ISBN 92-894-0432-9

2000 EDITION

Manual of concepts on land cover and land use information systems

EUROPEAN COMMISSION

eurostat

THEME 5
Agriculture and fisheries

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (http://europa.eu.int).

Cataloguing data can be found at the end of this publication.

Luxembourg: Office for Official Publications of the European Communities, 2001

ISBN 92-894-0432-9

Printed in Luxembourg

PRINTED ON WHITE CHLORINE-FREE PAPER

TABLE OF CONTENTS

1. INTRODUCTION

1. INTRODUCTION

The objective of the manual of concepts is to raise the awareness and sensibility of “experts” and “non experts” concerning differences in technical, methodological and conceptual aspects of land cover and land use information systems.

The manual should contribute to a better understanding, a more objective validation of current information and improve the necessary sensibility for a harmonised land cover and land use information system at European level.

In order to reach this objective the Manual of Concepts focuses on:

- definition of basic terms (land, land cover, land use),
- explanations of fundamental concepts and principles of essential elements of land cover and land use information systems (e.g. classification systems, data collection tools),
- explanations, definitions and descriptions of technical terms, which are often ignored but fundamental regarding information return,
- identification of data requirements and related problems.

The manual provides “experts” and “non experts” with a common “understanding” and “language” concerning different elements of land use and land cover information systems.

2. DEFINITION OF GENERAL TERMS

2. DEFINITION OF GENERAL TERMS

2.1. Land

Land is a term widely used throughout the world but definitions are not frequently given.

The interdepartmental working group on land use planning (IDWG-LUP) at FAO proposed in 1994 the following definition: *“A delineable area of the earth's terrestrial surface, embracing all attributes of the biosphere immediately above or below this surface, including those of the near surface climate, the soil and terrain forms, the surface hydrology including shallow lakes, rivers, marshes and swamps, the near-surface sedimentary layers and associated groundwater and geohydrological reserves, the plant and animal populations, the human settlement pattern and physical results of past and present human activity (terracing, water storage or drainage structures, roads, buildings, etc.)”.*

The question of the area to be taken into consideration is also introduced through the former definition: “all attributes of the biosphere immediately above or below the surface”. For land cover, the question is easily solved: the reference area for land is above the surface (see definitions next chapter). For land use, the situation is more complex: from a pragmatic point of view and considering the importance and the significance in terms of economic value of multiple-use aspects for "urban" areas, the understanding of land should also embrace uses above and below ground level. Particular problems may be found with mine deposits, subways beneath urban areas, water resources, mushroom beds, etc. For example: areas used for oil extraction; is it the whole field of extraction (and to a certain extent the whole oil concession), or is it limited to the oil well itself?

In order to limit potential problems it is proposed to restrict its application to reasonable cases. It is suggested to allow "urban" uses above and below ground level (case of buildings with shops at ground level, flats and offices above, car parks below). The most important problem remains in fact with extraction activities of natural resources (including water resources). Considering this difficulty (and also the difficult task to collect appropriate data) it is suggested to restrict the extension of such uses to their physical impact at ground level (oil well, entrance of mushroom beds etc.). This principle may be generalised for any kind of uses.

Given the context of global information systems of land cover and land use and the difficulty to establish clear thresholds between land and water (particularly for wetlands), it is recommended **to extend the concept of land to inland water areas and tidal flats.**

This approach and proposal of definition is to be clearly separated from the concepts utilised by statisticians for the determination of land area used for statistical purposes. Eurostat has recently proposed (EUROSTAT 1999) the concept of Land area to be used for statistical purposes, therefore excluding lakes, rivers and coastal areas. This is easily understood in the context for example of calculation of population densities where 17% of a country such as Netherlands is covered by water areas. The Eurostat approach is driven by the necessity to provide harmonised statistical data, the best example being the calculation of densities of population (figures for the Netherlands are changing dramatically if water areas are included within the total area of the country).

2.2. Land objects

If the question of units or objects is self-evident for many scientific fields, for land it is somewhat not the case: the meaning of an object is a complex problem since the categorical classification of a part of the earth’s surface pre-supposes that the area is clearly defined in space (DUHAMEL & VIDAL 1998). Objects are easily identifiable if the spaces are plots of farmland or built-up areas, as they have physical boundaries. However, these boundaries become blurred in semi-natural or natural environments. Delimitation problems are compounded in transitional zones. For example in the Mediterranean environment there are indistinct transitions in the biophysical *continuum* between forest, scrub and dry grassland. Delimitation problems also arise when use is made of category definitions based on cover or use percentages.

Mixed objects

Three types of mixtures exist on land:

- Spatial mixtures “by juxtaposition” that are highly dependent of scale and observation units. Statisticians apply the *prorata* rule (case of associated crops and natural biotopes). Many “nomenclatures” attempt to solve the problem by creating mixed classes, resulting in legends since they do not comply with the principle of absence of overlap);

Figure 2.1: Juxtaposition of cover and use

- Spatial mixtures in the third dimension are created by “superposition” of different covers or uses. Whatever the resolution of observation, decisions for discriminating the different covers or uses in the third dimension need a rule. A better resolution for observing superpositions does not have any impact: mixtures in the third dimension are therefore scale independent. Statisticians again apply the *prorata* rule (case of crops associated with tree covered areas). Sometimes specific rules such as dominance are adopted: example of dominant uses for urban.

Figure 2.2: Illustration of superposition

- Temporal mixtures of covers or uses are highly dependent on the period *dt* of observation. Agriculture statisticians apply the main crop rule, main being defined through the main commercial value. If this rule is not applicable it is replaced by the crop remaining the longest time on the parcel).

2.3. Land Cover and Land Use

Many existing information systems are mixing land cover and land use where natural and semi-natural vegetation are described in terms of land cover and agricultural and urban areas in terms of land use. However, these are two different issues and distinction between land cover and land use is fundamental though often ignored or forgotten. Confusion and ambiguity between these two terms lead to practical problems, particularly when data from the two different dimensions need to be matched, compared and/or combined.

Land cover corresponds to a physical description of space, *the observed (bio)physical cover of the earth's surface* (DI GREGORIO & JANSEN 1997). It is that which overlays or currently covers the ground. This description enables various biophysical categories to be distinguished - basically, areas of vegetation (trees, bushes, fields, lawns), bare soil (even if this is a lack of cover), hard surfaces (rocks, buildings) and wet areas and bodies of water (sheets of water and watercourses, wetlands). This definition has impacts on development of classification systems, data collection and information systems in general. It is said that Land Cover is "observed". This means that observation can be made from various "sources of observation" at different distances between the source and the earth's surface: the human eye, aerial photographs, satellite sensors.

For **land use,** various approaches are proposed into the literature. Two main “schools” may be distinguished. Land use in terms of *functional dimension* corresponds to the description of areas in terms of their socio-economic purpose: areas used for residential, industrial or commercial purposes, for farming or forestry, for recreational or conservation purposes, etc. Links with land cover are possible; it may be possible to infer land use from land cover and conversely. But situations are often complicated and the link is not so evident. Another approach, termed *sequential*, has been particularly developed for agricultural purposes. The definition is a series of operations on land, carried out by humans, with the intention to obtain products and/or benefits through using land resources. For example a sequence of operations such as ploughing, seeding, weeding, fertilising and harvesting (MÜCHER et al. 1993).

Contrary to land cover, land use is difficult to "observe". For example, it is often difficult to decide if grasslands are used or not for agricultural purposes. The information coming from the source of observation may not be sufficient and may require additional information. In the case of agricultural use, farmers may bring information, for example if cattle are present or not, if they are grazing. It is also possible to use characteristics on the spot indicating the presence or absence of cattle. For the FUNCTIONAL approach, inference from land cover may be helpful. For the SEQUENTIAL approach, a more exhaustive recording of various attributes will be needed, for example a multi-temporal approach. In the following, land use will be understood as FUNCTIONAL.

2.3.1. Interrelationship between land cover and land use

A very useful comparison can be made with approaches for classifying commodities where objects are described according to the material they are made of and the function they serve.

It is sometimes possible to determine **functional aspect** from **biophysical aspect** (DUHAMEL & VIDAL 1998). A parcel of land covered by a field of wheat can reasonably be associated with agricultural use. Similarly, it is possible to infer biophysical aspect from functional aspect. An area used for forest production can reasonably be assumed to correspond to a biophysical class of the “tree” type. However, for others, one biophysical category may correspond to a large number of functional categories. Areas of grass may, for example, correspond to a lawn in an urban environment, an airport runway, a sown meadow, rough pasture, a golf course - or even a church roof in Iceland. Conversely, one and the same functional class may cover several biophysical categories: for example, a residential area consists of lawns, buildings, tarmac roads, trees and bare soil.

There are methodological and technical arguments in favour of the systematic separation of the two approaches. Even if it is difficult to justify when analysing both user needs and the possible costs of simultaneously acquiring, using and managing data obtained through separate approaches, importance of the knowledge for the two dimensions may be illustrated with the following example adapted from (LUND 1998):

Let us imagine the following information system addressing uniquely the land cover dimension at different observation dates:

	t1	**t2**	**t3**	**t4**
LAND COVER	TREES (Chestnut trees)	TREES (Chestnut trees)	TREES (Chestnut trees)	TREES (Chestnut trees)

The sequence TREES>TREES>TREES>TREES yields a simple interpretation of non change throughout the period t1-t4 considered.

Let us consider the same area observed through an information system uniquely addressing land use:

	t1	**t2**	**t3**	**t4**
LAND USE	Forestry use (Timber)	Agricultural use (Chestnut production)	Agricultural use (Grazing area)	Forestry use

The general sequence FORESTRY>AGRICULTURE>AGRICULTURE>FORESTRY could be rapidly interpreted such as a deforestation sequence between t1 and t2 and afforestation between t3 and t4. With a better detail the sequence “Timber>Chestnut production>Grazing Area>Forestry use “provides a more clear picture of the possible land cover: between t1 and t2 it is clear that timber and chestnut production are corresponding to a tree cover. This is not so clear between t2 and t3 (chestnut production>grazing area)

	t1	**t2**	**t3**	**t4**
LAND COVER	TREES (Chestnut trees)	TREES (Chestnut trees)	TREES (Chestnut trees)	TREES (Chestnut trees)
LAND USE	Forestry use (Timber)	Agricultural use (Chestnut production)	Agricultural use (Grazing area)	Forestry use

The conclusion is that the simultaneous recording of both land cover and land use hampers any false interpretation or inference between land cover and land use and yields a richness in terms of contents of information useful for many applications.

2.4. Land cover – land use change

Land use and land cover classes represent analytical units, which allow to establish a first quantitative link between human activities, environmental impacts and its geographical (spatial) dimension. Information on land cover and/or land use change are of special value integrating the temporal dimension.

This is of overall interest for both politicians – for the evaluation of land related policy measures- and for the research community – discovering the underlying causes and consequences.

Land cover and land use change is commonly divided into two broad categories: conversion and modification (STOTT, A. & HAINES-YOUNG, R. 1996; ALUN, J. & CLARK, J. 1997; BAULIES, X.I. & SZEJWACH, G. 1997):

- **Conversion** refers to a change from one cover or use category to another (e.g. from forest to grassland)
- **Modification** represents a change within one land use or land cover category (e.g. from rainfed cultivated area to irrigated cultivated area) due to changes in its physical or functional attributes.

The pure land cover and land use information gains a significant added value through the analysis, identification and description of ongoing processes.

Based on land cover and land use change information, certain processes can be retrieved, which might also serve as simple indicators. Some processes are listed below:

Intensification

A flow representing the transition of land cover and land use types associated with low intensity use to a higher intensity (e.g. semi natural - arable land)

Extensification

A flow representing the transition of a land cover or land use type associated with high intensity of use to a lower intensity of use (e.g. improved grassland to semi- natural cover).

Afforestation

A flow representing the planting or natural regeneration of trees

Deforestation

A flow representing the clearance of trees.

Development

A flow involving the transformation of open land to urban, industrial or transport uses.

Reclamation

A flow involving the creation of open land to areas previously developed (e.g. reclamation of mineral workings)

These processes can further be categorised concerning direction (conversion or modification), magnitude (amount of change) and pace (rate of change).

It should be mentioned that the added value of land cover and land use changes and the retrieval of undergoing processes and indicators highly depend on the thematic and spatial detail of land cover and land use information. It has to be stated that land cover and land use information in general provide quantitative information (e.g. land cover/use categories, area estimates). Additional information concerning qualitative characteristics (e.g. farming practise, ecological value) is necessary to sufficiently describe the implicated processes.

This concerns also the analysis of the driving forces underlying land cover and land use changes such as economic, political, social, environmental and others forces. They are manifold and interrelated, which makes the analysis a highly complex one.

3. CLASSIFICATION SYSTEMS

3. CLASSIFICATION SYSTEMS

3.1. Definitions

Many terms with different senses are used in the field of classification, nomenclature and taxonomy. Terms are often used interchangeably and lead to many ambiguities for readers and users of information in a given field. Meanings of these terms often come from natural sciences such as biology, botany and zoology.

3.1.1. Classifications

The term classification is universally employed by scientists with a meaning historically linked to the classification of organisms, following an old Aristotelian tradition. However, the term classification is ambiguous since it covers two main senses (SUTCLIFFE, 1993):

- The sense of establishing a classification of all the objects in the universe of discourse U (responding to the instruction: classify the objects) resulting in a **classification of all objects**, the end-product being the organisation of sub-classes of U.
- The sense of using the classification so-established to decide the membership status of individual objects (responding to the instruction: classify the object X). This **identification** process is what the remote sensing community calls classification.

The widely used standard definition of classification of (SOKAL 1974) adapted from (SIMPSON 1961) is: "*the ordering or arrangement of organisms into groups or sets on the basis of their relationships*". This definition has been extended towards more general issues where the term *organisms* is replaced by the term *objects.* It corresponds to the process of establishing a classification of all objects but neither cover expressively the identification process nor the naming of groups and the rules governing application of these names. Moreover, the term relationship, as used in biology, covers two kinds of relationships: affinities in terms of similarity as judged by the characters of organisms (called phenetic relationship) and relationships by ancestry (called phyletic relationship), which do not have any equivalence in our domain of investigation.

3.1.2. Nomenclatures

Nomenclatures are lists of categories, summarising information in a highly reduced form while attempting to maintain a maximum information content. A nomenclature will normally cover a particular field of interest (the Universe of Discourse of SUTCLIFFE). Definitions of nomenclatures in biology deal with the legalistic aspect of taxonomy: the naming of groups and organisms and rules governing the application of these names.

3.1.3. Legends

Legends are frequently confused with classification systems and nomenclatures. In principle, a legend is an application of a nomenclature for a specific purpose: thematic mapping, i.e. using a defined scale and appropriate mapping units. If nomenclatures are exhaustive, covering the whole "universe of discourse" or domain, the legend is a sub-set, a portion of the nomenclature. Some classes of the nomenclature may be taken directly, others are omitted and others are combined, mixed into composite categories (mixed classes) according to minimum legible delineation constraints. If classification systems are and should be independent from scale and cartographic representation, this is not the case of legends. To a certain extent classification systems are partitioning land cover or land use types, contrarily to legends, which are partitioning the territory according to certain attributes or variables of land cover and/or land use. For example, CORINE Land cover is proposing a "nomenclature" which is in fact a legend where in particular classes are overlapping through mixed classes.

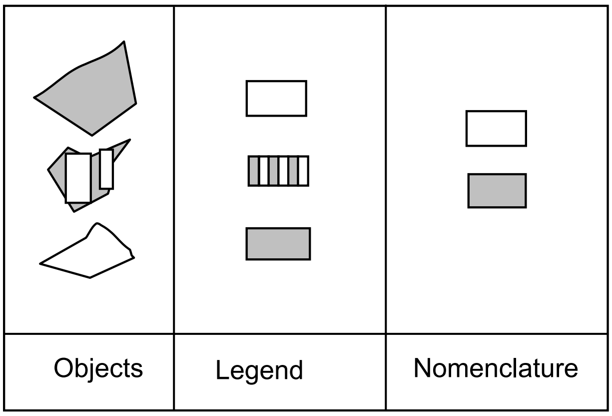


Figure 3.1: Objects, legends and nomenclatures

3.1.4. Towards classification systems

In front of such overlapping definitions and considering the need to develop fundamental means for structuring information and facilitating communication and exchange among users in a given discipline (land cover and land use in our case), it is necessary to know which functions a “classification system” would need in order to meet the requirements of scientists and users. Three main functions are proposed (mainly adapted from SUTCLIFFE 1993):

Classification = Assignment of all objects in a hierarchical series of nested categories that have been arranged to show relationships with one another,

Nomenclature = Naming and describing the groups of objects. The end-product is a list of names and descriptions linked by one-to-one mapping correspondence and generally presented according to the structure of the classification so-established.

