4 Data Analysis

4.1 Database

Edgar Kaufmann

4.1.1 Storage of the NFI Data

The NFI data consists of four different types of data: 1) Terrestrial assessed raw data (assessments of site and stand attributes, and tree measurements); 2) Raw data that were assessed on aerial photographs (e.g., land-use categories, mean stand height); 3) Data from external data sources (e.g., maps); and 4) Variables that are calculated from raw data of diverse origin (e.g., site quality, stem volume).

The data required for the analysis of the first and second NFI inventory were stored in a relational database (ORACLE system). A relational database system ensures data integrity and provides various tools to describe and manage the data. It allows the user to conveniently access the data with the help of a user-friendly interface and an easy to comprehend query language (SQL). The relationships between all objects are clearly and precisely defined. The data description corresponds at any given time inevitably and completely to the data structure and is managed by the system itself. The high degree of transparency and the easy way to link objects with each other facilitates in checking the data considerably.

4.1.2 Architecture of Relational Database Systems

Data in relational databases are strictly recorded in two-dimensional tables (consisting of rows and columns). Each column of a table is reserved for one attribute (variable). Each attribute has a name and contains exactly one value per row. The number of attributes in a row (record) is fixed. A row consists of global attributes which have key functions and occur in several tables, and local attributes which only occur in one table (for the structure of tables in relational database see for example FLEMING and VON HALLE 1989). Tables are linked to one another by global key attributes which can be found in each of the linked tables (for an example of the rules of logical connection and so-called normalized relational database see KORTH and SILBERSCHATZ 1986).

A key is composed of one or several attributes and is used to either identify a row as a search key within a table, or to link tables. Local attributes depend clearly and completely on the value of the identification key of a table.

Each table has an identification key (primary key). A specific combination of attribute values of a primary key exists exactly once in a table. All attribute values of global attributes in a table also exist in hierarchically superordinated tables. If a primary key is composed of several variables, all of which must be exported into the tables intended to be linked, usually an artificial key in the form of a number is generated, which is exported instead of the entire set of variables.

A specific value of a search key, which is composed of global attributes and which can have more than one occurrence with the same key value in a table, exists exactly once in a hierarchically superordinated table.

Logical, physical, and external structures are separated from each other in a relational database system and can be changed independently from one another (see for example, YANNAKOUDAKIS 1988). Logical structures encompass, among other things, the description of tables, the relationship of tables to one another, definition of data types and the domain (valid range) of attribute values, the definition of key structures, and integrity criteria. Physical structures refer, among others, to the allocation of computer memory, physical data

management, access algorithms and the corresponding address tables. The term "external structures" refers to special user views that are not implemented in the system and which are virtually generated (e.g., virtual placement of data elements that are actually located in several different tables into one table).

With the implementation of the database, a description of the data structure, the so-called "data dictionary" is created within the system itself (see for example WERTZ 1986). This description represents the condition of the database exactly and cannot be changed by the database administrator (who is responsible for the design and update of the database). Logical, physical, and external structures are stored in system tables and can be retrieved at any time. Categorical data are stored in the NFI database in coded form. The meaning of the codes for these variables is explained in separate tables. This part of the data description has to be managed by the database administrator.

The physical access to the data in a ORACLE database system is index sequential. Logical and physical addresses are managed with address tables (with binary tree structure, e.g., see KORTH and SILBERSCHATZ 1986). Such tables are explicitly generated by the administrator by indexing key variables. This significantly expedites the data access and is indispensable for an efficient data retrieval. Within an index sequential access, an entire block of data (usually substantially more than one individual record), which is stored physically close to each other, is retrieved. Thus, the access time is also decreased when records with logically related values of keys, which are often used for the access, are also physically stored close to one another.

