

Möglichkeiten der geographischen Datenerfassung in EDV-Informationssystemen

André Kilchenmann

1. Einleitung

Seit eh und je stellte sich die Frage, wie gesammeltes Datenmaterial am zweckmäßigsten aufbewahrt werden sollte, in welcher Form es am zweckmäßigsten zu speichern sei. Früher blieb nichts anderes übrig, als solches Zahlenmaterial in Tabellenform auf Papier zusammenzustellen und es in dieser Form, wenn möglich, zu publizieren. Heute steht uns für solche Aufgaben das ganze System der elektronischen Datenverarbeitung (EDV) zur Verfügung. Das Datenmaterial wird möglichst sinnvoll auf irgendwelche Datenträger übertragen (Lochkarten, Lochstreifen, Magnetbänder, Plattenspeicher), welche ohne weitere Arbeit von einem Computer gelesen und auf irgendeine Art (graphisch, tabellarisch, mathematisch) verarbeitet werden können. Diese moderne Einrichtung von sog. Datenbanken weist eine ganze Menge von bedeutenden Vorteilen auf, so daß in letzter Zeit auch in der Schweiz an verschiedenen Stellen mit dem Aufbau von solchen Informationssystemen begonnen wurde. Es sei hier kurz auf drei Beispiele verwiesen: auf den Informationsraster des ORL-Institutes an der ETH (1), auf den Datenraster der Region Zürich am Geographischen Institut der Universität Zürich (2) und auf die Datensammlung HELVETIA (Prof. Dieter Steiner, University of Waterloo und A. Kilchenmann, Universität Zürich). Eine gute Zusammenstellung über verschiedene andere Datenbanken, die zum Teil im Entwicklungsstadium sind, zum Teil aber bereits voll in Gebrauch stehen, findet sich in der Publikation von Tomlinson (3). Ob die allgemeine Bezeichnung «Environment Information Systems» oder «Information Systems for Regional Development» (4) lautet, sollte nicht über die Tatsache hinwegtäuschen, daß es sich durchwegs um geographische Informationssysteme handelt.

Man kann häufig feststellen, daß die Namengebung innerhalb von Informationssystemen nicht völlig durchdacht wurde. Dies führte zu Verwirrungen bei der Verwendung einiger Ausdrücke. In dieser kurzen Arbeit soll versucht werden, ein Schema über die Möglichkeiten der Aufnahme und Speicherung von geographischen Daten (= raumbezogenen Daten) aufzustellen, welches drei Dinge berücksichtigt:

1. die Art der statistischen Basiseinheiten
2. die Art der Raum- oder Lagebestimmung
3. die Art der Datenorganisation

Vorerst wollen wir einige Begriffe zusammenstellen, welche bereits ziemlich einheitlich im Gebrauch sind.

2. Grundbegriffe

Grundsätzlich kann eine Datenbank *qualitative* und *quantitative* Angaben enthalten. Der Unterschied liegt lediglich im Meßniveau, welches wir wählen oder wählen müssen (5).

Die detaillierteste Informationseinheit bezeichnet man als *Merkmal* (oder auch Einzelmerkmal). Mehrere sich gegenseitig ausschließende Merkmale nennt man eine *Merkmalsgruppe*, und mehrere Merkmalsgruppen, die miteinander in einem sachlichen Zusammenhang stehen, bezeichnet man als *Merkmalskategorie*.

Merkmale, die auf metrischen Skalen gemessen werden, werden meistens als *Variablen* bezeichnet.

Die Grundeinheiten, auf die sich die Merkmalsgruppen beziehen, wollen wir als *statistische Basiseinheiten* bezeichnen.

Unter dem Begriff *Zonen* verstehen wir Gebiete, welche rechtskräftig durch einen amtlichen Plan für bestimmte Nutzarten ausgeschieden worden sind.

Als *Areal* bezeichnen wir ein Gebiet, in dem ein bestimmtes Merkmal oder eine bestimmte Merkmalsgruppe dominiert (z. B. Laubwaldareal, Wieslandareal). Als *Region* bezeichnen wir eine Gesamtheit von zusammenhängenden Gebieten, welche sich durch eine gewisse Homogenität in der Merkmalszusammensetzung auszeichnen. Falls solche Gebiete mit ähnlicher Merkmalskombination teilweise verstreut vorkommen, sprechen wir von einem *Regio-entyp*.