Identification = Allocating an individual to a previously classified and named group. In other terms, using the basis of the classification so-established to decide the membership status of individual objects

Therefore, it is recommended to utilise the concept of **classification system** embracing:

- the demarcation of a universe of discourse (U): namely land cover and land use domain (DOMAIN);
- the establishment of a classification of all the objects in the universe of discourse (U) resulting in the organisation of sub-classes of U through a hierarchical series of nested categories that have been arranged to show relationships to one another (CLASSIFICATION);
- a system for naming and describing groups linked to the structure of the classification so-established (NOMENCLATURE);
- and procedures for allocating any object to one and only one previously classified and named group (IDENTIFICATION).

3.2. Classifications systems

3.2.1. An approximation of reality

Classification systems are tools, describing selected aspects of the real world. Categories chosen do not represent a one-dimensional partition of the real world but a multidimensional one. The partition of the real world through a classification system highlights certain aspects of reality: the same reality might well be described according to several classifications. (DUPRAT 1972)

Generally speaking classification systems are presented in tree form, i.e. hierarchically. A hierarchical system is an arrangement of objects into a series of groups, which are assigned to a succession of categories of ranks of different seniority. Groups of objects are defined by the selection of shared characteristics that make the members of each group similar to one another and unlike members of other groups. Each of the successive partitions means that "objective" characteristics have been taken into consideration, and implies a conscious choice. Different users may wish to partition the field of interest at a given level according to different criteria. This creates serious difficulties, when comparing information, since a class which one system regards as unique may fall into two or more categories in a classification system that is set up on different principles. For example, it is impossible to reconcile exactly a classification of trees that subdivides *broad-leaved* and *needle-leaved* species with one that identifies only *deciduous* and *evergreen* categories (WYATT 1997).

It is also to be noticed that a category of a classification can be homogeneous according to one character (a monothetic class), or two, or ...none (following the concept of polythetic classes proposed by biologists). In fact, if the process of aggregation is taken beyond a certain level of significance, categories no longer correctly represent meaningful entities: this is the case of an aggregate which would mix for example agricultural and urban areas within a classification system.

Therefore a classification system is the result of a structure and an order, coming from a system of values, revealing an intention. The purpose for which the classification is designed necessarily shapes its structure and content. This is why each user, in general, builds an individual classification adapted to his specific needs: spontaneous development of classifications therefore leads inevitably to incompatibility.

3.2.2. A compromise

A classification system should be the result of an ongoing dialogue between:

- **A systematic approach** imposing structure on information according to logical principles (completeness, absence of overlap, unambiguous definitions of classes, rules governing the representation of objects within the classification).
- **A pragmatic approach** taking account user's needs and existing sets of information.
- **A contextual approach** addressing specific constraints linked to the domain of investigation. For land cover and land use some constraints are inherent to the geographic dimension of information.

3.2.3. Some properties

Spatial consistency

The principle of ***spatial consistency*** requires that classification systems are designed in such a way that results are compatible between different sites, regions or countries in the geographical area under consideration.

Temporal consistency

The types of land use or land cover should be recorded at the instant time of observation (by the observer, the enumerator or satellite sensor). The classification system must therefore not take account of previous or future states (e.g. planned building sites: typically an intended use). Results must be considered as stocks, not flows (flows being measured by comparing two sets of stocks). This is the principle of ***temporal consistency***. However, particular attention has to be paid to areas where changes occur over a short period. The United Nations- Economic Commission for Europe (UN-ECE) nomenclature shows a bad example through the existence of a category 3.9.4 (Land intended for future construction) with a definition: land areas designated in public land development plans as land for construction, but where construction works have not yet started. The meaning of this information is questionable if the construction has not started or if the data is not updated after several years.

Compatibility with existing classification systems

A land cover or land use classification system is invaluable for compiling international or global statistical information on land cover / land use. However, creating a classification system for general usage can be justified only if it provides a fairly high degree of compatibility with existing information systems. Attention should be particularly paid to the fact that a proposed classification system should be as ***compatible*** as possible ***with major existing systems*** so as to allow meaningful conclusions to be drawn by reference to data from different relevant sources. For example a land use classification system should be linked as much as possible with socio-economic classifications since land use is dealing with socio-economic purpose. “Bridging tools” should be provided to enable such comparisons.

Independence from data collection and processing tools

In theory, the classification system should be constructed independently of the resources available for collecting information. In particular, it should be, as far as possible, free of cartographic restitution scales. However, practical experience has shown that it is difficult to construct a classification system that is not influenced in some respect by the observation method used. Many classifications have been adapted to observation tools. For land cover, classification systems have been built for taking into account specific tools such as aerial photographs or earth observation data. When tools are changed, the classification system is not capable of adaptation, and problems of continuity of time series of information arise. Frequently, compromises are to be found between availability of tools and user needs (both evolving).

Completeness

Every classification system refers to a certain “segment” of reality. This segment (the universe of discourse) must be described exhaustively: all physical “entities” in a classification of goods, all land cover types in a land cover classification, all landscape types in a landscape classification. The formal definition of completeness is that (BUNGE 1983):

- A class must be found for any object to be classified.
- The union of all classes on the basic level must equal the original collection.

Absence of overlap

Classes must be mutually exhaustive, without overlap. Absence of overlap is essential for a consistent application of a classification system: if overlap exists, then a choice must be made between two or more possible classes. Formally, the absence of overlap means that (BUNGE 1983):

- anything can be classified in only one class.

For land cover and land use this principle may be termed the principle of ***semantic consistency*** (EUROSTAT 1992). One fundamental consequence is the fact that mixed classes should be systematically excluded.

Identification rules

Established rules must be applied to identify and classify the objects in the appropriate classes according to criteria which should be clearly defined. In particular, cases of overlap, decisions on mixtures, decisions on parts and accessories in the case of goods classifications, should be resolved.

Naming rules

Text of headings should be elaborated from the **content** taking into account the characters of the notion and relations with other notions. The text of the heading should be:

- general, it cannot describe everything and should be an approximation and reflect the characteristics of the concept;
- concise, so long as it can be clearly understood, and abbreviated, otherwise, the user will abbreviate it;
- not circular (definition of arable land cannot be arable land).

It should also correspond to user knowledge and needs.

Definitions are verbal descriptions of a concept in terms of known concepts. Definitions are preferably given by **intension** (comprehension), providing the restrictive characters which are making distinction between the class and the others situated at the same level of the hierarchy. This is opposed to definitions by **extension** (generic) which consists in trying to enumerate all objects. Of course a combination of definitions both by intension and extension is helpful. **Prototypes** or types of categories (designated specimen in the biological terminology) could also be helpful: for example pictures of land covers unanimously recognised as a prototype or good representative of the class. Good examples are to be found in the Corine Land Cover technical guide (1990).

Explanatory notes: Explanatory notes provide the means for giving more detailed descriptions of the "objects" within a category. This is sometimes the easiest way for classifying since it indicates inclusions, exclusions, definitions and decisions to be taken. When an object is excluded from a category, it should be indicated where it belongs. After having identified the "kernels", the problems of boundaries between categories must be addressed. This is generally done (or not...) by listing, for each category, the **boundary conditions** that are included or excluded. In the case of exclusions, there should be cross-references to the appropriate category.

Index

An alphabetical index, to be built after the classification system, is a very helpful tool for identifying where to include an object within the existing classification (IDENTIFICATION). For example it is important to know in which class the object identified as a dam is pertaining: it is the role of the index, which has the role of enumerating all possible objects being part of the classification.

Principles of coding

Codes are attached to images. In general, codes are structured if the nomenclature is hierarchical. The main principle used is that the code of the lower level has to repeat the code of the higher level. In principle, it is possible to use any symbols for coding (numbers, characters, other signs). One important thing to be addressed is the meaning of the 0 (zero). As a general rule the zero should be used when, at a given level, no subdivision is made.

Rules for inclusion of new objects

Irrespective of the efforts made to ensure that the classification system is complete, it is inevitable that new objects have to be included at a later stage. In order to achieve this, the classification rules attached to the nomenclature must be adapted and the revised rules applied to all the objects affected by the change, including any objects previously classified. Attention should also be paid to the management of the “jurisprudence”: before including a new object it is necessary to be sure that it is a new object (importance of existing index!) and consequently the textual part of the nomenclature has to be carefully modified.

Aggregates

Having a textual part, particularly the denomination of the class, corresponding to its contents is a difficult exercise since information is disappearing through aggregation processes. It is necessary, for various purposes and for having meaningful questions and answers, to reach a certain level of detail: at most aggregated levels, categories lose their descriptive meaning and are just documentary milestones (DUPRAT 1972). Aggregates, even accepted into the common language, are in general units without single characteristics and with vague definitions.

Heterogeneity of semantic fields

Apart from purely translation problems, semantic fields of terms do not always match in a one-to-one mapping correspondence. An example is given by (ECO 1988), adapted from (HJELMSLEV 1957). The semantic field of forest-wood-tree shows that in different languages (French, German, Danish, Italian, English) there is no direct one-to-one mapping correspondence between the lexical units. It shows that in different languages the semantic value of each word is different: the partition of the *continuum* “forest-wood-tree” results into two words in Danish, three in French and German and four in Italian and English (without any clear distinction between timber and wood). This does not ease the harmonisation of concepts.

<table>
<tr><th>French</th><th>German</th><th>Danish</th><th>Italian</th><th>English</th></tr>
<tr><td>Arbre</td><td>Baum</td><td rowspan="2">Trae</td><td>Albero</td><td>Tree</td></tr>
<tr><td rowspan="3">Bois</td><td rowspan="2">Holz</td><td rowspan="2">Legno</td><td rowspan="2">Timber</td></tr>
<tr><td rowspan="3">Skov</td></tr>
<tr><td rowspan="2">Wald</td><td>Bosco</td><td>Wood</td></tr>
<tr><td>Forêt</td><td>Foresta</td><td>Forest</td></tr>
</table>

Figure 3.2: Heterogeneity of semantic fields (ECO 1988)

3.3. Existing approaches of classification systems

3.3.1. Hierarchical systems

The hierarchical aspect of nomenclatures is an artificial construction since real objects are linked through a complex net of relationships, likeness, affinities or neighbourhoods. In order to transform this net into a tree, some links or relations considered as important or significant have to be chosen according to a particular point of view.

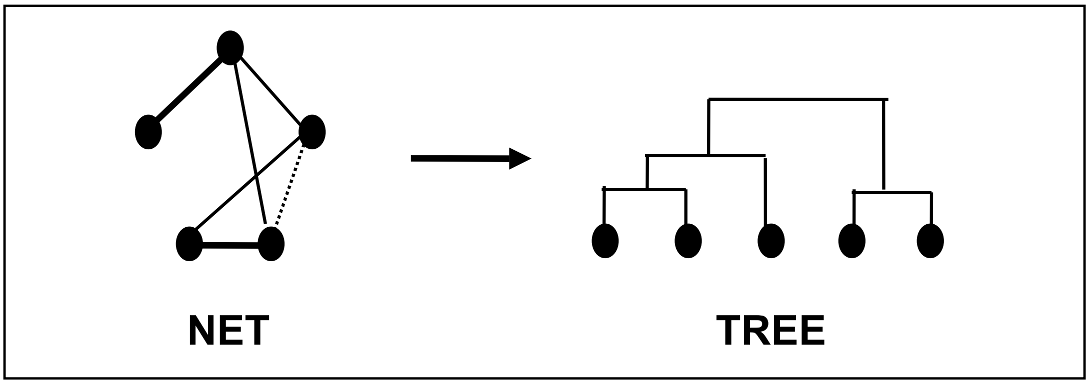


Figure 3.3: Illustration of net and tree systems

These links or relations are of different types, adapted from (WÜSTER 1971):

- logical links (industrial: chemical, iron transformation, car industry …);
- ontological links = whole and part (residential area - house);
- co-ordination links (maize-wheat);
- material-product links (vineyard - wine);
- time links (bare soil - crop).

These links and relations may be combined and presented from broader to narrower concepts. To a certain extent links between land cover are mainly ontological (whole-and-part: forest > stand > tree). Conversely, land use links are mainly logical (industrial: chemical, iron transformation etc...).

Clustering

The objective of this method is to constitute categories through the grouping of "similar" objects. This involves the comparison of a large number of characteristics of one object with the same characteristics of other objects. Objects with many features in common will be clustered together. A nomenclature is therefore derived from the grouping obtained. Various methods exist: factorial analysis, hierarchical ascending classifications, dynamic clustering. These are largely applied in the field of botany, zoology, language taxonomy through the term numerical taxonomy, see (SNEATH & SOKAL 1973). These methods may be useful if enough information on observation units is available. This is not often the case for land cover and land use.

Top-down tree (Descending tree)

This is the most self-evident solution. Many classification systems are built following this *a-priori* approach: the domain of study is divided into categories and sub-categories, according to certain objectives and purposes. The method has strong disadvantages: the tree is a rigid structure leading to difficulties if modifications are to be made without alteration to the former structure of information. The only possibility for modifying is creating more detailed levels on the basis of the categories already existing. In this case when the existing tree is no more adapted to the needs, additional levels do not solve the problem. Other disadvantages are the unequalled development of the sectors and the exaggerated importance of aggregated concepts created from the first partition. The United Nations-Economic Commission for Europe (UN-ECE) for example has developed a system aimed at describing land cover and land use. The nomenclature is hierarchical and comprises 3 levels. The two first levels are considered as basic and the third one is optional. The first level of the nomenclature is focused on land cover and the other levels introduce functional aspects. The nomenclature is quite well adapted to temperate and Nordic landscapes but is extremely difficult to utilise for describing southern European landscapes. As an example, the category 5 (Dry open land with special vegetation cover) is subdivided into 4 categories (Heathland, Dry tundra, Montane grassland and Others). This partition does not allow a good classification of Mediterranean landscapes which are included under the sub-category Others. Inclusion of these types of “landscapes” (savannah, chapparal, prairies) can only be done by subdividing the category ‘Others’ at a fourth level (which gives unbalanced categories at 4th level) or by complete restructuring of the whole category.

Discriminant analysis

Another possible approach is the discriminant analysis where from an a-priori partition in different classes attempts are made for optimising the separation of the classes and minimising the variance within classes through the use of a series of descriptive variables. This approach is helpful for allocating individuals into classes but it requires availability of numerical data.

3.3.2. Non-hierarchical systems

Elementary kernels and Systematic Intersection

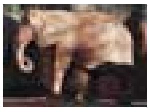
This method aims at bringing out empirically from different existing classification systems some kernels which will set up the beginning of main categories. It implies the gathering of existing nomenclatures on the domains to be considered and their confrontation and analysis in order to answer the different user requirements. This method has the objective of comparing existing nomenclatures assuming that major aggregates could be common for many users or approaches. General aggregates, commonly accepted through various nomenclatures, would constitute the core (kernels) and items, which may be allocated to different aggregates according to the different nomenclatures, would constitute the margins. For example, temporary and artificial pastures would be part of “arable land” for agriculture nomenclatures and would be part of another category for other users (POIRET 1997). The number of nomenclatures to consider is however to be limited since the more classifications are taken into account, the fewer kernels will be identified, since the probability of discovering objects not belonging to the core increases as different points of view are encountered.

A similar approach is the one termed systematic intersection of nomenclatures. In (RADERMACHER 1988) a systematic intersection of two basic nomenclatures is proposed: one on land cover, one on land use. The results of this work is a table where 9 land covers and 14 land uses have been systematically crossed, resulting into 126 classes of which 75 were empty (not relevant classes such as “water used for forestry”). Of the 51 remaining classes, 9 groups have been chosen according to specific purposes (mainly ecological purposes). Of these 9 classes, 3 classes retain a purely cover dimension (the cover aspect has been judged fundamental for wooded areas, other natural areas and water independently from the use). The six other classes mix cover and use or indicate another dimension: for example the category “Areas without any ecological relevance”.

Systematic approaches through classifiers

Instead of a universal classification system on land cover and land use, there is a need to develop tools aimed at facilitating the linkage between systems; tools which should be, as far as possible, independent of the various constraints linked to construction of nomenclatures and the specific domain of land cover and land use which a tool could consist of a combinatory system applied on a common basis. This basis would be just a set of necessary characteristics to describe the objects. These characteristics, once identified and defined uniformly will allow, through combinations, the definition of the objects and the grouping of the objects for all possible systems.

Describing an object is to account for its characters (EKHOLM 1996). Characters may have different "expressions": character states. Different types of characters may be encountered: two-state characters: present / absent, positive / negative, +/-, A/not A, 1/0; multi-state characters: quantitative states (continuous or discrete), qualitative states (terminological). Characters may be used in the decision rule of classifying objects into a given classification system. The following table 3.1 based on an example of elephants, whales and anteaters illustrates this decision chain (DUHAMEL 1998).

	ANIMALITY	SIZE	SKIN COLOUR	PRESENCE OF TRUNK	NAME (IMAGE) in English	CODE NOMENCLATURE X
	YES	BIG	GREY	YES	ELEPHANT	AH B200
	YES	BIG	GREY	NO	WHALE	AV Z300
	YES	MEDIUM	BLACK and WHITE	YES	ANT EATER	AB K541

Object	**Concept** (standardised through a set of characters)	**Image**

Table 3.1: Example for the application of classifiers

Four characters have been arbitrarily chosen for illustrating the principle of description of objects through relevant characters: a two-state character ANIMALITY (states YES/NO), a three state-character SIZE (states BIG, MEDIUM, SMALL), a multi-state character SKIN COLOUR and a two state character PRESENCE OF TRUNK (States YES/NO). It can be said that the possession of the 4 following character states [ANIMALITY: YES; SIZE: BIG; COLOUR OF SKIN: GREY; PRESENCE OF TRUNK: YES] define objects termed "ELEPHANTS". These characters may be systematically used as **diagnostic criteria, or classifiers**. These classifiers are conceptual representations of characters of objects, a decision rule that specifies for each observation (object) which class to assign. Classifiers represent characteristic properties relevant to the objectives of the nomenclature.

