPECTION_DATE
PERSON_IN_CHARGE
RESULT
Relationships:
Always IS_MADE_IN one ARTIFICIAL_REGENERATION
Identifiers:
INSPECTION_NR
IS_MADE_IN ARTIFICIAL_REGENERATION

Entity: NATURAL_REGENERATION_CUTTING
Attributes:
NUMBER_OF_SEED_TREES
CLEARING
YEAR_OF_CLEARING
SOIL_PREPARATION_METHOD
DATE_OF_SOIL_PREPARATION
Relationships:
Sometimes MAY_NEED many COMPLEMENTARY_PLANTING
Always IS one CUTTING
Identifiers:
IS CUTTING

Entity: OCCURRED_DAMAGE
Attributes:
OPERATION_CODE
AREA
TREE_SPECIES
SOURCE

PHENOMENON_FORM
GRADE
STARTING_YEAR
CLOSING_YEAR

Relationships:

Always IS many OPERATION_OF_HISTORY_POLYGON

Identifiers:

OPERATION_CODE

Entity: OPERATION_OF_HISTORY_POLYGON

Attributes:

POLYGON_ID
RESEARCH_FOREST
BLOCK_NR
AREA
OPERATION_TYPE
OPERATION_CODE

Relationships:

Sometimes MAY_BE one DRAINING
Sometimes MAY_BE one PRUNING
Sometimes MAY_BE one FERTILIZATION
Sometimes MAY_BE one CUTTING
Sometimes MAY_BE one SEEDLING_STAND_IMPROVEMENT
Sometimes MAY_BE one OCCURRED_DAMAGE
Sometimes MAY_BE one ARTIFICIAL_REGENERATION
Always IS_MADE_IN one OPERATION_HISTORY_POLYGON

Identifiers:

OPERATION_CODE
OPERATION_TYPE
POLYGON_ID

Entity: PLANT_TYPE

Attributes:

NURSERY
DATE_OF_BIRTH
PLANT_TYPE
PLANT_CLASS
HANDLING_OF_PLANTS
CONDITION_OF_PLANTS
MEAN_HEIGHT

Relationships:

Always IS_USED_IN one REGENERATED_TREE_SPECIES

Identifiers:

PLANT_TYPE
IS_USED_IN REGENERATED_TREE_SPECIES

Entity: PRUNING

Attributes:

OPERATION_CODE

DATE
AREA
TREE_SPECIES
NUMBER_OF_TREES
HEIGHT
MEAN_DIAMETER

Relationships:
Always IS many OPERATION_OF_HISTORY_POLYGON

Identifiers:
OPERATION_CODE

Entity: REGENERATED_TREE_SPECIES

Attributes:
SPECIES_NR

Relationships:
Sometimes MAY_HAVE many PLANT_TYPE
Always INCLUDES many SEED_TYPE
Always IS_USED_IN one ARTIFICIAL_REGENERATION

Identifiers:
SPECIES_NR
IS_USED_IN ARTIFICIAL_REGENERATION

Entity: REPAIR_PLANTING

Attributes:
DATE
TREE_SPECIES
SEED_ID
ORIGIN_OF_SEED
REGENERATION_METHOD
AMOUNT_OF_MATERIAL

Relationships:
Always IS_USED_IN one ARTIFICIAL_REGENERATION

Identifiers:
TREE_SPECIES
DATE
IS_USED_IN ARTIFICIAL_REGENERATION

Entity: SEED_TYPE

Attributes:
SEED_ID
ORIGIN_OF_SEED
SITE_LOCATION
GERMINATION_PERCENTAGE
SEED_QUALITY
DESCENT

Relationships:
Always IS_USED_IN one REGENERATED_TREE_SPECIES

Entity: SEEDLING_STAND_IMPROVEMENT

Attributes:
- OPERATION_CODE
- DATE
- AREA
- METHOD
- NUMBER_OF_CUT_TREES
- NUMBER_OF_LEFT_TREES
- TYPE_OF_HERBICIDE
- AMOUNT_OF_HERBICIDE

Relationships:
- Always IS many OPERATION_OF_HISTORY_POLYGON

Identifiers:
- OPERATION_CODE

Entity: FOREST_STAND_HISTORY

Attributes:
- RESEARCH_FOREST
- BLOCK_NR
- COMPARTMENT_NR
- SUBCOMPARTM_NR
- STARTING_DATE
- CLOSING_DATE
- AREA
- PERSON_IN_CHARGE
- LAND_COVER_CLASS
- LAND_USE_CLASS
- SUBCLASS
- SOIL_TYPE
- PEAT_FORMATION
- DRAINAGE_SIT
- SITE_CLASS
- TAX_CLASS
- LOGGING_ACCESS
- DEVELOPMENT_CL
- QUALITY_CLASS
- SITE_TYPE

Relationships:
- Always CONTAINS many HISTORY_TREE_STRATUM
- Always HAS many STATE_OF_STATE_HISTORY_POLYGON

Identifiers:
- STARTING_DATE
- SUBCOMPARTM_NR
- COMPARTMENT_NR
- BLOCK_NR
- RESEARCH_FOREST

Entity: HISTORY_TREE_STRATUM

Attributes:
- TREE_SPECIES

CANOPY_LAYER
NUMBER_OF_STEMS
BASAL_AREA
MEAN_DIAMETER
MIN_DIAMETER
MEAN_HEIGHT
MEAN_AGE_AT_BH
MEAN_TOTAL_AGE
VOLUME
TIMBER_VOLUME
PULPWOOD_VOLUME

Relationships:

Always IS_INCLUDED_IN one FOREST_STAND_HISTORY

Identifiers:

CANOPY_LAYER
TREE_SPECIES
IS_INCLUDED_IN FOREST_STAND_HISTORY

Entity: STATE_OF_STATE_HISTORY_POLYGON

Attributes:

POLYGON_ID
RESEARCH_FOREST
BLOCK_NR
COMPARTMENT_NR
SUBCOMPARTM_NR
STARTING_DATE
CLOSING_DATE
AREA

Relationships:

Always REFERS_TO one FOREST_STAND_HISTORY
Always REFERS_TO one STATE_HISTORY_POLYGON

Identifiers:

STARTING_DATE
SUBCOMPARTM_NR
COMPARTMENT_NR
BLOCK_NR
RESEARCH_FOREST
POLYGON_ID

ISBN 951-40-1357-3
ISSN 0358-4283