4.1.3 Logical Structure of the NFI Database

Figure 1 shows the logical structure of the NFI database. The data can be divided into three hierarchical levels (I–III) with sub-levels (IIa, IIb, IIIa). Each box represents a table. A box contains the table name (in capital letters; for the content see Table 1), the global attributes of a table, which together form the primary key of the table (non-italic), and attributes, which represent an important search key (italic). Bold printed variable names refer to global key attributes (see Table 2), the other names refer to local key attributes (see Table 3). The levels have a "1 to mc" relationship to each other. This means that for each record ("1") in a table at a certain level, there exist several records ("m" for many), one record, or no records ("c" for conditional) in a subordinated table. Conversely, for each record in a table at a certain level, there exists exactly one record in a superordinated table. Hierarchical relationships are created using the tables in the dark-gray tables at the very left side. Through a dark-gray marked table, all tables in the lower level can be linked to one another. Cross-indexing at the same level is a "1 to c" relationship. For each record in the table on side "1" there exists one or no record on side "c" (i.e., the primary key values of the table on the "c" side are a subset of those in the table on the "1" side).

4.1.3.1 Data Levels

Level I contains data that do not depend on a certain inventory. They do, however, depend on the geographic location of a sample point and have one record per sample point. The tables in the sub-level Ia do not depend on the inventory either, but have many records per sample point.

Level II contains inventory specific sample plot data that have one record per sample point and per inventory. The tables in the sublevel IIa contain several records per plot and inventory (e.g., several young growth plots per sample plot, several timber extraction phases per sample point). The young growth data at the sub-level IIb have several records per sample point, inventory, and young growth plot. The young growth tables at the sub-level IIb are linked with the table JWSALFI (Figure 1 light-gray) at the sub-level IIa.

Level III contains individual tree data with one record per sample plot, inventory, and tree. The table "BA" is linked through the search key "clnr,invnr" to table "WA." It is linked through the primary key "invnr, banr" to table "TB" at the same level and to the tables on level IIIa. Level IIIa contains damages, remarks, and assortments with several possible records per tree.