This approach may be used for building classification systems (as exists for example in the field of soil taxonomy). One fundamental question is the one linked to the problem of hierarchy of classifiers, which means that one can lay down *a priori* rules, separating categories into sub-categories and so on. This is an *a priori* method in fact resulting in hierarchic schemes where *a priori* and unequal weighting is proposed. We are therefore again faced with a traditional way of buildings tree classifications where successive partitions taking into account "objective" characteristics / characters are following this scheme: a structure, a rank is given at different successive partition levels on a series of characters. The solution is to develop, instead of hierarchical and rigid schemes, a "flat" combinatory system applied on a common basis of classifiers of equal weight, this basis just being a set of necessary characteristics to the description of the objects. These characteristics, once identified and defined uniformly allow, through combinations, the definition of the objects and the grouping of the objects for all possible nomenclatures. It is difficult to develop such a system without proposing a lot of classifiers allowing discrimination of all the objects on land cover and land use. Some "ranking" is generally proposed to limit the number of classifiers. Therefore, two successive approaches are proposed:

- general classifiers of higher rank which may be freely combined (faceted);
- descriptors of lower rank which are specific to a particular domain (adopting a principle of economy to eliminate redundancy and limiting the number of classifiers).

4. DATA COLLECTION TOOLS

4. DATA COLLECTION TOOLS

The following chapter illustrates different data collection tools available for primary data gathering used for land cover and land use information. The section aims at describing the basic principles, the underlying concepts and the differences in conceptual ideas, accompanied by a brief summary of the advantages and the limits. The technical explanations of data collection tools are necessary for the assessment of information systems and quality of the information return.

4.1. Space remote sensing imagery

4.1.1. Definition

"Remote sensing is the science (and to some extent, art) of acquiring information about the earth's surface without actually being in contact with it. This is done by sensing and recording reflected or emitted energy and processing, analysing, and applying that information" (Canada Centre for Remote Sensing (CCRS): Fundamentals of Remote Sensing 1999).

4.1.2. Principles

The principle of remote sensing can be briefly summarised as follows:

The earth's surface is illuminated by a wide spectrum of electromagnetic radiation coming from the sun. Table 4.1 shows different parts of the electromagnetic spectrum ranging from the ultraviolet part up to the far infra-red region.

Spectral range		**Wavelength in µm (~10^{-6} m)**
Ultraviolet:		**< 0.3**
Visible part of the spectrum	**Violet:**	**0.4 - 0.446**
	Blue:	**0.446 - 0.500**
	Green:	**0.500 - 0.578**
	Yellow:	**0.578 - 0.592**
	Orange:	**0.592 - 0.620**
	Red:	**0.620 - 0.7**
reflected IR		**0.7 - 3.0**
thermal IR		**3.0 - 100**

Table 4.1: The electromagnetic spectra

The radiation coming from the sun is interacting with the atmosphere and does not reach the earth's surface completely uninfluenced. Passing through the atmosphere a large part of the energy is absorbed. The stratospheric ozone for example absorbs major parts of the ultraviolet radiation. The water content in the atmosphere is responsible for the absorption of specific parts of the infra-red radiation. Only for some wavelengths the atmosphere is to a large extent transparent, particularly in the visible proportion of the electromagnetic spectrum. These "atmospheric windows", where the energy transmission is effectively undisturbed, are used in remote sensing.

All objects on the earth's surface (targets) interfere with the radiation. Targets reflect, transmit or absorb the incoming electromagnetic waves. The process taking place depends on the physical and chemical structure of the target and the wavelength involved.

The reflected part of the spectrum is the most important for remote sensing applications dealing with land. Over the different wavelengths, targets reflect in a specific, and in some cases unique, way (figure 4.1).

This characteristical spectral response of objects enables their identification by means of remote sensing. Comparing the response patterns of different features of the earth's surface in different spectral ranges makes the distinction between objects possible.

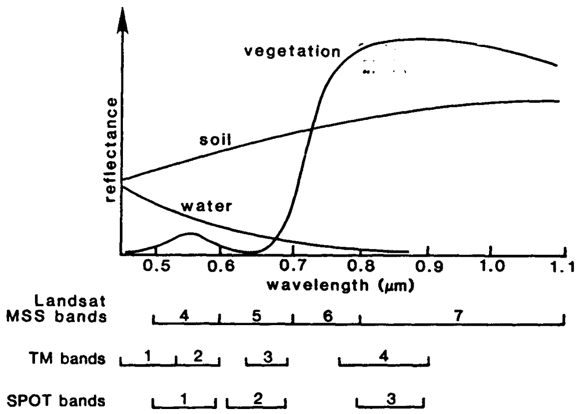


Figure 4.1: Spectral response of some common surfaces and spectral bands of satellite sensors

The figure above shows the spectral response of some common surfaces:

- **Green vegetation**: Chlorophyll, a chemical compound in leaves, strongly absorbs radiation in the red and blue wavelengths but reflects green wavelengths. Therefore leaves appear green in the summer, when chlorophyll content is at its maximum. In autumn, there is less chlorophyll in the leaves, so there is less absorption and proportionally more reflection of the red wavelengths, making the leaves appear red or yellow. The internal structure of healthy leaves act as a strong reflector of Near-Infra-red wavelengths. The Near-IR / Red ratio is the basis for many vegetation indices, used for vegetation monitoring. The specific reflection properties of plants enable the identification of different plants.

- **Water**: Water absorbs more the longer wavelength in the visible range and the near infra-red radiation than shorter visible wavelengths. Thus water typically looks blue or blue-green due to stronger reflectance at these shorter wavelengths, and darker if viewed at red or near infra-red wavelengths. If there is suspended sediment present in the upper layers of the water body, then this will allow better reflectivity and a brighter appearance of the water.

- **Soil and minerals**: The reflection patterns of soils exhibit stronger spectral features. The reflection depends mainly on the mineral composition, the grain size, the water and the organic content of the soil. The dryer and purer the soils, the lower emissivities in this range of the spectrum.

The spectral response can be quite variable even for the same target type, and can also vary with time and location. In addition the spectral response is influenced by the atmospheric conditions during the image acquisition because the radiative transfer is strongly dependent on the water and dust content in the atmosphere.

Knowing where to "look" spectrally and understanding the factors influencing the spectral response of the features of interest are crucial for correct interpretation of remotely sensed images and their results.

4.1.3 The measurement process: scanning

Remote sensors acquire data using scanning systems, which employ a sensor with a narrow field of view that sweeps over the terrain to build up and produce a two-dimensional image of the surface (raster image, figure 4.2).

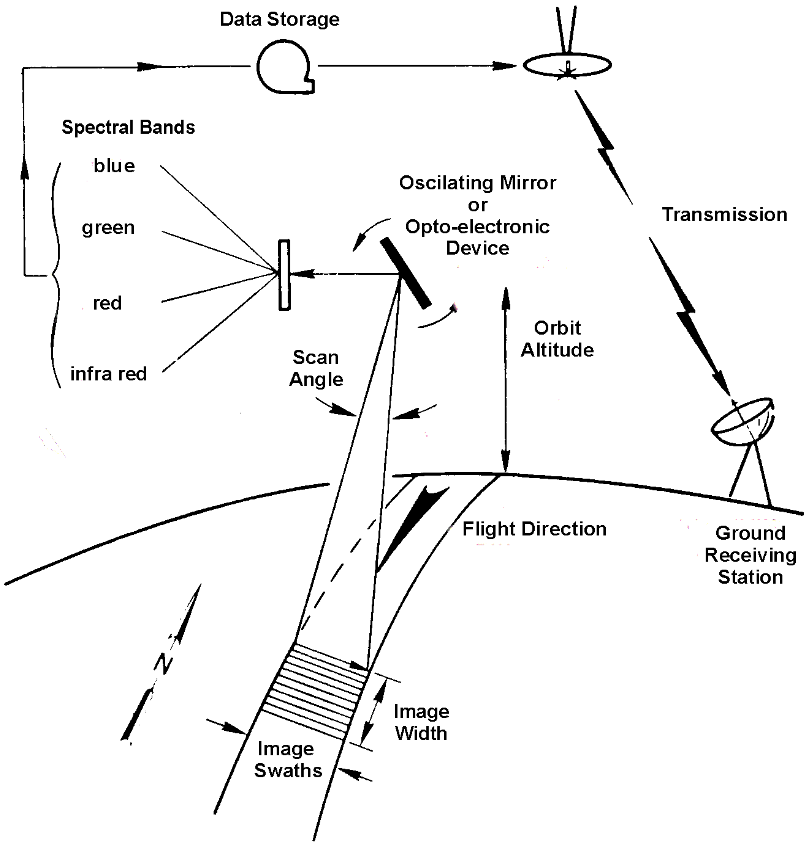


Figure 4.2: Principle of data recording, scanning process

Scanning systems can be used on both aircraft and satellite platforms and have essentially the same operating principles. The scanning systems measuring the reflected (or emitted) energy simultaneously over a variety of different wavelength ranges (spectral bands) are called multispectral scanner. Both, the size of the raster cells (or Picture Elements or pixels) and the wavelength ranges measured depend on the technical specification of the sensor (spatial and spectral resolution).

A Scanning system has several advantages over photographic systems. The spectral range of photographic systems is restricted to the visible and near-infra-red regions while a multispectral scanner can extend this range into the thermal infra-red. They are also capable of much higher spectral resolution than photographic systems. Multispectral photographic systems use separate lens systems to acquire each spectral band simultaneously. Photographic systems record the energy detected on an emulsion on film by means of a photochemical process that is difficult to measure and to make consistent. Because scanner data are recorded electronically, it is easier to determine the specific amount of energy measured, and they can record over a greater range of values in a digital format.

Satellite platforms

Some of the known satellites and imaging sensors are shortly described in the following:

- Landsat

Since 1972 the American satellite Landsat delivers multi spectral imagery of the earth's surface. Landsat Satellites 1–3, operational until 1983, were equipped with the Multi Spectral Scanner (MSS) with the following specifications: 4 bands, image size 185km*185km, pixel size 80m*80m. Landsat 4 (launched 1982) and Landsat 5 (launched 1984) carried in addition to the Multi Spectral Scanner the Thematic Mapper sensor (TM), characterised by 7 spectral bands, image size of 185km*185km, pixelsize 30m*30m). The payload of Landsat 7, operational since April 1999, consists of an enhanced Thematic Mapper sensor (ETM). In addition to the multispectral bands, similar to those of Landsat 5, the ETM sensor scans the earth in a panchromatic band with a pixelsize of 15m*15m. During the operational phase of the Landsat satellite programme an extensive archive of satellite images was created, which offers a retrospective view and the analysis of changes.

- SPOT

The French SPOT satellite programme ('Système Probatoire d'Observation de la Terre') was initiated in 1986 with SPOT 1, followed by SPOT 2 in 1990 and SPOT 3 in 1993. The HRV (Haute Resolution Visible) sensor on board delivers imagery of 3 spectral bands with a pixelsize of 20m*20m and a panchromatic band with 10m*10m pixelsize taken from an orbit altitude of 830 km. The SPOT 4, launched in 1998, carries also the so called "Vegetation Instrument" producing imagery with 4 bands (blue, red, near infra-red and short wave infra-red) with a pixelsize of 1km*1km and a width of 2250 km. It is planned to launch in 2000 SPOT 5 with a sensor producing panchromatic images with 5m*5m pixelsize.

- IRS

The first Indian Remote Sensing satellite (IRS-1A) was launched in 1988. The sensors carried by the most recent IRS-1D platform produce a panchromatic image with a 5.8m*5.8m pixelsize, a 4 band imagery with 23.5m*23.5m pixelsize (LISS) and a 2 band imagery with 188m*188m pixelsize (WiFS).

- IKONOS

The IKONOS satellite, the first commercial high resolution satellite, is operational since end of 1999. The satellite, flown at an altitude of 681 km, carries two independent sensors, scanning the earth's surface in a strip of 11 km width and up to 1000 km length. With a ground resolution of 1m*1m (panchromatic) and 4m*4m (multispectral) the sensor provides high quality images which are close to aerial photography, enabling new applications of satellite remote sensing imagery in a wide field of subjects.

- NOAA - AVHRR

The Advanced Very High Resolution Radiometer (AVHRR) on board the NOAA (National Oceanic and Atmospheric Administration) satellites was originally designed for meteorological purposes. Therefore the technical specifications are quite different from those satellites, developed for land applications (like Landsat, SPOT or IRS). The AVHRR sensor records the spectral reflectance in the red and near infra-red wavelength and the emitted energy (temperature) in the middle and long infra-red region. Information about the temperature, particularly of clouds, is of specific interest for meteorologists. The broad spatial resolution of the AVHRR sensor of 1.1 km and the wide scan angle of 55° enable a synoptic view of huge areas of about 2000 km width (Landsat TM: 185 km). The technical properties and the fact that at least 2 satellites are in operation at the same time, enables the provision of a daily coverage of the entire earth and allows a continuous monitoring, which is of particular importance for weather forecast and other meteorological subjects.

Figure 4.3: IRS III image of the Rhine valley north of Karlsruhe (2nd November 1997 (NIR, Red, Green = RGB; spatial resolution: 23 m); Source: Euromap, Neustrelitz (www.euromap.de)

Figure 4.4: IRS 1 C Panchromatic image of Amsterdam's airport Schiphol 17th May 1998 (spatial resolution: 5,8 m) Source: Euromap, Neustrelitz (www.euromap.de)

- RADARSAT, ERS

From the conceptual and technical point of view radar satellite systems are quite different from those mentioned up to now. In contrast to sensors, measuring radiance coming from the sun and reflected or emitted from the earth's surface (passive systems), radar sensors actively send their own signals and record the reflected proportion of that signal. The Structural properties of the targets determine the way of the reflection, thus enabling their identification. The advantage of radar systems is the capability to penetrate clouds, so that the image acquisition is independent from the atmospheric conditions. Radar systems are of minor importance for land cover/land use mapping applications and therefore not exhaustively treated here. They are more frequently used e.g. for geological purposes, sea ice or oil spill detection. The Canadian RADARSAT and the European ERS-1 and -2 satellites are two radar satellite systems working operationally.

4.1.4 Remote sensing images: properties

The technical specifications of the sensors and the orbit characteristics of the satellite platforms mainly determine the capabilities and the potential applications of remote sensing images. The most relevant issues are explained in the following chapter.

Spectral resolution

As mentioned above, features or targets of the earth's surface can be characterised and distinguished by the spectral reflectance over a variety of wavelengths. Satellite sensors measure the reflected radiation of the surface in different spectral intervals, so-called spectral bands or channels, in order to capture these differences. The capability of a satellite sensor to identify targets on the earth's surface depends to a great extent on the number of spectral bands, the so-called spectral resolution.

The spectral resolution describes the ability of a sensor to distinguish between fine wavelength intervals. Remote sensing systems record the reflected proportion of radiation in several separate wavelength ranges (so-called spectral bands or channels) at various spectral resolutions. The finer the spectral resolution, the narrower the wavelength range for a particular channel or band, the better different objects can be detected and distinguished. Advanced multi-spectral sensors, called hyperspectral sensors, detect hundreds of very narrow spectral bands throughout the visible, near infrared and mid-infrared portions of the electromagnetic spectrum. Their very high spectral resolution facilitates fine discrimination between different targets based on their spectral response in each of the narrow bands.

Table 4.2 shows the spectral resolution of common sensors used for land applications. With 7 spectral bands, the Landsat Thematic Mapper sensor enables the best discrimination of objects, while SPOT and IRS operates only with 3 respectively 4 spectral bands.

SPOT, IRS and the latest Landsat ETM sensors also operate in a panchromatic mode, scanning the earth in a broad wavelength range (visible spectrum) and with a higher spatial resolution than the multi-spectral bands.