Meistens bestehen die gegenwärtigen Datenbanken aus *starrer* Information (sog. «stock information»), d. h. aus Informationen, welche lediglich etwas über Zustände aussagen, die formale Aspekte der statistischen Basiseinheiten widerspiegeln.

Es ist zu hoffen, daß in zunehmendem Maße *fließende Information* (sog. «flow information») in die Datenbanken integriert wird. Unter diesem Begriff müssen wir Informationen über Wechselbeziehungen zwischen den statistischen Basiseinheiten ver-

stehen, also Angaben über funktionale Aspekte (z. B. Gütertausch, Verkehrsbewegungen, soziale Interaktionen).

Grundsätzlich wollen hier kurz darauf hinweisen, daß *Daten vier «Dimensionen»* aufweisen (6):

1. Besonders wichtig ist der *Raum*, d. h. der Standort oder die Lage der statistischen Basiseinheiten. Wir werden später auf die verschiedenen Möglichkeiten der Raum- oder Lagebestimmung im Detail eingehen.
2. Die zweite Dimension der Daten bilden die *Charakteristika* (Merkmale). Unterschiede ergeben sich hier wegen der verschiedenen Meß-Skalen, die verwendet werden können.
3. Die einfachste Dimension von Daten ist die *Zeit*. Sie ist linear faßbar und von den beiden ersten Dimensionen völlig unabhängig.
4. Die *Genauigkeit* ist eine weitere Dimension der Daten. Sie wird häufig vernachlässigt. Bei jeder Analyse von Daten sollte man genaue Vorstellungen über die Genauigkeit und Zuverlässigkeit der drei anderen Dimensionen besitzen.

Mit anderen Worten: Von jeder Zahl, welche sich in einer Zelle einer Datenbank befindet, sollte man wissen:

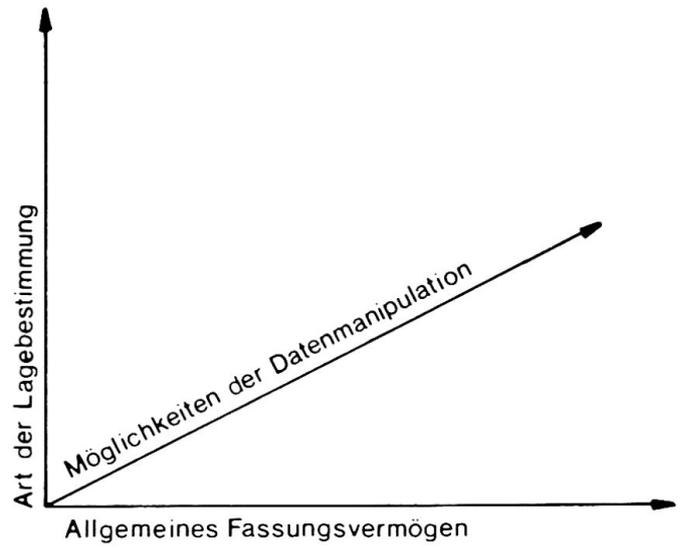
1. An welcher Stelle der Erde die Zahl gilt;
2. was die Zahl bedeutet;
3. wann die Zahl gültig war (oder ist);
4. wie genau diese Angaben sind.

Bei jeder detaillierten Besprechung von Informationssystemen müssen wir genau zwischen den folgenden *vier Hauptproblemen* unterscheiden:

1. Datenerfassung;
2. Datenspeicherung;
3. Datenabfrage – Datendarstellung – Datenmanipulation;
4. Datennachführung.

Die «*Commission on Geographical Data Sensing and Processing*» der IGU (International Geographical Union) ist im Moment daran, diese Detailprobleme zu studieren. Sie hat als Fernziel den Aufbau einer internationalen geographischen Datenbank im Auge.