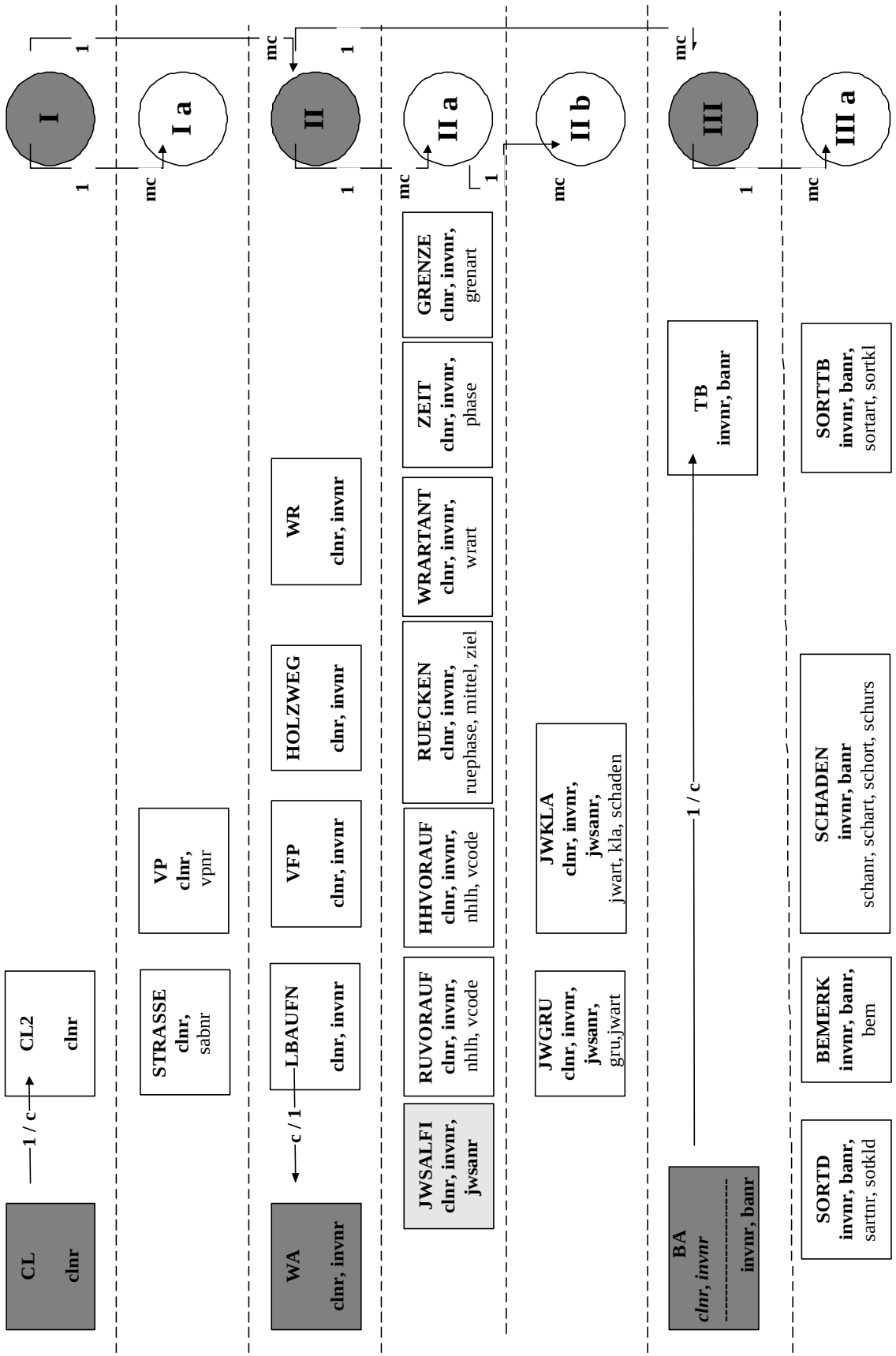

Figure 1. Logical structure of the NFI database. I-III: hierarchical levels. Ia, IIa, IIb, IIIa: sublevels. Capitalized letters: Table names (see Table 1). In bold small letters: global key attributes (see Table 2). In normal small letters: local key attributes. Italicize: Important search key.

Table 1. Tables of the NFI database

Name of table	Content
CL	Contains information that refers to a certain geographic location and that is independent of individual inventories (e.g., canton, elevation above sea level, x-coordinate, y-coordinate, production region). These are not project or NFI specific data. These data exist in the 500m-grid.
CL2	Contains mainly NFI specific attributes of the same type as in CL, often attributes that are derived from external data sources (e.g., site quality TI).
STRASSE	Contains information about the position, classification, and length of those roads relevant for the timber harvest in an area of 500 meters x 500 meters around the sample plot.
VP	Contains the permanently marked points for the terrestrial assessed permanent forest sample plots.
WA	Contains site and stand attributes of the terrestrial sample plots that are related to the sample plot and are specific to the inventory, information about timber harvest, and forest ownership. Contains also other attributes of the sample plot that are calculated from individual tree measures like timber volume per hectare and number of stems per hectare.
LBAUFN	Contains data assessed in aerial photographs, which are required for the NFI analysis. Contains stand information, information about the topography, and the position of the sample plot center.
VFP	Contains the permanently marked points of the terrestrial forested sample plots used in the inventory.
HOLZWEG	Contains data derived from the forest transportation survey (e.g., transport distance).
WR	Contains data that describe the forest edge.
JWSALFI	Contains sample plot related to young growth data, which were assessed for each young growth sample plot (e.g., type of regeneration).
RUVORAUF	Contains input data necessary to calculate the expenditure for timber removal as well as the calculated expenditure.
HHVORAUF	Contains input data necessary to calculate the expenditure for timber harvest as well as the calculated expenditure.
RUECKEN	Contains the information about timber removal from the enquiry at the forest service.
WRARTANT	Contains the proportions with which the different tree species occur at the forest edge.
ZEIT	Contains the time expenditure that was needed for the different work steps at the terrestrial sample plots.
GRENZE	Contains information about the position and the type of the borders that were assessed for the terrestrial sample plot.
JWGRU	Contains information about the crown closure of the young growth plots.
JWKLA	Contains the young growth – individual tree information.
BA	Contains for all trees with a DBH of at least 12 cm all individual tree information – the raw data, like the DBH as well as derived attributes, such as the stem volume. All trees that were assessed on the matched grid NFI1-NFI2 in the first as well as in the second NFI (with the code value for the variable HISTORY ranging between 1 and 6, see Chapter 3.2.5) have an occurrence for invnr = 100 (NFI1) and an occurrence for invnr = 210 (NFI2).
TB	Contains all individual tree information for the tariff sample trees.
SORTD	Contains assortment volume of the assortments in which an individual tree can be divided into.
BEMERK	Contains all remarks about the individual tree attributes.
SCHADEN	Contains all damages observed on the trees.
SORTTB	Contains assortment volumes of the assortments in which a tariff sample tree can be divided into.