Sensor	Spectral resolution: Channel	Spectral resolution: Wavelength Range (in μm)	Spatial resolution	**Recommended maximum working scale (approx.)**
Landsat MSS	**MSS 1**	**0.5 - 0.6 (green)**	**80m *80 m**	1:500.000
	MSS 2	**0.6 - 0.7 (red)**		
	MSS 3	**0.7 - 0.8 (near infrared)**		
	MSS 4	**0.8 - 1.1 (near infrared)**		
Landsat TM/ ETM	**TM 1**	**0.45 - 0.52 (blue)**	**30 m*30 m**	1:200.000
	TM 2	**0.52 - 0.60 (green)**		
	TM 3	**0.63 - 0.69 (red)**		
	TM 4	**0.76 - 0.90 (near IR)**		
	TM 5	**1.55 - 1.75 (short wave IR)**		
	TM 6	**10.4 - 12.5 (thermal IR)**	**120 m* 120 m**	
	TM 7	**2.08 - 2.35 (short wave IR)**		
Landsat ETM	**Panchromatic**	**0.52-0.9**	**15 m*15 m**	**1:100.000**
SPOT Pan	**Panchromatic**	**0.51 - 0.73 (blue-green-red)**	**10 m*10 m**	**1:50.000**
SPOT XS Multispectral	**Band 1**	**0.50 - 0.59 (green)**	**20 m*20m**	**1:100.000**
	Band 2	**0.61 - 0.68 (red)**		
	Band 3	**0.79 - 0.89 (near infrared)**		
SPOT Vegetation	**Channel 1**	**0.50-0.59 (green)**	**1000 m* 1000 m**	**1:1.5 Mio**
	Channel 2	**0.61-0.68 (red)**		
	Channel 3	**0.79-0.89 (near infrared)**		
	Channel 4	**1.58-1.75 (short wave IR)**		
IRS Pan	**Panchromatic**	**0.5 – 0.75 (blue-green-red)**	**5.8 m*5.8m**	**1:15.000**
IRS LISS	**Band 2**	**0.52 - 0.59 (green)**	**23.5 m * 23.5 m**	**1:100.000**
	Band 3	**0.62 - 0.68 (red)**		
	Band 4	**0.77 - 0.86 (near infrared)**		
	Band 5	**1.55 - 1.7 (short wave IR)**		
IRS WiFS	**Band 3**	**0.62 - 0.68 (red)**	**188 m * 188 m**	**1:500.000**
	Band 4	**0.77 - 0.86 (near infrared)**		
NOAA AVHRR	**Channel 1**	**0.58- 0.68 (red)**	**1100 m* 1100 m**	**1:1.5 Mio**
	Channel 2	**0.725 – 1.1 (near IR)**		
	Channel 3	**3.55 – 3.93 (middle IR)**		
	Channel 4	**11.3 – 11.3 (thermal IR)**		
	Channel 5	**11.4 – 12.4 (thermal IR)**		
IKONOS	**Panchromatic**	**0.45 – 0.90**	**1 m * 1 m**	**1:5000**
multispectral	**Channel 1**	**0.45 – 0.52 (blue)**	**4 m *4 m**	**1:15.000**
	Channel 2	**0.52 – 0.60 (green)**		
	Channel 3	**0.63 – 0.69 (red)**		
	Channel 4	**0.79 – 0.90 (near IR)**		

Table 4.2: Spectral resolution, spatial resolution and recommended working scale (approximation) of common remote sensing instruments

Spatial Resolution

Remote sensing images are composed of a matrix of picture elements, or pixel, which are the smallest units where spectral response is measured. The size of the pixel varies according to the technical design of the sensors.

Images where only large features are visible are said to have coarse or low resolution. In fine or high resolution images, small objects can be detected. Commercial satellites provide imagery with resolutions varying from a few metres to several kilometres.

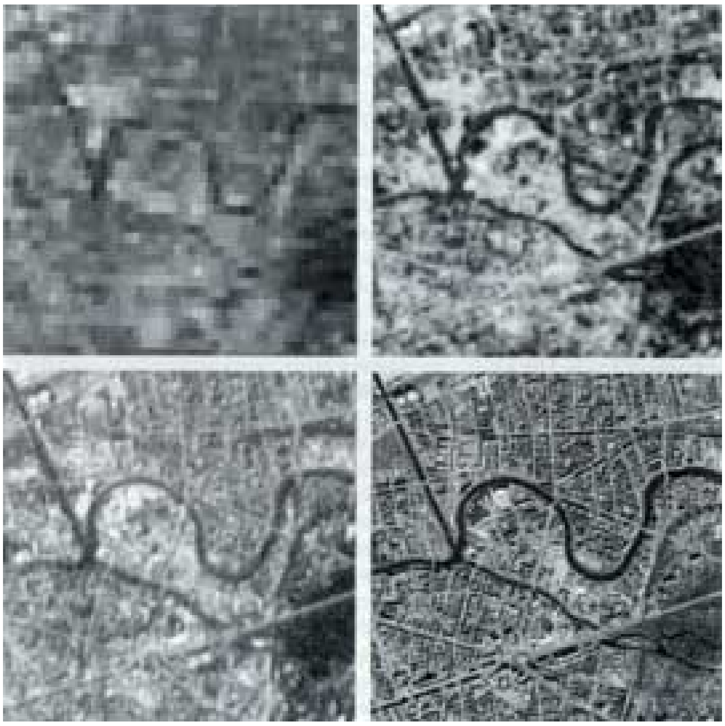


Figure 4.5: Subset of satellite images with different spatial resolution of Berlin (D): upper left: MSS (80m); Upper right: TM (30m); lower left: SPOT Multispectral (20m); lower right: SPOT panchromatic (10m) (ALBERTS 1991)

The spatial resolution has important implications for the identification of objects on the surface, the scale of the analysis, the locational precision and accuracy. The higher the spatial resolution, the more complete and precise the shapes of the objects are captured, the more can be identified based on their shape and the more accurately the location, extent and area of objects can be determined (see figure 4.5).

For a homogeneous feature to be detected, generally its size has to be equal to or larger than the pixel. If the feature is smaller, it may not be detectable because the average brightness of all features in that pixel will be integrated over the pixel area and then recorded. However, smaller features may sometimes be detectable if their reflectance dominates within a particular pixel thus allowing sub-pixel detection.

The spatial resolution of the data must be compatible to the project objectives. New developments in remote sensing are directed towards high spatial resolution images of about 1 to 2 meters, enabling the identification of small features.

Radiometric Resolution

Radiometric resolution refers to the dynamic range, or the number of different output levels used to record the radiant energy for a single measurement. The dynamic range of the most common satellite data is 7 bits or 128 different levels (Landsat MSS, IRS), or 256 levels (8 bits) for Landsat TM, ETM.

The greater the radiometric resolution, the more accurately the remotely sensed data can represent variations in surface leaving radiance. Many image processing software tools are designed to process 8-bit data, and other byte sizes might require special handling.

Temporal Resolution

In addition to spatial and spectral resolution, the concept of temporal resolution is also important to consider in a remote sensing system. The temporal resolution depends on the revisit period, which refers to the time it takes for a satellite to observe and image the same area on the ground at the same viewing angle a second time after completion of one entire orbit cycle (see figure 4.6 and 4.7).

The revisit period of a satellite sensor is usually several days (see table 4.3). Therefore the absolute temporal resolution of a remote sensing system to image the exact same area at the same viewing angle a second time is equal to this period.

Remote Sensing platforms	**Temporal Resolution**
IKONOS	4 Days
Landsat 1,2, and 3	18 Days (Every 251 Orbits)
Landsat 4,5, and 7	16 Days (Every 233 Orbits)
SPOT	26 Days
NOAA-AVHRR	1 day

Table 4.3: Temporal resolution of some common satellite platforms

However, because of some degree of overlap in the imaging swaths of adjacent orbits for most satellites and the increase of this overlap with increasing latitude, some areas of the earth tend to be re-visited more frequently. Also, some satellite systems (e.g. SPOT, IKONOS) are able to point their sensors by panning to image the same area between different satellite passes, separated by periods from one to five days. Thus, the actual temporal resolution of a sensor depends on a variety of factors, including the satellite/sensor capabilities, the swath overlap, the orbit altitude and the geographic latitude of the area of interest.

The most prominent factor influencing the temporal resolution is the fact, that cloud free conditions are required during the image acquisition. The chance to get cloud free images for a certain region in successive orbits is limited. The figures in table 4.3 are more theoretical than practical.

However, the ability to collect imagery of the same area of the earth's surface at different periods in time is one of the most important elements for applying remote sensing data. Spectral characteristics of features may change over time and these changes can be detected by collecting and comparing multi-temporal imagery. For example, during the growing season, most species of vegetation are in a continual state of change and the ability to monitor those subtle changes using remote sensing is dependent on when and how frequently imagery is recorded.

By imaging on a continuous basis at different times it is possible to monitor the changes that take place on the earth's surface, whether they are naturally occurring (such as changes in natural succession of vegetation cover or flooding) or man-indicted (such as urban development or deforestation).

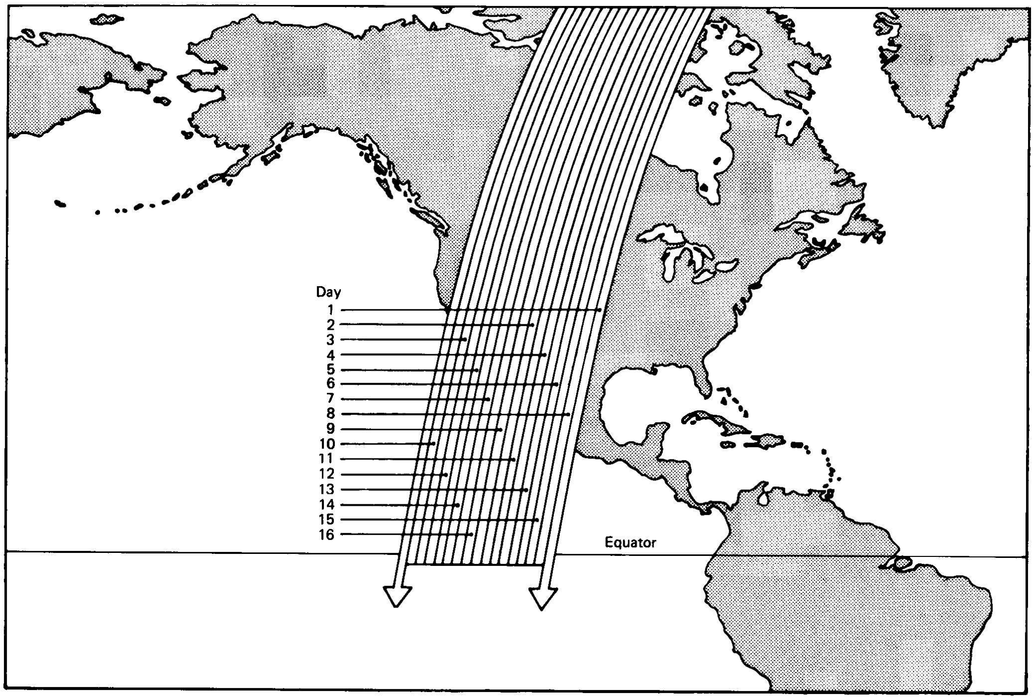


Figure 4.6: Timing of adjacent Landsat 4 and 5 tracks. Adjacent swaths are recorded 7 days apart (LILLESAND & KIEFER 1994).

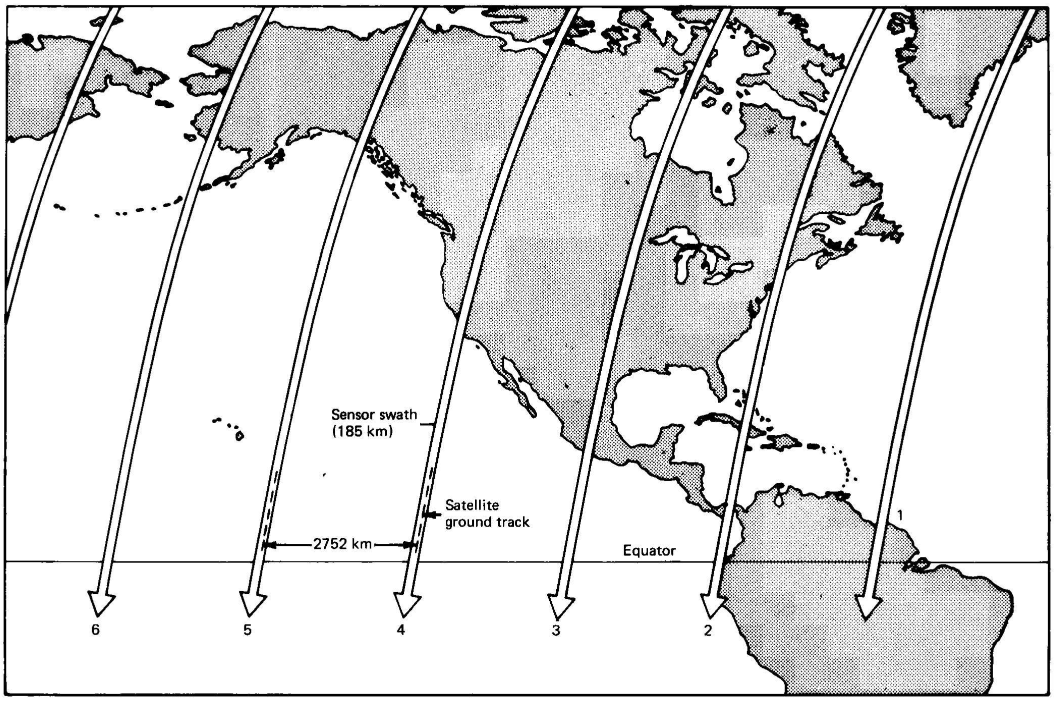


Figure 4.7: Spacing between adjacent orbit tracks (paths) of Landsat at the equator (LILLESAND & KIEFER 1994).

4.1.5 Digital Image Processing

The use of digital space remote sensing images involves digital image processing procedures. Only in a few cases data can be used directly. The obligatory steps in image processing prior to the data analysis are the following:

- Radiometric correction
- Geometric correction
- Image enhancement

Radiometric corrections are necessary due to sensor irregularities over time and the unwanted atmospheric influences. Processing is also required when physical units (like reflected or emitted radiation) are to be calculated. In particular, when remote sensing data are used for monitoring purposes, great care has to be taken on the radiometry.

Geometric correction deals with the conversion of the image matrix to “real world co-ordinates” and map projections.

Image enhancement concerns the improvement of the appearance of the imagery to assist in visual interpretation and analysis.

4.1.6 Image Analysis

Interpretation and analysis of remote sensing imagery involves the identification and/or measurement of various targets in an image in order to extract useful information. Targets in remote sensing images may be any feature or object that can be observed in an image. The target must be spatially and spectrally distinguishable i.e. it must contrast with other features around it in the image. Image Analysis is performed manually (analogue image analysis, e.g. visual interpretation) and/or digitally.

Manual interpretation and analysis dates back to the early beginnings of remote sensing for air photo interpretation. Digital processing and analysis is more recent with the advent of digital recording of remote sensing data and the development of computers. Digital image analysis is performed when the full spectral information (multi-channel data sets) is to be used. A variety of methods and approaches for specific purposes have been developed.

Concerning land cover information, digital image classification procedures are performed. Based on the spectral signature of required classes or categories (coniferous forests, grassland) the pixel’s spectral information is statistically assigned to one of them. The output of such a classification is a “thematic map”.

Image analysis does not rely solely on the digital image information but includes also auxiliary information like topographic maps, thematic maps, and digital terrain models. Above all, it should not be forgotten that ground truth surveys are essential for validating the results. The use of satellite images does not make ground surveys superfluous.

Both manual and digital techniques for interpretation of remote sensing data have their respective advantages and disadvantages. Generally, visual interpretation based on paper copies requires little specialised equipment. Digital processing and analysis needs at least a standard PC and special software (e.g. Erdas Imagine, ER Mapper). Visual interpretation is often limited to analysing only several channels of data or a single image at a time due to the difficulty in performing visual interpretation with multiple images. The computer environment is more amenable to handling complex images of many channels or from several dates. In this sense, digital analysis is useful for simultaneous analysis of many spectral bands and can process large data sets much faster than a human interpreter. In contrast to computer based analysis, the human factor enables the integration of multiple contextual information based on the a priori knowledge of the interpreter about the area under investigation. However, visual interpretation is to some extent a subjective process, meaning that the results may vary between different interpreters. Digital analysis is based on the manipulation of digital numbers in a computer and is thus more objective, generally resulting in more consistent results. However, determining the validity and accuracy of the results from digital processing can be difficult.

It is important to reiterate that visual and digital analyses of remote sensing imagery are not mutually exclusive. Both methods have their merits. In most cases, a mix of both methods is usually employed when analysing imagery. In fact, the ultimate decision of the utility and relevance of the information extracted at the end of the analysis process must still be made by human interpreters.

4.1.7 Advantage and Disadvantage of Remote sensing imagery

There is no doubt that remote sensing data represent a data source which contributes to a deeper understanding of processes on the earth's surface.

Remote sensing data provide a synoptic overview of large areas. The position, distribution and spatial relationships of features on the earth's surface are clearly evident. Thus spatial relationships can be examined.

Remote sensors look over a broader portion of the spectrum than the human eye, enabling the detection and identification of various environmental features of the earth's surface or the atmosphere, particularly when sensors focus on a very specific bandwidth.

Through repetitive looks at the same area, the data represent a unique data source for monitoring purposes and change detection. The use of remote sensing data for monitoring need some input of methodological work and pre-processing capabilities concerning geometric and radiometric correction, which is time and cost intensive.

Remote sensing data play an important role in regional planning and land use planning, particular where no basic data are available and the relatively small scale is still sufficient.