Zum Abschluß dieses Kapitels soll darauf hingewiesen werden, daß von dieser IGU Commission ein Vorschlag zur *Klassifikation von Informationssystemen* vorliegt. Jedes Informationssystem kann in bezug auf seine Lage im folgenden dreidimensionalen Raum eingereiht werden (7):



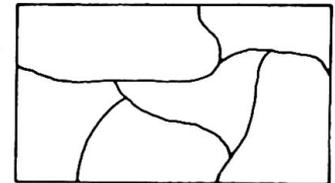
3. Arten der statistischen Basiseinheiten

Grundsätzlich kann man sich drei Arten denken:

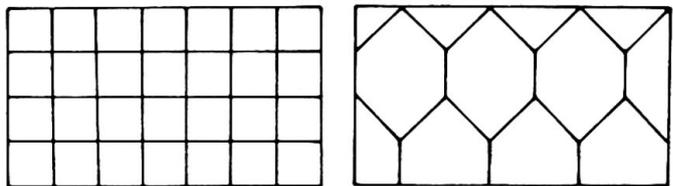
- a. *Punkte*,
- b. *Linien*,
- c. *Flächen*.

Die Flächen lassen sich in drei Flächentypen unterteilen:

- 3.1. *Administrative Flächen* (zusammenhängend, blattfüllend)

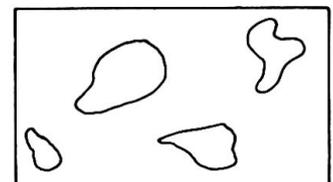


- 3.2. *Rasterflächen*, (Rechtecke, Quadrate, Hexagone)



Bei einem unendlich feinen Raster wird die Rasterfläche zum Punkt.

- 3.3. *Einzelflächen* (nicht zusammenhängend, nicht blattfüllend)



4. Arten der Raum- oder Lagebestimmung

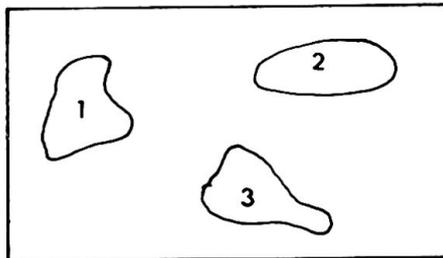
Bei der Art der Lagebestimmung müssen grundsätzlich zwei Möglichkeiten unterschieden werden:

4.1. Die Lage der statistischen Basiseinheit wird *extern* festgelegt. Eine codierte Bezeichnung (Zahlencode, Namenscode, Adreßcode) identifiziert jede Basiseinheit. «Extern» muß jedoch nicht bedeuten, daß die Lage damit außerhalb des Informationssystems festgelegt ist. Der «Code» kann (und sollte) im System selber, beispielsweise im Datensatz, als eines der Merkmale enthalten sein.

4.1.1. *Zahlencode*, beispielsweise:

1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	.	.	.			

1	2	3	4	5	6	7
.
8	9	10	11	12	13	14
.
15	.	.	.			
.	.	.	.			

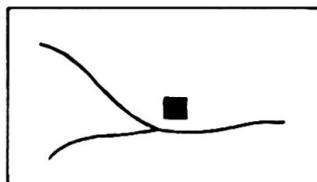


4.1.2. *Namenscode*
beispielsweise:



4.1.3. *Adreßcode*, beispielsweise:

Blümlisalpstr. 10, Zürich
Schweiz, Europa, Erde



4.2. Die Lage der statistischen Basiseinheit wird *intern* festgelegt. In diesem Fall wird ein Koordinatenpaar als direkte Identifikation der Basiseinheit verwendet.

Koordinatensystem

Es sind zwei Arten von Koordinatensystemen zu unterscheiden:

a) Feldkoordinaten

	1	2	3	4	5	6	7	8	9	.	.	.
1												
2												
3												
4												
5												
6												
7												

b) Punktkoordinaten

	1	2	3	4	5	6	7	8	9	.	.	.
1	•	•	•									
2												
3												
4												
5												
6												
7												

5. Art der Datenkoordination

Bei diesem Unterscheidungslevel legen wir das Augenmerk auf die Frage: Wie oder in welcher Form sind die Daten gespeichert, oder, als was sind die Daten gespeichert? Die drei einfachsten Antworten und Möglichkeiten lauten: als Linie, als Tabelle, als Raum (im Sinne von «spatial»).