Table 2. Global key variables for the NFI database.

Global key variable	Definition
CLNR	Artificial number, which identifies unambiguously a point that is defined by the X and Y coordinates of the Swiss Federal Office of Topography.
INVNR	Inventory number (invnr = 100 for NFI1; invnr = 110 for check assessment NFI1; invnr = 210 for field survey NFI2; invnr = 220 for check assessment NFI2; invnr = 230 for regional inventories).
JWSANR	Identification number of a young growth plot on a sample plot. The number is unambiguous for one sample plot.
BANR	Tree number. This number identifies a tree unambiguously in the entire database. If a tree dies, this number is not used any longer.

Table 3. Local key variables for the NFI database.

Locale key variable	Definition
SABNR:	Identification number of a road section. This number identifies a road section in the entire database.
VPNR:	Identification number for a permanently marked point. This number identifies a permanently marked point in the entire database.
NHLH:	Tree species group: Broadleaf or conifer.
VCODE:	Type of expenditure calculation.
RUEPHASE:	Phase of the timber extraction.
MITTEL:	Timber extraction method.
ZIEL:	Place to which timber is skidded after cut.
WRART:	Tree species for forest edge survey.
PHASE:	Work stage at the terrestrial sample plot.
GRENART:	Type of border (forest edge, accessibility boundary).
GRU	Denotation of a young growth group for the closure assessment.
JWART	Tree species for the young growth survey.
KLA:	Young growth class.
SCHADEN	Young growth plant damage (e.g., browsing).
SARTNR	Type of the timber assortment (short or longer stemwood).
SORTKLD	Assortment classes.
BEM	Remark.
SCHANR	Damage identification number (number unambiguous for one tree).
SCHART:	Type of damage.
SCHURS:	Cause of damage.
SORTART:	Type of the timber assortment (short or long stemwood).
SORTKL:	Assortment class

4.1.3.2 Tables and Key Variables

Table 1 gives an overview of the tables in the NFI database and of their contents. In Table 2, the global key variables and their meanings are listed. Table 3 lists the local key variables. The individual attributes of the tables and their meanings are presented in the variable documentation (Chapter 6).

4.1.4 Literature

FLEMING, C.C.; VON HALLE, B. 1989. Handbook of Relational Database Design. New York: Addison-Wesley.

KORTH, H.F.; SILBERSCHATZ, A. 1986. Database System Concepts. New York: McGraw-Hill.

WERTZ, C.J. 1986. The Data Dictionary: Concepts and Uses. Amsterdam: North-Holland.

YANNAKOUDAKIS, E.J. 1988. The Architectural Logic of Database Systems. New York: Springer.