However, the use of satellite data and the ability for detection and identification of e.g. land cover classes depends on the spectral and spatial resolution of satellite sensors. The spatial resolution determines the scale of work. Common satellite imagery enables mapping at a scale of 1:50.000 or 1:100.000. In a highly structured landscape the spatial resolution of e.g. 20m*20m does not enable a sufficient discrimination of objects composing such an area. Consequences of the relatively broad spatial resolution are that maps derived from satellite imagery are at scales, which are not always appropriate. With new high resolution satellite systems, like IKONOS, this limit can drastically be reduced, enabling map production up to scales of 1:5.000.

4.2. Aerial photographs

Aerial photographs are black and white or colour pictures of the earth's surface taken by a film camera onboard an airplane, helicopter or balloon (NASA 1999).

It is important to distinguish between the terms 'images' and 'photographs' in remote sensing. An image refers to any pictorial representation, regardless of what wavelengths or remote sensing device has been used to detect and record the electromagnetic energy. A photograph refers specifically to images that have been detected as well as recorded on photographic film (Canadian Center for Remote Sensing 1999).

4.2.1. Principles

In contrast to remote sensing sensors the electromagnetic energy reflected or emitted from the earth's surface is recorded by a camera on a film. The photographic process uses chemical reactions on the surface of light-sensitive film to detect and record energy variations. Photos are normally recorded over the wavelength range from 0.3 µm to 0.9 µm - the visible and reflected infrared. In general the camera looks vertically straight to the ground, pointing straight down to the photo centre point (nadir).

The size of the photo, i.e. the area captured depends on technical and optical parameters of the camera (focal length, frame size) and the platform's altitude.

The scale of aerial photographs ranges from 1:60.000 (small scale) up to 1:1.000 (large scale).

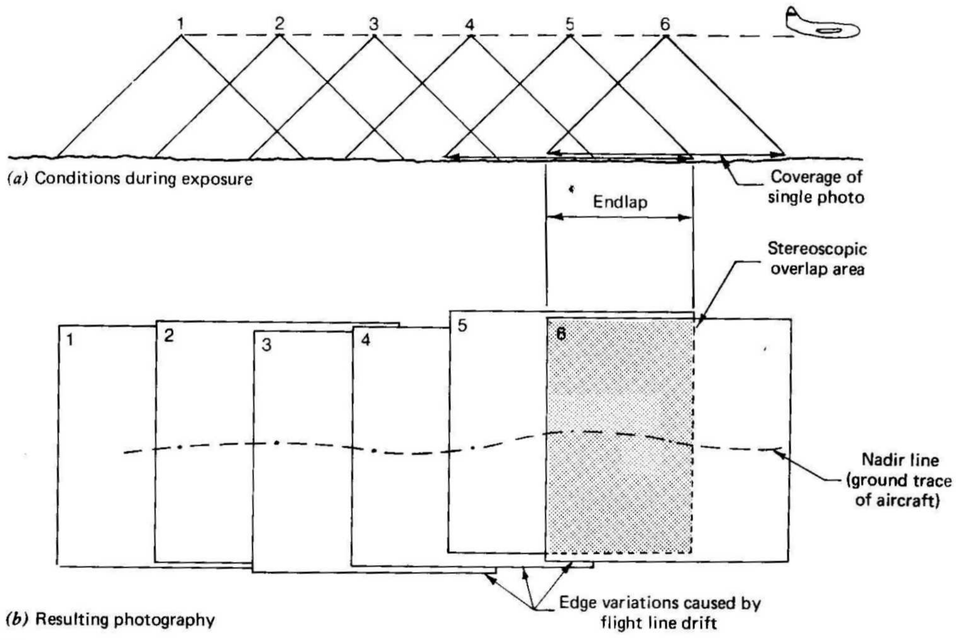


Figure 4.8: Photographic coverage along a flight strip (LILLESAND & KIEFER 1994)

A photograph could also be represented and displayed in a digital format by a simple scanning process. The image is subdivided into small equal-sized and shaped areas (pixels) representing the brightness of each area with a numeric value or digital number.

Although both maps and aerial photos present a "bird's-eye" view of the earth, aerial photographs are not maps. Maps are orthogonal representations of the earth's surface, meaning that they are directionally and geometrically accurate. Aerial photos, on the other hand, display a high degree of radial distortion. The topography is distorted, and until corrections are made for the distortion, measurements made from a photograph are not accurate.

Typically the area surveyed is traversed along back and forth flight allowing a 60% overlap of successive and a 20-40% overlap between lines (see figure 4.8). Due to the overlap of neighbouring photos, features on the surface are seen from two different viewpoints. The resulting parallax enables a three dimensional (stereoscopic) view of the terrain and allows retrieval of information about the height of single objects or the altimetry of entire relief (contour lines).

4.2.2. Types of Aerial Photography

Black and White

The film emulsion of black and white photos is adjusted slightly from regular film.

The first reliable black and white film was developed at the beginning of the century. With the aeronautical development the age of aerial photography was initiated. Due to a long tradition, national mapping agencies dispose of huge historic archives of aerial photographs, which reveal details in the changes of the landscape.

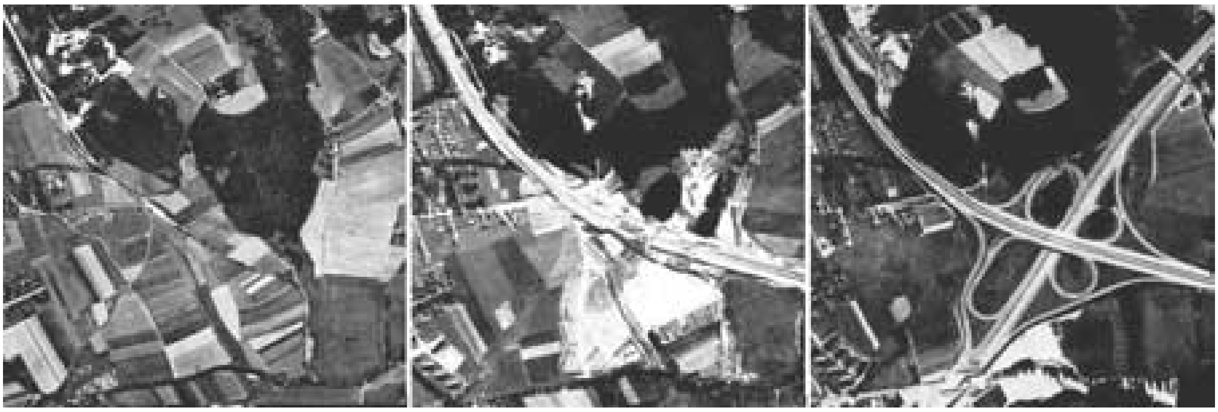


Figure 4.9: Example of a black and white aerial photograph time serie of Westkreuz Germany 1952 – 1959 – 1964 (HANSA LUFTBILD www.hansaluftbild.de)

Colour infrared film (CIR)

Colour infrared film is often called "false-colour" film. In contrast to black and white or colour films, the film emulsion is sensitive for infrared radiation. Healthy green vegetation is a very strong reflector of infrared radiation and appears bright red on colour infrared photographs. Objects that are normally red appear green, green objects (except vegetation) appear blue. The primary use of colour infrared photography is vegetation studies and forestry (see example figure 4.10).

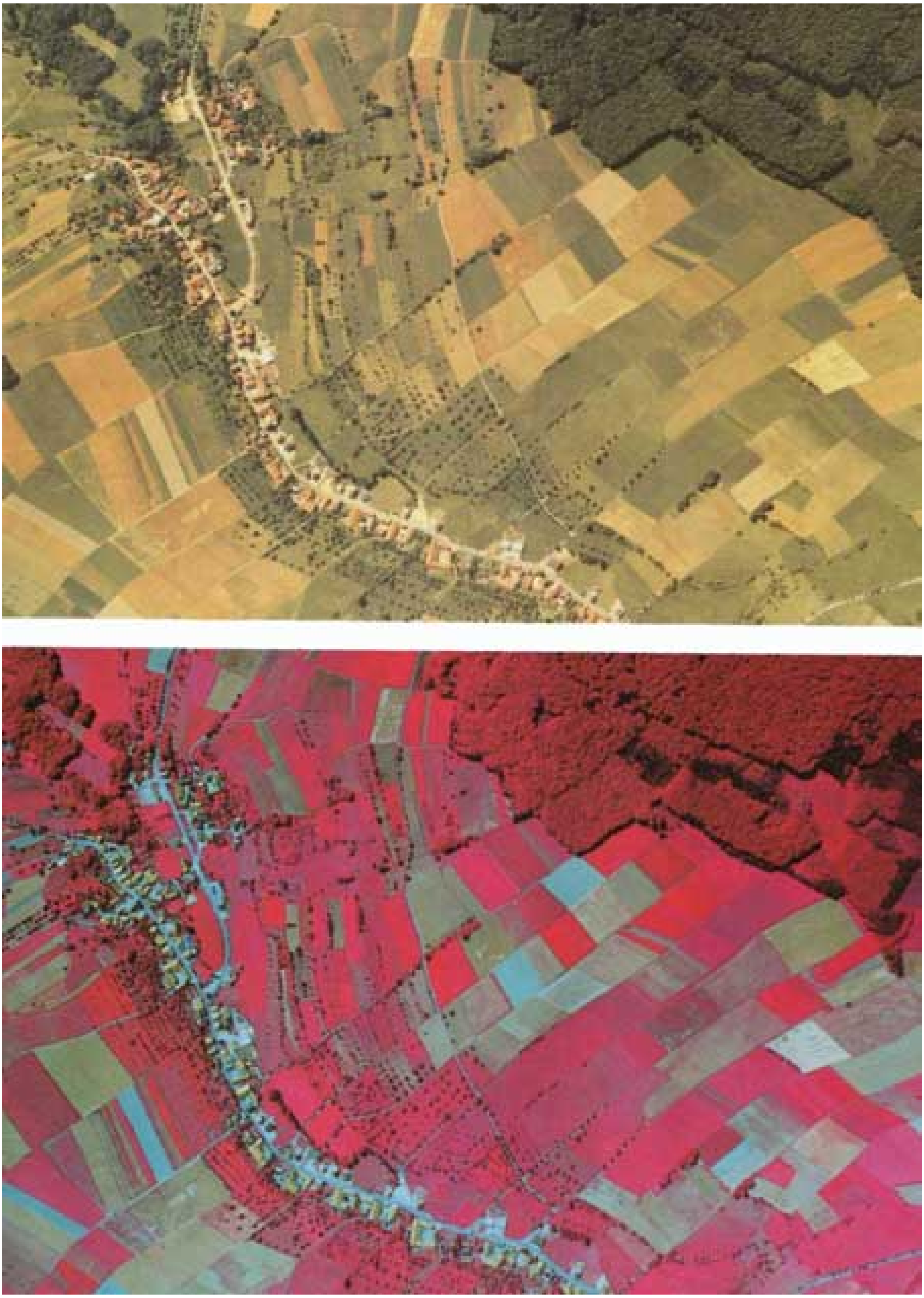


Figure 4.10: Example of a colour and a CIR aerial photography, Odenwald, Germany, date 30.7.67, original scale 1:14.000 (ALBERTZ 1991)

Orthophoto

An orthophoto (black and white, Colour or CIR) refers to an aerial photo, which is geometrically rectified, so that image displacements caused by camera tilt and relief of terrain are removed. Thus orthophotos have a consistent scale throughout the photo and due to the georeferencing process they represent a uniform spatial reference in a harmonised format. The photos are offered in digital formats, so that they can be integrated in GIS systems without any further processing.

Due to the high information content (pixel size up to 0.25 cm) they serve as a very precise data source for information retrieval.

In the framework of the IACS (**I**ntegrated **A**dministration **C**ontrol **S**ystem) an extensive archive of digital orthophotos was created, covering Italy, Greece, Portugal, Ireland, Denmark, Belgium, Finland, the southern parts of France and around 400.000 km^2 of Spain. The scanned back and white images (acquisition scale 1:40.000) are used to support farmers in preparing their declaration forms but also used by national authorities for validation and control in order to detect incompatible declarations of the farmers.

4.2.3. Basic Elements of Air Photo Interpretation

Aerial Photographs are mainly interpreted visually. Tones and tonal variations (grey or colour) and the resulting pattern made by these determine the pictorial content of an aerial photo. The following basic elements can aid in identifying objects on aerial photographs.

Tone (also called Hue or Colour) refers to the relative brightness or colour of elements on a photograph.

Shape refers to the general outline of objects. Regular geometric shapes are usually indicators of human presence and use. Some objects can be identified almost solely on the basis of their shapes.

Texture: The impression of "smoothness" or "roughness" of image features is caused by the frequency of change of tone in photographs. It is produced by a set of features too small to identify individually. Grass, cement, and water generally appear "smooth", while a forest canopy may appear "rough".

Pattern (spatial arrangement) formed by objects in a photo can be diagnostic. Consider the difference between (1) the random pattern formed by an unmanaged area of trees and (2) the evenly spaced rows formed by an orchard.

Site refers to topographic or geographic location. This characteristic of photographs is especially important in identifying vegetation types and landforms.

Association: Some objects are always found in association with other objects. The context of an object can provide insight into what it is. For instance, a nuclear power plant is not (generally) going to be found in the midst of single-family housing.

4.2.4. Advantages - Disadvantages

The advantages of aerial photographs are easy handling, which does not require any sophisticated hard- or software equipment. In general the large image scale enables the exact identification, description and delimitation of even small objects. Due to the high information content aerial photos are an excellent data source and reference and they are frequently used for supporting large scale ground surveys. With the availability of orthophotos as a standard product offered by mapping authorities this primary data source gathers additional added value.

Aerial photographs at large scale cover only small proportions of the earth's surface, so that a great number of photos might be necessary to cover larger areas. This in turn is cost intensive.

However, with the new generation of high resolution satellite sensors (e.g. IKONOS) the difference between aerial photography and space remote sensing images regarding spatial and spectral characteristics becomes smaller, enabling the provision of high quality data for the benefit of the user.

4.3. Sample Surveys - Area Frame Surveys

An alternative approach to acquire e.g. land cover or land use data or information about the agricultural sector (yields etc.) is the use of area frame surveys. In contrast to remote sensing based surveys, where the entire area is mapped, area frame sampling is based on the selection and observation of representative "area" samples. The purpose of sampling is to enable a valid generalisation to be made without studying the whole area under investigation.

Area frame surveys are designed for specific purposes (set of variables to be studied) and commonly used in agricultural statistics (see for example: FAO 1996; FAO 1998; COTTER & NEALON 1987).

4.3.1. Principles

Area frame sampling can be considered as a statistical approach, aiming at the retrieval of statistical figures for a predefined reference or restitution unit. The aim of area frame sampling is to interfere from sample parameters to true values of the population (entire area under investigation). A large set of statistical rules has been developed to enable the sampling results to be applied to the whole population.

Area frame surveys are defined "… as a sample survey in which at least at the final stage land areas are sampling units" (FAO 1996). "The units of an area frame survey are directly bound to a geographical area" (GALLEGO 1995).

In the following some elementary features of area frame sampling are explained enabling a "non statistician" a first insight to the principles.

Sampling Units

Area frame sampling basically deals with the division of land into pieces (primary sampling units) out of which a set of representative samples is taken. Three main types of sampling units can be distinguished:

- Points

  The point is the simplest sampling unit, represented as a dot in a map or an aerial photography. The point can exactly be located by its geographical co-ordinates.

- Lines

  Lines can also be used as sample units to select features along a transect, represented as a line in a map or an aerial photograph.

- Quadrates or polygons

  Quadrates or areas are commonly used by ecologists in the field and they are of particular interest for geographical sampling of vegetation. In contrast to points or lines, quadrates are difficult to handle because the exact location of the borders may be difficult to recognise. Aerial photographs are used to overcome this problem. With regards to the shape various approaches can be found, having all their specific advantages and disadvantages: regular square or irregular, with or without recognisable physical boundaries etc.

The selected set of samples are surveyed, i.e. the variables of interest (area acreage of crops or land cover/land use) information are gathered and finally projected for the whole territory or reference unit.

Sampling Design or Sampling Plan

There are different procedures for selecting the units, which are to comprise the sample.