Bevor wir diese Typen besprechen, wollen wir kurz auf die Terminologie eingehen, wie sie in der Datenverarbeitung verwendet wird. Die Eingabe von Daten auf die modernen Datenträger basiert auf der Einheit des «record», des *Datensatzes*, welcher aus sog. Datengruppen mit Datenfeldern zusammengesetzt ist. Ein Datensatz kann am einfachsten als Zeile oder Spalte einer Matrix (im mathematischen Sinn) verstanden werden. Ein oder mehrere Datensätze bilden eine sog. *Datei* (engl. «file»), und mehrere Dateien bilden eine sog. *integrierte Datei*. Von einer *Datenbank* spricht man in der Regel erst, wenn zur gesamten Datensammlung noch verschiedene Manipulationsprogramme gehören.

Bei unserer Typisierung auf dem Niveau der Datenorganisation verwenden wir den Namen «File». Für die Speicherung von spezifisch geographischen

Daten (= raumbezogenen Daten) lassen sich drei einfache Arten von Filetypen unterscheiden.

5.1. Linientyp (Linienfile, «linefile»)

Ein Linienfile enthält den Verlauf einer Linie. Dieser Typ stellt nicht eine Matrix dar, sondern lediglich einen Zahlenvektor. Ein Linienfile besteht also nur aus einem einzigen Datensatz. Als Beispiel kann man sich die Erfassung eines Straßenverlaufs (z. B. die N 3) vorstellen, der durch die Koordinatenwerte von einzelnen Punkten auf der Straßelinie räumlich festgelegt wird und folgendermaßen gespeichert wird:

N 3: $x_1, y_1/x_2, y_2/x_3, y_3/ \dots \dots \dots /x_n, y_n$

5.2. Tabellentyp (Tabellenfile, «statistical file»)

Ein Tabellenfile enthält die Daten einer Tabelle. Es werden also die Zeilen oder Spalten einer Tabellenmatrix gespeichert. Eine solche Tabellenmatrix, welche spezifisch geographische Daten («geographical facts») enthält, wird in der Literatur häufig als «geographische Matrix» bezeichnet.

Je nachdem, ob man die Zeilen oder die Spalten als einzelne «records» auffaßt, lassen sich die folgenden beiden Fälle unterscheiden:

5.2.1. *Regionaler Tabellenfile*: Jeder Datensatz enthält die Angaben über alle Merkmale einer statistischen Basiseinheit (Gemeinde, Land, Rasterfläche usw.).

5.2.2. *Thematischer Tabellenfile*: Jeder Datensatz enthält die Angaben für je ein Merkmal von allen statistischen Basiseinheiten.

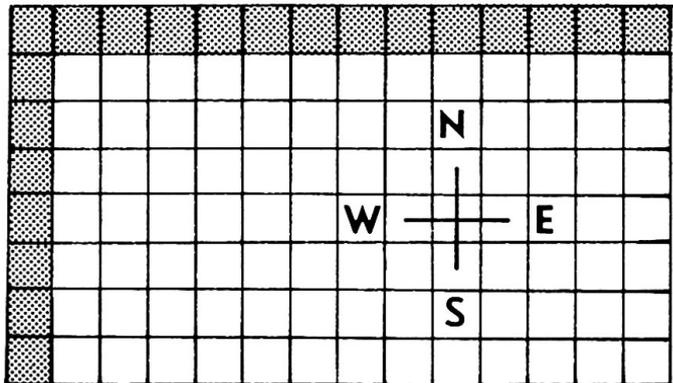
Eine solche Datei enthält also für alle statistischen Basiseinheiten Werte für verschiedene Merkmale, und eine integrierte Datei würde aus mehreren (eventuell verschieden großen) Tabellen bestehen.

5.3. Raumtyp (Raumfile, «grid file»)

Ein Raumfile enthält die Verteilung eines Merkmals im Raum. In einem Raumfile werden die Daten also so gespeichert, wie sie im Raum liegen, d. h. lagekonform. Bei der formalen Matrix handelt es sich um ein Rasternetz, das direkt auf der Karte liegt. Auch in diesem Fall sind wieder zwei Möglichkeiten zu unterscheiden, je nachdem, ob die Spalten oder die Zeilen als Datensatz aufgefaßt werden:

5.3.1. *N-S Raumfile*: Die Zeilen als Datensatz (bei allgemeiner Nordorientierung)

5.3.2. *W-E Raumfile*: Die Spalten als Datensatz (bei allgemeiner Nordorientierung)