- Random sampling procedures
- Systematic sampling
- Stratified sampling

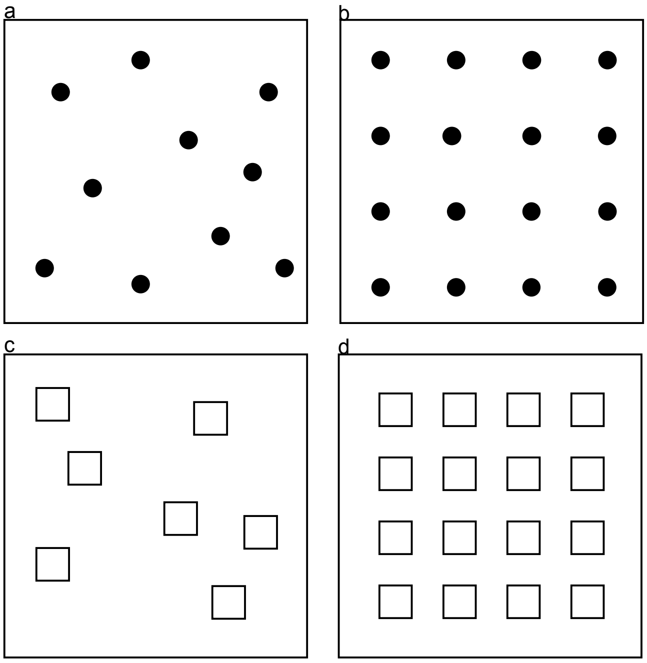


Figure 4.11: Examples of sampling units (a,b: points; c,d: quadrats) and selection by random (a,c) and systematic sampling (b,d)

Random Sampling

Random Sampling is the simplest selection of sampling units, which are to be observed. A simple random sample is one selected by a process giving every possible sample the same chance of selection (probability sampling). Figures above illustrates how an area is subdivided into sampling units (grid cells) from which a sample is taken and surveyed.

Systematic sampling

Systematic sampling is a method whereby only the first individual unit is selected at random or fixed based on specific requirements. All subsequent sampling units are chosen at a uniform interval from the first.

Stratified sampling

A stratified sampling is a more advanced way to select the samples, and is particularly useful if the phenomena under investigation are unevenly distributed or the study area is very heterogeneous. The approach is based on a division of the entire study area into several predefined sub-areas (strata), which are more homogeneous in respect of spatial distribution of the relevant attributes and characteristics. Within the different strata independent samples are selected, either randomly or systematically.

Most operational area frame surveys consider a stratification in order to improve statistical precision and to adapt the sample design to the local specific situation. The homogenisation is done according to *a priori* knowledge regarding specific attributes, reflecting the geographic distribution. In agricultural surveys such **strata** are for instance certain land cover/use characteristics. The delineation of strata and segments is commonly eased by the use of satellite images or aerial photographs.

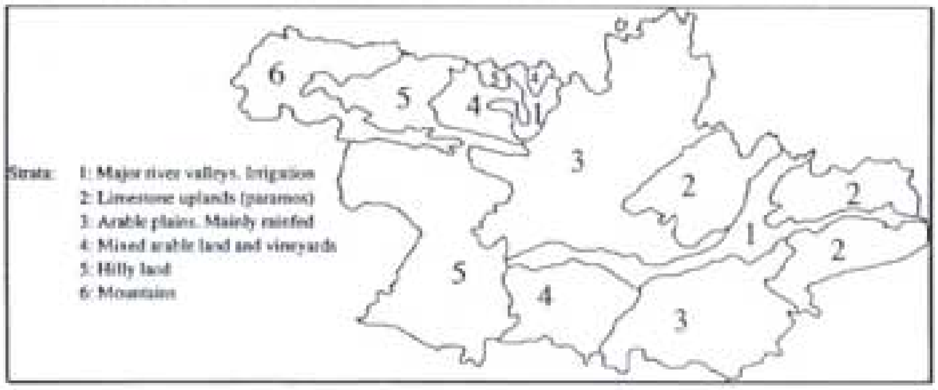


Figure 4.12: Example of a territorial stratification based on land use and terrain relief (TAYLOR et. al 1997)

The sampling units (points, lines, areas) and the selection procedure (random, systematic and stratified) represent the basic approaches. There are no strict or standardised rules to be used for defining and selecting the samples. In practice the sample design is determined by the variables to be observed, the required statistical precision, the financial and human resources available as well as the temporal delivery of results. The different approaches developed represent a compromise between these different aspects.

Sample Size

Crucial in the sampling plan is the number of sampling units selected for observation. The sampling size determines the required precision of the estimated characteristic of the population. However, the decision on the sample size also depends on the characteristics of the territory and the financial and human resources available for carrying out the survey.
It can be considered that with an increasing sampling size the precision of the estimate rises.

Output

In contrast to a mapping approach a sampling design provides only statistical figures related to a specific region. For each sampling design special statistical measures are used to extrapolate from the sample to the entire area under investigation (e.g. expansion factor). The art of the sampling lies in choosing the minimum size of the sample, which will give a reliable answer within a certain degree of confidence. The “best” estimator is an unbiased estimator with the smallest sampling variance.

The quality or precision of the estimates can be evaluated by means of different statistical measures, like the standard error of mean (SEM). For the various sampling methods an extensive set of statistical measures have been developed to evaluate the results and to improve successively the sampling design.

4.3.2. Advantages - Disadvantages

The advantage of area frame sampling is that only parts of the territory are observed. This allows for the gathering of very detailed and specific information, which cannot be provided by e.g. remote sensing. Area frame sampling is conducted mainly for agricultural estimates of crop areas but also employed for ecological purposes[1].

The advantage of area frame sampling is its relatively simple conduction. It enables periodic (annual or seasonal) information return and reliable timely.

In addition accuracy estimates are possible which judge reliability and efficiency.

From the conceptual point of view area frame sampling produces statistics which are valuable for general political decision making. For concrete planning purposes (e.g. land use planning) at local level, where exhaustive map data are required, this information is of relatively low importance.

The information return of area frame sampling refers only to predefined restitution units, which are in general administrative units. A spatial aggregation to higher levels is possible, but no disaggregation into smaller units is possible. This limitation is absent in exhaustive maps, where data can be grouped using almost all kind of boundaries (e.g. administrative units, landscape units, watersheds etc.).

[1] Ecological Area Frame Sampling in Germany, Countryside Survey UK.

4.4. Administrative data

4.4.1. Data Integration using Administrative and Statistic registers

Bearing in mind the definition of land use as presented in the introduction of this manual, it should be clear that only a limited set of land use parameters could be observed by remote sensing and aerial photographs.

The statistics will have to deal with the functional dimension and describe the use of areas by the socio-economic purpose. This information can be compiled by conducting surveys or censuses and/or by using already compiled information from administrative and statistical registers, especially for built-up areas (settlements).

The use of administrative data for statistical purposes is much depending on the original purpose of the registers. The purpose of the registers must be 'important' enough to guarantee high quality of the data in terms of accurateness and actuality. In addition, the land use information held in the registers must be relevant for that purpose. For example in ownership cadastral registers the land use information is collected but is not the most important part of it. The ownership will surely be updated but the land use information mustn't be. Another example is the Integrated Administrative Control System IACS where the land use information is most relevant but the register data does not cover the entire statistical population of farmers but only the ones requesting support.

The introduction of GIS-tools and geo-referenced information from registers has opened up for new and important data sources for land use statistics. Register information based on activity-data and technical information about buildings and ground properties can be used alone or combined with observations from satellite-images/orthophotos, to form a good total data source for deriving land use statistics.

4.4.2. Principles

The main principle is to establish a link between physical entities of land and the dominating socio-economic use that has given the land its characteristics. This can be achieved by linking both geo-coded and geo-referenced register information from a farm register, a building and dwelling register or a business register (for example: kind of activity, employees, turnover) with geo-referenced information from a building- and property register (for example: kind and size of building and property).

The street address or the cadastral code of the ground-property will normally serve as the key link between different registers so that information on activity etc. can be linked to building or ground properties with co-ordinates. Where high quality and harmonised use of addresses is not available, approximations may be used for geo-coding, for example small grids or census tracts/basic statistical units, parts of roads.

It is very important to have clear strategies for the final use and purposes of the statistics derived. Sometimes less accurate geo-referencing can be “good enough” for the end user and yield cost-efficient solutions.

Larger areas, which are dominated by a specific land use class of sites, can be aggregated in a GIS to for example industrial areas, residential areas etc.

Areas classified by land use can finally be combined and linked together by digital railroad- and/or road databases in models where interaction between entities of land can be followed through net work analyses.

4.4.3. Advantages-disadvantages

The advantages of this method for collection of land use information, is the direct link between statistics concerning human activity and the land on which these activities influence. A link between classification on land use and the activity classification of SITC or NACE can be ensured.

As more and more information becomes geo-referenced, it is possible to link demographic and economic parameters at micro level directly to the land use. This forms a comprehensive data set for monitoring and analysing status and changes of the use of land as well as we are able to yield important background information on pressure and driving forces.

Further more this method for aggregating land use statistics is cost efficient and the introduction of GIS gives a new dimension to the use of administrative and statistical registers. These registers are also continuously updated and thus they form good sources for covering of the dynamics of build-up areas.

The disadvantage is that the information stored in the registers is not always as good as it should be with respect to completeness, quality and timeliness. Registers do also not cover all the areas that actually should be classified as built-up. Geo-referenced information about green parks in urban areas as well as large areas covered with asphalt but with no buildings on it - such as parts of industrial areas, parking lots and big harbours - is normally difficult to find in a register. To some extent this is expected to be taken care of by using data from topographic maps on large scale in juxtaposition with cadaster maps.

Not all information from the registers is possible to use, because the process of geo-referencing may fail. Matching registers normally returns some unmatched information that can be very costly to deal with by manual routines.

The generalisation from a single building's site to the larger areas with near homogenous use of land is depending on a set of criteria and preconditions that always can be discussed. In addition a generalisation process means that some information will be lost.

5. INFORMATION REQUIREMENTS - DATA SPECIFICATION

5. INFORMATION REQUIREMENTS - DATA SPECIFICATION

Policy formulation and decision making are based on information. The better the quality of the information meaning adequate for the policy and sufficiently complete regards thematic detail and geographic coverage, the more realistic and useful decisions can be. In general, politicians and decision-makers do not express their needs for information in terms of data specifications. In the best case they describe the problem and the information required to formulate policy and to justify certain decisions. It is the task of the statistician **to translate**/interpret this 'information requirement' into concrete data specifications, and then to provide the adequate figures using the statistical information system. On the other hand it is extremely difficult for a statistician to assess in advance the potential information demand, its importance and the priorities of the user community.

To facilitate this crucial process **a dialogue** is absolutely necessary between policy/decision-makers and statisticians. The statistician needs **multi-disciplinary expertise** in the different policy fields to understand the problem and to be able to correctly specify the indicators required by politicians, and to **define** the **specifications** of the necessary **data** to inform these indicators.

For example, decision-makers at European level need to monitor the impact of the Common Agriculture Policy to decide on the continuation or change in the policy. In the dialogue to be held the problem will be further specified. For example if the economic situation of farmers is subject of monitoring to justify direct income subventions, statistics on the net income of farmers, the employment of family-members etc. can be of interest. If ecological issues are of concern, e.g. ground water pollution, statistics on fertilisers used, soil types and categories of land use and land cover are of interest. The statistician must be able to find the adequate thematic data at the adequate geographic breakdown to inform these indicators within the statistical information system.

5.1. What is the 'level' of information required?

When talking about data requirements of policy makers or decision making, a distinction should be made concerning the different policy levels (regional, national and international policies) and the kind of decisions to be taken (concrete actions or “only” a set up of specific frameworks and programmes). They all have very specific data or information demands regarding two basic aspects: a **thematic** and a **geographic** aspect[2].

From a **thematic viewpoint**, the 'level' of information refers in general to the **degree of detail.**

For example a 'low' level of information regarding land cover categories refers to a simple distinction between 'Built-up area' versus 'Non-built-up area'. A similar 'low' level of land use categories may refer to the 'Artificial Area', 'Utilised Agricultural Area' or 'Forest area'. To provide statistics on a low level of detail on land *cover*, the specifications for the required data are much more simple to describe than for low level data on land *use*.

High level information would for example refer to a detailed and highly disaggregated classification of land cover in agriculture.

Thematically high disaggregated information is the basis for a reasonable decision making process, both at international level as well as for local policies. The notion 'European Level' does not necessarily mean that the level of detail required is low. On the contrary thematic detail in general required for such purposes is quite high, e.g. surfaces set-aside (e.g. non-cropped agricultural areas eligible for financial support and now possibly used for e.g. renewable energy plants - high level of detail in category and space - set-aside for a certain period - high level of detail in time).

The following table shows the nomenclature of CLUSTERS (Classification of Land Use Statistics - Eurostat Remote Sensing), a pilot nomenclature developed by Eurostat in the framework of a its programme on Remote Sensing and Statistics (Eurostat 1996).

[2] The notion 'level' used in statistics should not be confused with the 'scale' used in cartography.

Table 5.1: Example for the thematic level of detail: CLUSTERS: Pilot Nomenclature on Land Use Statistics (Eurostat 1996)

	Level I		Level II		Level III		Level IV
A	Man-made areas	A1	Residential areas Public services	A11	Residential areas	A111	Continuous and dense residential areas
						A112	Continuous resid. areas of moderate density
						A113	Discontinuous resid. areas of moderate density
						A114	Isolated residential areas
						A115	Collective residential areas
				A12	Public services, local authorities	A120	Public services, local authorities
		A2	Industrial or commercial activities	A20	Industrial or commercial activities	A201	Heavy industry
						A202	Manufacturing industrial activities
						A203	Commercial and financial activities and service.
						A204	Agricultural holdings
		A3	Technical and transport infrastructures	A31	Technical infrastructures	A311	Technical network, productive structures
						A312	Water and waste treatment
				A32	Transport	A321	Road transport
						A322	Rail networks
						A323	Airports and aerodromes
						A324	River and maritime Transport
		A4	Extractive industries building sites Tips and wasteland	A41	Extractive industries	A410	Extractive industries
				A42	Building sites, tips and wasteland	A421	Building sites
						A422	Tips
						A423	Wasteland
		A5	Land developed for recreational purposes	A50	Land developed for recreational purposes	A501	Cultural sites
						A502	Sport facilities
						A503	Green or leisure areas
B	Utilised Agricultural areas	B1	Tilled and fallow land	B11	Cereals	B110	Cereals
				B12	Root and industrial crops	B121	Root crops
						B122	Non-permanent industrial crops
				B13	Vegetables and flowers	B131	Dry pulses
						B132	Fresh vegetables
						B133	Floriculture
				B14	Fallow land, inc. Green manure	B140	Fallow land, including green manure
		B2	Areas under grass used For agricultural purposes	B21	Temporary and artificial grazing	B210	Temporary and artificial grazing
				B22	Permanent pastures and grazing	B220	Permanent pastures and grazing
				B23	Rough grazing	B230	Rough grazing
		B3	Permanent crops	B31	Fruit trees and berries	B310	Fruit trees and berries
				B32	Citrus fruit	B320	Citrus fruit
				B33	Olive trees	B330	Olive trees
				B34	Vines	B340	Vines
				B35	Nurseries	B350	Nurseries
				B36	Permanent industrial crops	B360	Permanent industrial crops
C	Forests	C1	Wooded forest areas	C11	Deciduous trees	C110	Deciduous trees
				C12	Sclerophyllous trees	C120	Sclerophyllous trees
				C13	Conifers	C130	Conifers
				C14	Intensively managed plantations	C140	Intensively managed plantations
		C2	Non-wooded forest areas	C21	Clear-cut zones	C210	Clear-cut zones
				C22	Other unprod. forestry areas	C220	Other unproductive forestry areas
D	Bush or Herbaceous Areas	D1	bushes	D10	Bushes	D101	Brushy areas in temperate, mount., arctic reg.
						D102	Xerophyte bushes
		D2	Herbaceous vegetation	D20	Herbaceous vegetation	D201	Grassland in temperate, mount. arctic regions
						D202	Steppes and dry meadows
E	Surface with Little or no Vegetation	E0	Surface with Little or no Vegetation	E01	Bare soil	E011	Rocks and scree
						E012	Dunes and beaches
				E02	Glaciers and eternal snow	E020	Glaciers and eternal snow
				E03	Burned areas	E030	Burned areas
F	Wet surfaces And surfaces Under water	F1	Wet surfaces	F10	Wet surfaces	F101	Bogs and marshes
						F102	Moors
						F103	Other wet areas
		F2	Inland waters	F20	Inland waters	F201	Inland water courses and bodies of water
						F202	Ponds for fish
		F3	Coastal waters	F30	Coastal waters	F301	Estuaries, lagoons
						F302	Bodies of water with fish or shellfish

Turning to the **geographic point of view**, the 'level' of statistical information refers to the geographic breakdown, for example a restitution unit within a statistical regional reference system. The restitution unit is the unit in which statistical data are represented, they may have various dimensions of time (interval) and space (regions). Regarding the spatial aspect, statistical data is 'geo-coded' to a geographic extent, e.g. a specific town or locality, a region or a whole country or a group of countries like the European Union. It is anticipated that the statistical information is **harmonised and comparable** across the whole geographic extension of this Regional Reference System. At European level the Regional Reference System is the NUTS[3] (Nomenclature des Unités Territoriales Statistiques - Nomenclature of Statistical Territorial Units).

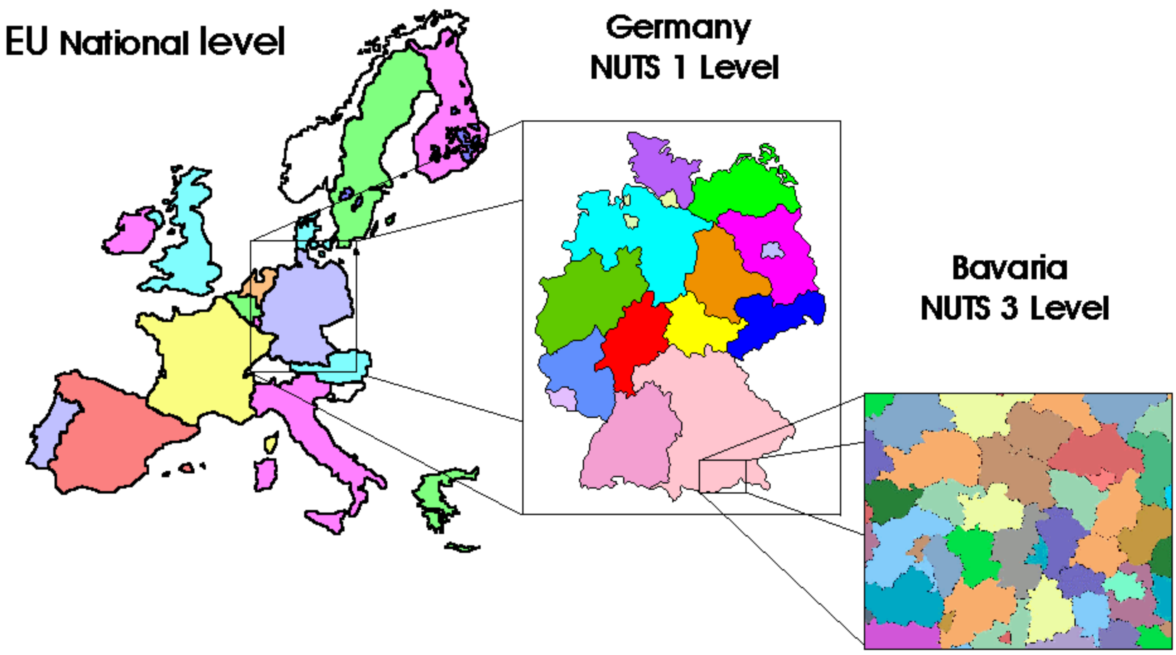


Figure 5.1: Example for the spatial level of detail: NUTS level 0 = EU Member States, Germany NUTS level 1 (= Länder) and Bavaria NUTS level 3 (=Kreise)

Regarding the spatial aspect of the information requirement, a distinction can be made between general policy formulation and concrete action planning. For international and national level, a spatially aggregated statistical information is sufficient. At local level, particularly for concrete planning purposes, a high spatial disaggregation is required, i.e. the more policy is approaching the local dimension, the higher the importance of a mapping approach.

The graphic below tries to illustrate and summarise the general trend: the higher the political level, the more spatially aggregated statistical information is required.

[3] When talking about statistics at 'European level' these figures need to be harmonised (comparable) across EU Member States at a certain regional breakdown (NUTS levels, whereas 'European' level = National data = NUTS level 0 = 15 member states, NUTS level 1 = 77 regions, NUTS level 2 = 206 provinces, NUTS level 3 = 1031 Counties, NUTS level 4 = 1074 districts, NUTS level 5 = 98433 communities).

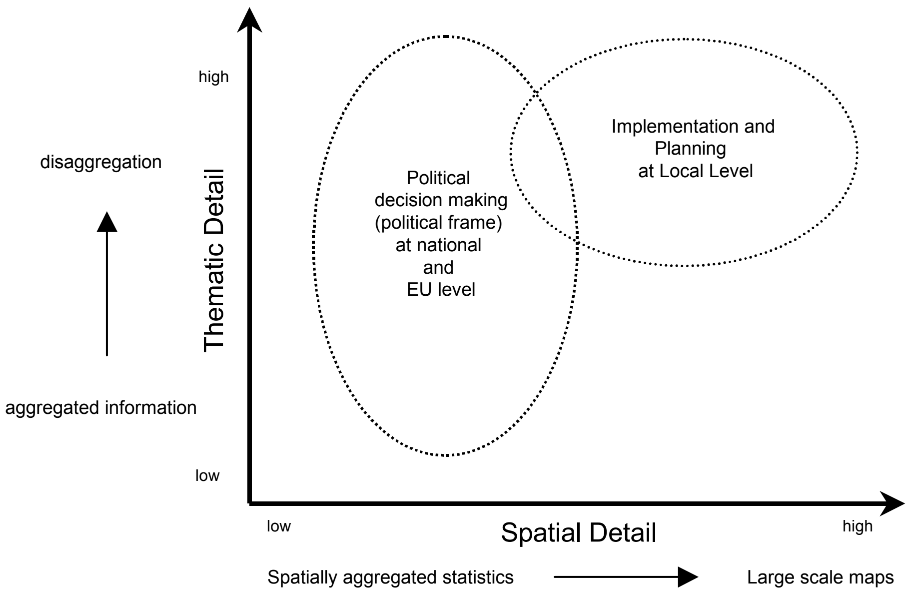


Figure 5.2: Data requirements and specification at different political levels

5.2. Multipurpose Information Systems

The statistical information systems aim at providing the **source (basic) data** to satisfy the information needs of various users in different thematic fields. A limited number of basic data sets, consistent in specifications over time and space can be regarded as a flexible data source for applications in different domains.

For example many economic and demographic statistical indicators required by decision-makers are based on basic data sets, e.g. the population number, age, education, income etc. This multiple use increases the cost-efficiency of the statistical system in general because many users are contributing to the cost of data collection, processing, diffusion etc., and so to the overall cost of the statistical system.

Regarding land cover / land use statistics the situation today is unfortunately somewhat different. A multi-purpose data set does not exist at the European level, but numerous information systems have been developed for specific data demands of specific user groups related to the description of specific features (e.g. agricultural, environment etc.). Concerning land use and land cover in Europe more than 70 are established (CROI 1999).

All of them are designed for the same purpose: to deliver land use and land cover information. But in fact their contents are so different that the information is generally not comparable with one another. The situation gets worse when trying to use land use and land cover information for other thematic fields, like environment.

The causes and consequences for this unsatisfactory situation can briefly be summarised as follows JANSEN, L.J.M.; A. DI GREGORIO (1999):

- Sectorial approach adopted to the specific regional conditions

  Many information systems follow a sectorial approach, designed for specific user requirements and adapted to the specific local conditions.

  In consequence, they are limited in the ability to define the whole range of possible land cover classes.

Due to the specific purpose a certain scale and a certain data collection system (e.g. remote sensing data) is often chosen. In some cases (e.g. land cover classes based on NOAA AVHRR Data) the derived classes might be strictly dependent on the means used for data retrieval.

- Consistency

  Most land cover/ land use information systems suffer from internal inconsistencies.

  Often the criteria used to derive classes are not systematically applied or class definitions are imprecise, ambiguous or absent. The type of diagnostic criteria and their arrangement to form a class is very often in contrast to the ability to define a clear class boundary. This is, however, a basic requirement for any system.

  The full combination of diagnostic elements describing a class is usually not considered, because the application of all combinations of the possible classifiers would lead to a vast number of classes which cannot be handled with the current methods of class description. Therefore, the current systems often create gaps in the systematic application of the diagnostic criteria used.

  The result of these inconsistencies is that the class definition between different classifications are unclear, or result in overlaps or gaps.

  Sometimes information systems are mere legends rather than a systematic classification, so that only a proportion or subset of the entire range of possible classes is described. The user cannot refer back to a classification system that leads to the impossibility of making inter-comparisons with other systems.

The above mentioned drawbacks illustrate that there is not as much compatibility between classification systems, or between classification and legend, as may be expected. However, no matter how useful the current classification systems may be for their specific tasks, the above drawbacks hamper the possibility of the use of such classification results by a large audience for a broad range of applications.

The development of multipurpose information systems should be reinforced in order to overcome this unsatisfactory situation.

The aim of a multipurpose information system on land use and land cover is manifold. Information retrieved through a multipurpose system should:

- provide comparable land use and land cover information through harmonised data
- generic land use and land cover information adaptable for specific purposes in various thematic fields, so that different users can rely on the same basic set of information

The potential benefit of such a system lays in its synergetic potential. In times of limited resources the effort to develop a unique and harmonised land cover / land use information system saves money, time and human resources.

6. DATA HANDLING AND MANIPULATION

6. DATA HANDLING AND MANIPULATION

6.1. Generalisation & Aggregation

6.1.1 Generalisation

Generalisation is "the reduction of detail or simplification of reality".

In CARTOGRAPHY generalisation refers to geometric and thematic features of maps. Maps are representations of reality and so cannot retain 'all detail', or if they could, they would be too complex to understand. The **process of generalisation** consists of simplification, selection and classification.

Simplification

All features intrinsically have three dimensions: length, width and height. As paper maps are 2 dimensional, the height (z) is lost. Geographic features can be simplified down to:

Area	2D	Polygons may be filled to differ from lines (e.g. lakes)
Line	1D	Arcs may be patterned to differ from bounds (e.g. roads)
Point	0D	Points (e.g. a landmark)

One method of simplification is COLLAPSING: As scale decreases (e.g. from 1:50.000 to 1: 500.000), features progressively lose dimensionality:

Areas -> Lines e.g. rivers, roads, at large scale presented as areas with 2 lines become one line-symbol

Areas -> Points e.g. city boundaries shrink to one point-symbol

Lines -> Points e.g. bridges

Detail of lines and areas and total number of points, lines, and areas also decrease.

Selection

Another way of generalisation is to select features according to the purpose of the map (topographic, thematic) and drop the others. How many elements remain depends, besides the theme, mostly on map scale: if map scale halves, then the total map area size is reduced by 4, and the number of elements: points, lines and areas are reduced accordingly.

Selection processes include:

- Thematic: Aggregation (merging of several elements, most common in areas)
  Elimination (removal of certain elements)
- Geometric: Smoothing (removal of details in shape or outline)

Classification

While most items have individual characteristics (no two forest stands or city sizes are identical), they must be grouped into classes for comprehension and ease of representation. There are three types of classifications or data groupings:

- Nominal: by 'type' or qualitatively e.g. forest species, oil versus gas pipelines, types of buildings;
- Ordinal: ranked in order (size implied) e.g. fire risk - high, medium, low; roads hierarchies (main, secondary, paths)
- Interval: quantitative, by size e.g. city populations, amount for rainfall; These can be expressed as totals, ratios, densities or percentages.

6.1.2. Aggregation

In *STATISTICS* generalisation or loss of detail in information is implied in aggregation. Two aspects are to be considered:

- the thematic dimension, e.g. combining several variables into one 'aggregate' or an 'indicator' (=compressing information) or by adding up for example figures on land cover of different hierarchical levels of a classification to an upper level (e.g. from 44 CORINE Land Cover classes on the 3rd level to 15 classes on 2nd level)
- the spatial dimension, e.g. adding up (or averaging) figures referenced to small restitution units to larger areas (upper levels of the regional reference system) without changing the meaning.

Census data are normally aggregated in the spatial dimension to guarantee statistical confidentiality.

There is a crucial distinction to be made in the use of maps in statistics:

- Thematic or statistical mapping

  Statistics are presented on generalised maps (digital of analogue format). Selected features, mostly boundaries of reference regions, and toponomy are sufficient for such statistical or thematic mapping. That does not mean that the statistical content of the maps is also 'generalised' meaning aggregated. Very precise statistics may be shown on very generalised and small-scale maps.

- Use of maps or geographic data as input to statistics

  Regarding the use of geographic data for preparation of statistical surveys, geometric precise and accurate information is required mostly on a large scale e.g. to identify buildings or parcels. Many more features are required by the surveyor to find the observation unit e.g. delineation of buildings, parcel boundaries from the cadastre, transport network, landmarks, waterways etc. This usage influences the quality of the resulting statistics.

6.2. Data Integration

"The process of unifying existing data sources into a single framework is called database integration." (DEVOGELE 1998).

In the field of land cover / land use information the data has a spatial aspect. Specific software tools in combination with data base management systems have been developed to handle geographic data: so-called geographic information systems. A Geographic information system (GIS) is *(GOODCHILD 1997):*

- a system for input, storage, manipulation, and output of geographic information
- a class of software
- a practical instance of a GIS combines software with hardware, data, a user, etc., to solve a problem, support a decision, help to plan

GIS enables data integration using the geographic reference of data. Information of different sources may describe the same area and can be integrated using the geographic dimension. Problems of major concern hereby are:

- understanding the meaning (semantics) of the data,
- developing correspondence between similar contents (i.e. establishment of the classification system) and
- choose adequate data processing techniques.

Data integration is to be seen in the context of the **information system**, in our case on land cover and land use. Such an information system requires traditionally the following elements:

- **Classification:** organising those groups of objects (classes) in a structure (mostly hierarchical) defining relationships of those classes to each other
- **Nomenclature:** giving names and detailed descriptions of properties of land cover and land use types to enable the identification of individuals (pieces of land, biotopes, buildings) to groups of objects (land cover, use classes).
- **Identification:** assignment of individuals to a group of objects

The implementation (creation) of such an information system maybe structured into phases:

- **Design and development phase** of the above mentioned elements as well as specifications such as data model and database structure, quality management[4], up-dating procedures, data administration and distribution policy
- **Data collection phase**: selection of appropriate methodology and techniques for data collection, application of the pre-defined rules to the objects to be classified, internal quality checks (data integration is one technique to create new information by combination of different source data)
- **Quality control**: application of data quality control procedures
- **Distribution**: Organisation of data distribution (physical distribution, user-support etc.)
- **Up-dating and maintenance**: Execution of up-dating procedures (not necessarily the same methodology as used for the first data collection phase) including quality management and control of the database.

[4] The quality management is closely linked to the project management and includes review and revision of the implementation, creation of a meta-database etc.

Various disciplines, such as agriculture, forestry, ecology, regional and town planning etc., require land cover or land use information according to sector-specific parameters, depending on the purpose. There exists nearly as numerous data sets as projects requiring land cover and land use information (in Europe more than 70 at national or international level (CROI 1999). These diverse 'information systems' contain data that are hardly comparable or compatible. This results in a low cost-efficiency ratio of the global situation.

It would be ideal to collect from scratch by means of field surveying, basic data with a very high precision and accuracy, as proposed for the German ATKIS database (Authoritative Topographic Cartographic Information System). Because this is a very costly and time-consuming exercise it has been decided by the German authorities to digitise the existing topographic maps (1:25.000 scale)[5].

Integration of already existing data sets from different sources is another way to improve cost-efficiency. The choice to apply data integration to create a new data set is much depending on the purpose and the objectives. These objectives need carefully to be analysed as well as the consequences of the data integration process on the quality and reliability of the resulting data.

6.2.1. Phases of Data Integration

The data integration process consists in general of the following phases (adapted from DEVOGELE 1998):

Compilation and preparation

This phase includes all the necessary activities aimed at the integration of different data sets:

- identifying and gathering the relevant data (including meta-data) from various sources,
- transformation of all data to fit a unique format (harmonisation of heterogeneous data models into a common one at the logical level, vector to raster conversion and vice-versa at the technical level),
- complement data descriptions to enable future users to fully understand the quality of the integrated database: e.g. explicit description of the map projection and reference system used, the transformation procedures applied etc.
- establishment of global (=database wide) and local (referred to each source) thesauri.

Investigations on correspondences

Investigations on correspondences between related categories of data and detection of conflicts consists of: identification and detailed description of all correspondences among the different data i) at the description (meta) level and ii) at the object (data-) level. Semi-automatic techniques and tools to analyse similarities exist and help to decide on correspondences based on the knowledge and understanding of the data meaning (=semantics). In addition the collaboration with experts of the domain of the source data helps to understand the data and to establish conversion tables.

Analysis of correspondence of different source data sets concludes in recommendations/guidelines concerning which data sets can be integrated, which manipulations need to be made for integration and which may not at all be integrated.

[5] The 'ideal' data collection is being done in parallel with the regular and continuous up-dating process applied for the base maps (1 : 5.000).

Harmonisation

At the thematic level, the categories or classes on land cover and land use are to be harmonised according to the results of the correspondence analysis. This is the most difficult and problematic part of data integration. At the technical level, geographic data is to be harmonised enabling synoptic analysis (e.g. overlaying): matching points and lines (polygon boundaries) from different sources related to similar objects e.g. river network of the topographic map with the one from a satellite image classification, boundaries of an arable land area from a land use data layer with those boundaries stemming from a forest area data layer etc. This is one of the most time-consuming steps in data integration.

Incorporation

Incorporation means in fact: Make one data set out of two (or three …).

For example centre **points** of buildings from a register with their attributes (=point data), maybe incorporated with building boundaries from a large-scale data set e.g. a 1:5.000 scale city plan (=line data) to create a Built-up **area** data layer (polygon data). To complement this new data layer another digital data set on wooded land at a specific scale with a specific definition, precision and accuracy - in terms of thematic reliability - and collected with a specific method might be incorporated.

Generalisation

Data integration obviously implies geometric and thematic generalisation to enable meaningful incorporation of data with different specifications and scales.

Generalisation procedures may also help to prepare the data for different applications. Figure 6.1 summarises the process.