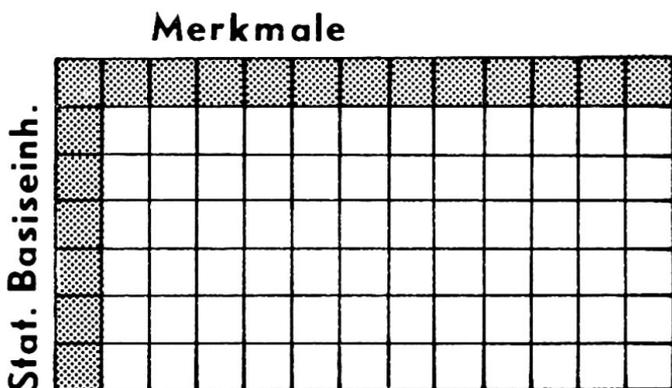
Horizontale Zeile:
Datensatz 1 eines N-S Raumfiles
Vertikale Kolonne:
Datensatz 1 eines W-E Raumfiles

Während im Tabellenfile beliebig viele Merkmalsgruppen eines Datensatzes enthalten sein können, erfaßt ein Raumfile nur eine einzige Merkmalsgruppe, weil beide Dimensionen der Matrix für die Lagebestimmung benützt werden. Im Tabellenfile wird die Lagebestimmung auf eine Dimension komprimiert (codiert).

Die Erfassung der Verteilung mehrerer Merkmalsgruppen erfordert also für Raumfiles mehrere Dateien. Der gesamte Aufwand an einzelnen Datenzellen ist indessen bei Raumfiles und Tabellenfiles bei gleicher Anzahl von statistischen Basiseinheiten und Merkmalsgruppen derselbe.

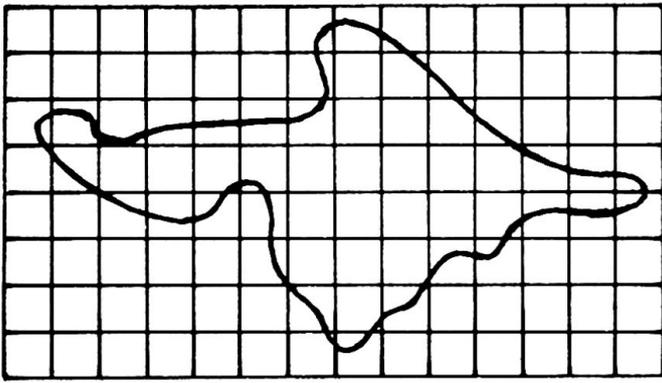
Raumfiles sind in erster Linie dann nicht zu empfehlen, wenn das untersuchte Gebiet nur schlecht durch ein Rechteck angepaßt werden kann.

Der Speicherplatz wird in diesem Fall schlecht ausgenützt (viele leere Zellen):

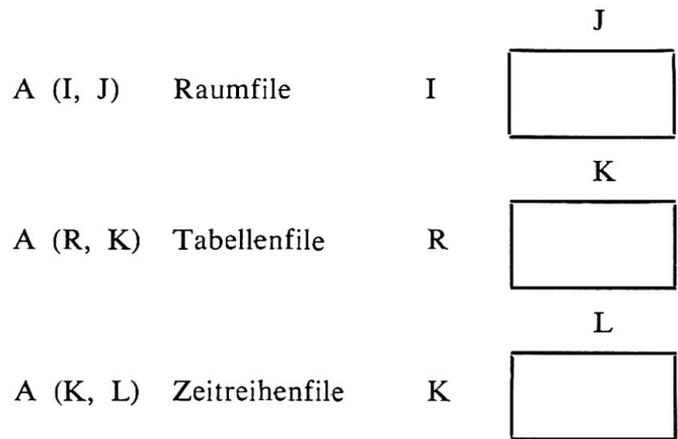


Horizontale Zeile:
Datensatz 1 eines regionalen Tabellenfiles
(eine Region)

Vertikale Kolonne:
Datensatz 1 eines thematischen Tabellenfiles
(ein Thema)



In der Praxis werden, wenn immer möglich, solche mehrdimensionale Arrays aus ökonomischen Gründen (z. B. Rechen- und Zeitaufwand) vermieden, und in zweidimensionale Arrays aufgeteilt, z. B.:



5.4. Mehrdimensionale Arrays

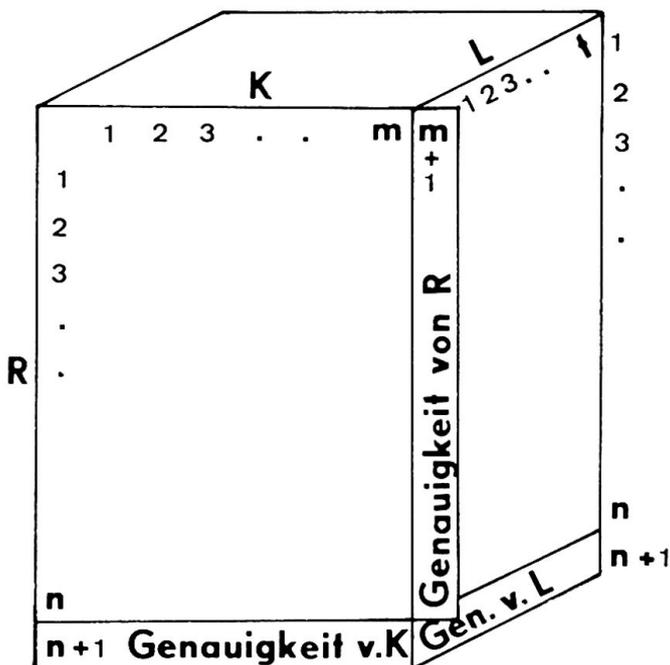
Computer- und programmieretechnisch ist es möglich, sämtliche Dimensionen der Daten, wie sie vorher kurz beschrieben worden sind (Raum, Charakteristika, Zeit, Genauigkeit), in einem sog. «Array» mit mehreren Indices zu speichern. «Array» wird auf deutsch mit Datenfeld übersetzt und folgendermaßen umschrieben (8). Ein Datenfeld ist ein zur Aufnahme von zusammenhängenden Daten definierter Bereich im Arbeitsspeicher.

Dies kann auf folgende zwei Arten geschehen:
 A (I, J, K, L) oder A (R, K, L)

wobei: I: Geographische Länge $I = 1, 2, \dots, n$
 J: Geographische Breite $J = 1, 2, \dots, n$
 K: Charakteristika $K = 1, 2, \dots, m$
 L: Zeitabschnitte $L = 1, 2, \dots, t$
 R: Lagecode $R = 1, 2, \dots, n$

Mit Einbezug der Genauigkeitsangaben kann man sich

A (R, K, L), mit
 $R = 1, n + 1 / K = 1, m + 1 / L = 1, t$
 geometrisch folgendermaßen vorstellen:

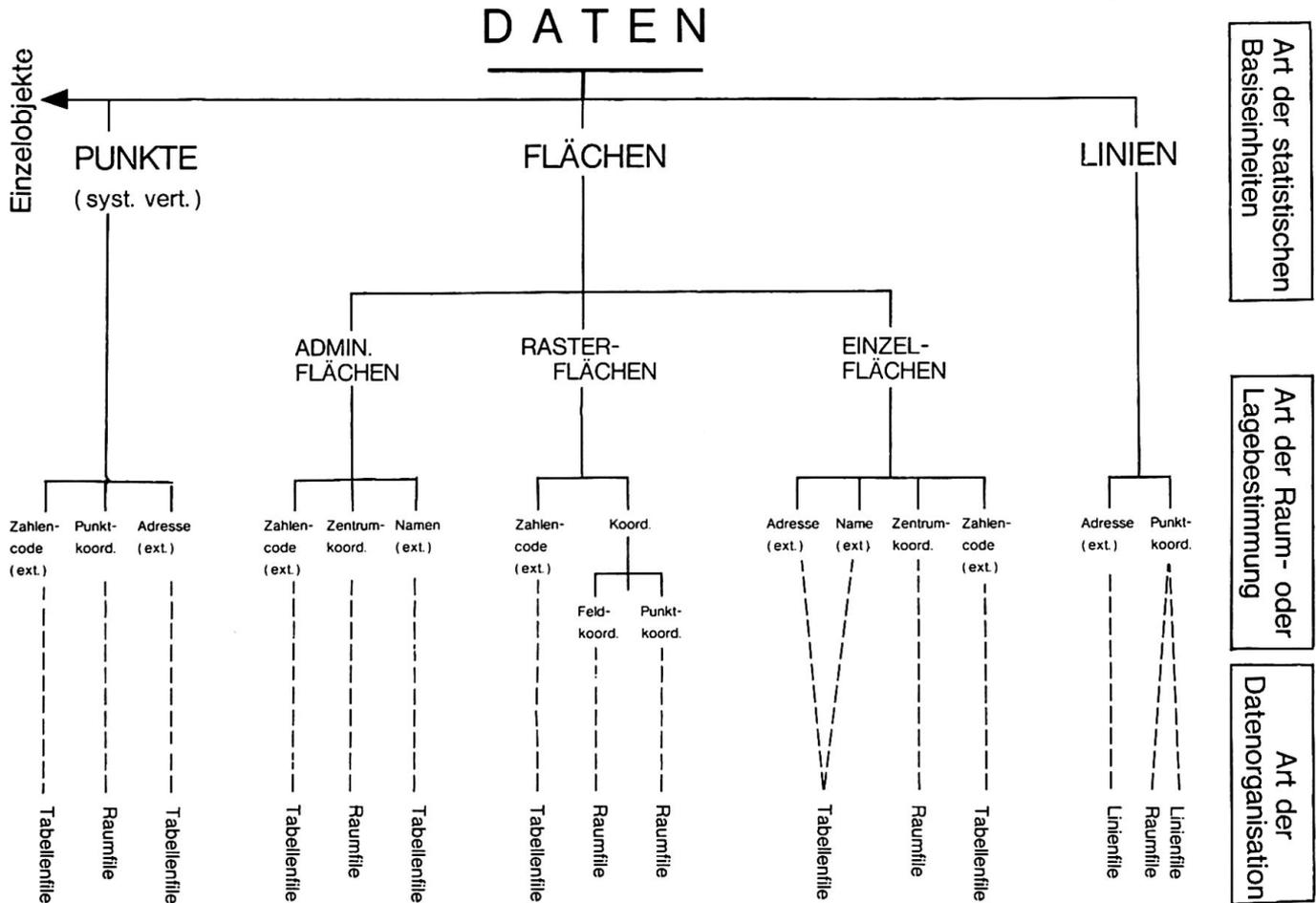


6. Zusammenfassendes Schema

Unter Berücksichtigung der Art der statistischen Basiseinheiten, der Art der Raum- oder Lagebestimmung und der Art der Datenspeicherung läßt sich das auf Seite 30 folgende Datenorganisationschema aufstellen:

Referenzen

- (1) Informationsraster, Arbeitsberichte Nr. 4.1. bis 4.6. zur Orts-, Regional- und Landesplanung; Institut für Orts-, Regional- und Landesplanung an der ETH, Zürich.
- (2) Brassel, Kurt: Das Datenraster Region Zürich; Diskussionsmanuskript No. 4, Geographisches Institut der Universität Zürich, 1972.
- (3) Tomlinson, R. F.: Environment Information Systems; Proceedings of the UNESCO/IGU First Symposium on Geographical Information Systems, IGU Publication, Ottawa, 1970.
- (4) Hägerstrand, T. und Kuklinski, A. R.: Information Systems for Regional Development – a Seminar; Lund Studies in Geography, Lund, 1971.
- (5) Kilchenmann, André: Geographie und moderne Wissenschaft; Diskussionsmanuskript Nr. 1; Geographisches Institut der Universität Zürich, 1971.
- (6) Salomonsson, Owe: Some problems of data identification, Data Integration and Data Organiza-



tion in Urban and Regional Information Systems, in Hägerstrand and Kuklinski, op. cit.

Kurzbemerkung zum Schema

(7) Tomlinson, R. F.: op. cit. p. 1-21.

Unsystematisch verteilte Punkte können immer auf den Fall von sehr kleinen Einzelflächen zurückgeführt werden, mit den gleichen Möglichkeiten, die unter dieser Kategorie aufgeführt sind!

(8) Schneider, C.: Taschenlexikon der Datenverarbeitung, Forkel-Verlag, Stuttgart.