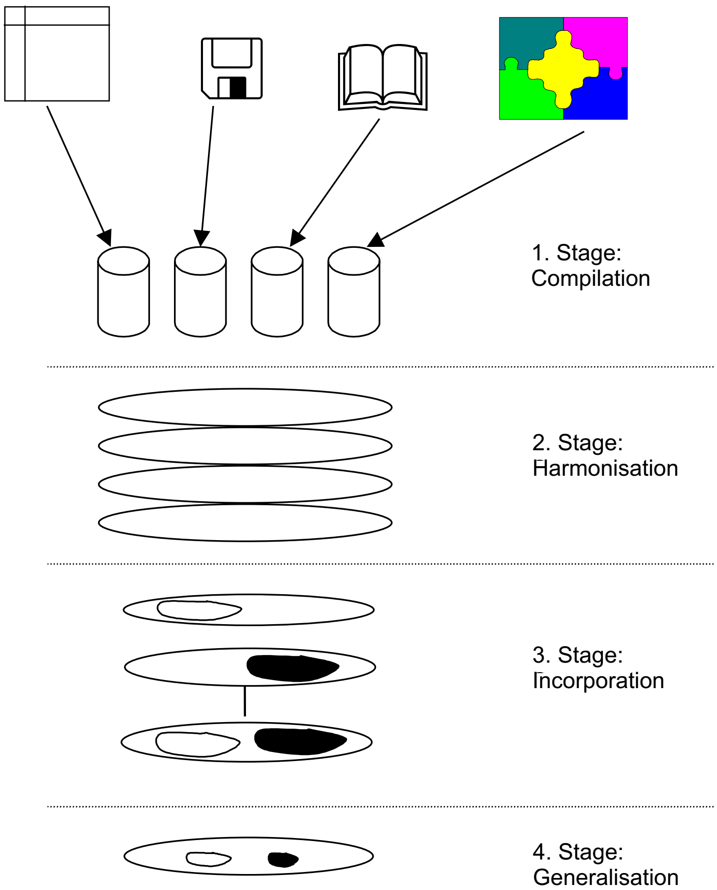


Figure 6.1: Scheme of data integration

6.2.2. Problems of Data Integration

As already mentioned, several types of problems need to be addressed in this context regarding technical issues of integration process, the concepts and the observation techniques applied to create the source data sets, and the data quality of the new data set.

In the following some of the problems are briefly discussed.

Conceptual Problems

Incompatible system parameters of source data

The design of the basic elements of the source data, the choice of the data collection approach, the parameters and the applied techniques shall be ruled by the specific requirements and the purpose.

The information gathered in the survey of existing European systems showed on the contrary, that a great number of systems have been designed to match with the potential of the data source and the data collection tools used to create data sets, especially where earth observation is used as data source (Land Cover Map of Great Britain, CORINE Land Cover, Forest Map of Europe using NOAA-AVHHR data etc.).

Classification systems

Differences in the class definitions in existing systems do mostly not allow direct comparability or even compatibility between different source data sets. The detailed definitions of land cover and land use classes need to be analysed to enable eventual reclassification of data. Great attention has to be given to this correspondence analysis regarding the targeted purpose of the new data set. A multi-disciplinary team should carry out the analysis of correspondence with expertise in the respective domains of the source data (e.g. agriculture statisticians, foresters, ecologists or town planners). The analysis concludes in rules for the re-classification and lead to conversion tables showing the link of the nomenclatures.

Several approaches to build bridges or interfaces between nomenclatures exist: ITE approach: creation of a baseline nomenclature; Eurostat approach of Kernels and Margins; FAO approach: defining classifiers for land cover and for land use.

Different data collection approaches: Mapping meets Statistics

Statistical sampling and cartographic mapping are conceptually different approaches to collect data on land cover or land use.

Mapping - in the true sense - has the advantage to cover exhaustively all the territory of interest. The geometric scale is of high importance on the potential applications of the data: large scale data (1:1.000 - 1:10.000) are highly precise in terms of geometry, and maybe also in the thematic detail what is required for ownership management and local / town planning; smaller scale data (from 1:50.000) contain generalised information in terms of thematic and spatial precision what is required for regional planning and environmental monitoring. Disadvantages of the exhaustive mapping approach are the high cost and the long time required for implementation covering large areas.

Sampling methods are not so costly and deliver, depending on the survey method, detailed information on land cover or land use in a short time period. Disadvantage of sampling data is a coarse spatial resolution in terms of restitution unit - one should not talk about scale in this context - depending on the sampling methodology and eventual a stratification - meaning that information is representative for an 'artificial' regional reference area and not for a precisely defined georeferenced location as with the mapping approach.

The consequences of integration of those different types of data need to be carefully analysed and taken into account in applications. For specific purposes, the integration of these different types of data can be extremely useful (e.g. the Finnish Multi-Source Forest Inventory, TOMPPO 1992).

Problems related to observation

Different data capture and processing techniques

The existing source data are in general captured with different techniques depending on the approach. In statistical surveys generally data is captured by field investigations and interviews using questionnaires or measurements of variables in the field. The results of statistical surveys are strongly dependant on the sampling strategy and design, as well as on the skills of the surveyor. Ground survey may result in very precise information on a limited number of samples (points, segments).

Mapping on large scales requires field surveying resulting in very precise and accurate data whereas for large areas mostly earth observation data such as aerial photographs or satellite imagery are visually interpreted or somehow digitally processed to follow-up or create data sets.

Results of photo-interpretation of land cover and land use are highly influenced by the skills and experience of the interpreter and the auxiliary information available. Such data is not recommended to derive area statistics from it because of bias introduced by systematic non-sampling errors (no compensation of omission or commission errors) and the minimum size of interpreted polygons (GALLEGO & CARFAGNA 1998). Parameters applied in automatic or semi-automatic processing are documented and do so allow a repetition of the process, but the setting of parameters requires the expertise of an image-analyst, which is again subjective.

Integration of data derived through different collection methods has a strong impact on the reliability of the resulting data set. If such integration is performed the data processing need to be precisely documented so that the final user may assess the usefulness of the integrated data set for his purpose.

Different scales and thematic detail of data

Maps show a simplified model of the reality on the area concerned. Objects are classified and symbols used to represent them on paper. Large-scale maps show more objects and symbols and more detail in terms of attributes with a relatively high geometric precision and accuracy. Small-scale maps present only generalised information and are in principle not very rich in terms of content and nomenclature.

In a digital geographic information system, the limitation of the paper size does not exist anymore. In an ideal digital GIS the data stored should be collected by field survey with a very high geometric precision and accuracy, so to say scale-less. Because of the cost and the long time required to cover large areas, today data is digitised from existing maps at various scales so the geometric precision / accuracy of the data depends on the original scale of the source data and the data capture method.

The integration of small-scale geographic data - meaning “low quality” data in terms of thematic richness (aggregated land cover or use classes) and geometric precision - into a data set increases the problem of classification systems (in terms of correspondence of classes) and does not contribute very much towards multi-purpose information. The higher aggregated the data is (= poor in detail) the less useful it is for further processing for multi-purpose applications. The smallest scale amongst the source data sets determines the geometric precision and accuracy of the resulting geographic data set.

For example, satellite imagery has a geometric resolution, corresponding to the pixel size, which depends on the sensor type, the orbit parameters of the satellite platform and the ground (altitude, relief energy). This data may be presented at different scales but the geometric resolution of e.g. of 30 m * 30 m for Landsat TM images, and the content do not change. The land cover information derived from such satellite data inherits this geometry. When integrating small objects from other data sources, e.g. geographic point data from a register, location / positional accuracy is very important.

The generalisation process can be seen as a normalisation of both: thematic and geometric content is normalised, what allows to integrate input data of different scales without too much geometric distortion of the output data as long as the scale of output data is smaller than the smallest scale of the input data.

Different time periods of data collection

The use of various source data collected not at the same date (or a short period of time) leads to time inconsistencies in the data set. Data on one region of the area observed were collected e.g. in 1988 and in other regions in 1994. Change detection may only be carried out if the date for each area unit is known. For example the Swiss Area Statistics are based on the photo-interpretation of points (based on a 1ha grid) in aerial photographs covering all Switzerland in a period of 1979/85 and 1992/97 (SWISS FEDERAL STATISTICAL OFFICE 1996).

Problems related to quality and accuracy

The quality assessment of data integration is a difficult issue. To allow the user to make an assessment of the "fitness for purpose" (= quality) of a data set, the most important information required is concerned with the procedures of creation of the data, element of the so-called 'meta-information'. Quality assurance must include analysis on error propagation regarding the errors contained in source data and introduced through the integration process and generalisation. Accuracy assessment needs to be applied in terms of geometry accuracy and thematic correctness.

Meta-Information and Data Genealogy

Each data set should be accompanied with detailed information on the approach, the date and the method of data collection, the data processing methods applied, accuracy assessments in terms of spatial and thematic aspects carried out by the data provider, technical formats, and so on (HUNTING 1997). This meta information is crucial for data integration. Only recently the development of meta data standards has been addressed by different institutions as the Comité Européen de Normalisation CEN with its CEN/TC 287 metadata standard, the International Standards Organisation ISO with its ISO/TC 211 or by MEGRIN with its Geographical Data Description Directory GDDD.

Thematic accuracy assessment

The assessment of the thematic accuracy of a data set is a complex but necessary part of its creation. In integrated data sets this assessment is much more difficult than in single source data because of the different error sources and the data processing itself. Concerning land cover and land use statistics, the correctness of the data is checked for example by double survey of a sampling of the interviews enabling to calculate the accuracy in a defined confidence interval, or as in the case of the Swiss area statistics by double interpretation of each point by different experts and field surveys. Regarding thematic accuracy of land cover / land use data derived using satellite imagery in general it is assessed using a 'ground truth' information on area samples derived from ground survey or other sources.

Concerning error propagation while integrating different types of data, a thorough analysis on potential error sources in spatial data bases is to be carried out: calculation errors, specification errors, sampling errors, measurement errors or stochastic errors.

Geometric accuracy assessment

The geometric precision and accuracy depends on the original data collection method and the scale of the source data. The assessment of geometric accuracy needs to take into account the data processing: digitising or scanning, resampling, geometric rectification to different geographic co-ordinates and projection systems, generalisation procedures etc.).

A sound methodology for the geometric accuracy assessment includes for example checkpoints using ground survey measurements, topographic maps or GPS (Global Positioning System, US satellite navigation system). Even already georeferenced data should at least be visually checked. If possible, a georeferencing protocol should be obtained indicating reference map, scale, date and projection system. Positional errors are best to be detected on linear features (roads, rivers) not using polygons because they are comprised by different objects boundaries which depend on nomenclature.

Errors introduced by generalisation

Aim of generalisation is to simplify, so to reduce the amount of spatial and thematic information (JAAKKOLA 1998). This implies inevitably the introduction of errors in geographic data sets. Statistical comparison of the input – non-generalised – data with the generalised output data to demonstrate the grade of generalisation is not enough.

The assessment of generalisation errors should include the analysis of the flows of 'area' during the generalisation process from one cover type to the other (similar to: analysis of flow of votes from one party to another in an election) and the loss of features like islands or lakes.

Technical problems

Different environments of the source data

The transformation and re-transformation of all data to e.g. the raster format does not yield the same geometric and thematic precision if the aspect of geometric scale is not carefully taken. Similarly the vectorisation of raster data does not yield exactly the same representation than the source data, if parameters are not carefully checked.

Resampling to the same pixel size

Resampling of data to the same grid size in the raster format bears the danger of mixing large and small-scale data. It eases data processing but the “real” precision and detail of data needs to be kept in mind (and documented in the meta-data) when data are used. For example the resampling of satellite imagery data of an original pixel size of e.g. 30m x 30m (Landsat TM imagery) to 25m x 25m does not increase the precision and the accuracy.

Digitisation of analogue data

There are several methods to put analogue paper maps in digital format:

- Scanning: Maps are scanned with specific parameters (e.g. 300 dpi (dots per inch) scanning resolution, 30-bit colour resolution). Scanned maps in raster format need further processing to be integrated in GIS databases: georeferencing, extraction of linear or spatial features, eventually vectorisation etc.
- Digitisation on a digitising tablet: Geographic objects are digitised and georeferenced from maps with specific parameters (e.g. 2540 lpi (lines per inch) resolution, 0.15 mm accuracy etc.). Depending on the digitising software, attributes may directly be added.
- Digitisation on screen: Similar technique as with the tablet but with a better accuracy.

Regarding the Minimum Legible Delineation of an analogue source map of a certain scale, the parameters of the digitising or scanning technique applied need to be taken into account in data integration process.

Rectification of data into the same geographic projection

Separate procedures are used for rectification (vector data rectified in vector format, raster data in raster format). Vector data have infinite precision at a point (or line) whereas raster data covers a certain area on the ground. There are hidden data model transformation problems in the data integration process.

7. QUALITY

7. QUALITY

In general the quality of a product means that the product must fit the purpose. The ISO norm 8402 on 'Quality management and quality assurance - Vocabulary' defines quality as "totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs".

In terms of land cover and land use information this implies that the information is:

- relevant and useful for the application
- reliable regarding the aspects of its thematic content and geometric accuracy, particularly in a mapping approach

The first aspect requires that the data is adequately described and the user has access to complete metadata about the dataset. The second aspect is based on accuracy measurements regarding both, thematic and geometric accuracy.

7.1. Metadata

Metadata is information about data. It is the detailed description of data specifications. Such 'information about information' enables the user to assess the usefulness of the data for his specific application.

The following information shall be contained in a metadata set for land cover and land use data:

- Data set identification (title, responsible organisation) scale, restitution unit
- Data set overview (general description of the context, potential applications, limits etc.)
- Data set specifications (coverage, nomenclature, classification principles, data collection approach and techniques, spatial reference, restitution unit or mapping unit, scale, source data, periodicity)
- Data quality parameters (quality assurance procedures and control measures)
- Distribution policy (pricing, copyright)
- Metadata reference (source of metadata)

There are initiatives to develop a Metadata Standard on Geographic Information by the “Comité Européen de Normalisation” (CEN TC 287). The "Multipurpose European Ground Related Information Network" MEGRIN has developed a 'Geographical Data Description Directory (GDDD) Data Model' and implemented a database accesible on the Internet (http://www.megrin.org/ GDDD/Overview.html). Eurostat has implemented a Data dictionary for its geographic information system GISCO.

7.2. Accuracy assessment

To enable quality assessment of data quality, measures need to be performed. These tests result in quantitative statements on the geometric and the thematic accuracy. An established method for quality assessment is the validation of the data on a sample using information from another source.

A method for assessment of positional accuracy of georeferenced data is as follows: control points are selected in the map (or image) data and their co-ordinates are compared to those of the same points determined from another source of information (=ground control points), e.g. a topographic map or GPS measurement in the field. The positional deviation of the image control points in reference to the ground control points is then calculated and given as the Root Mean Square error (RMS). The following table shows as an example the RMS of the rectification of satellite data with 10m x 10m pixelsize at a scale of 1:50.000 (Eurostat Pilot project land use mapping of the Rhône valley)

	X Map	Y Map	X Image	Y Image	Deviation X	Deviation Y
FRANCE						
1	953675	3175750	959,00	12510,50	- 1,17	- 0,2
2	966050	3224475	2182,00	7633,50	0,58	- 0,08
3	962250	3159000	1821,25	14184,12	2,31	2,06
4	997225	3167525	5329,12	13329,62	- 0,02	- 0,03
5	970250	3162875	2625,12	13797,38	- 0,35	0,26
6	972900	3148625	2895,75	15224,50	- 0,26	- 0,64
7	1024675	3212325	8064,75	8841,62	0,88	1,23
8	973100	3199925	2897,50	10090,25	0,45	- 0,95
…	…	…	…	…	…	…
…	…	…	…	…	…	…
32	1415300	4906975	10239,50	6197,00	- 0,89	1,10
33	1387725	4849825	7969,50	12133,50	0,63	- 0,37
Root Mean Square error					**0,90**	**0,96**

Table 7.1: Example for the geometric quality assessment: root mean square error of different ground control points (GÉOIMAGE 1995)

Thematic accuracy may also be assessed by selecting test points (or Pixels) and comparing the land cover or land use attribute with the same information from a field survey, or by photo-interpretation of other suitable source material (e.g. aerial photographs, existing maps). The result is a confusion matrix where the commission and omission errors and the producer and the user accuracy for each class are given. The following table shows a fictive example.

	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8	Type 9	Type 10	Type 11	Total	Com-mission errors	User accuracy %
Type 1	53				1		1					55	2	96.3
Type 2		30			1					1		32	2	93.7
Type 3		2	12	1								15	3	80
Type 4				10								10	0	100
Type 5					54	1						55	1	98.1
Type 6					1	27	1					29	2	93.1
Type 7						4	46					50	4	92
Type 8		2						16		1		19	3	84.2
Type 9		1				2	2		15			20	5	75
Type 10							2			35		37	2	94.6
Type 11	2				1					1	20	24	4	93.3
Total	55	35	12	11	58	34	52	16	15	38	20	346		
Omission errors	2	5	0	1	4	7	6	0	0	3	0			
Producer accuracy %	96.3	85.7	100	90.9	93.1	79.4	88.4	100	100	92.1	100			

Table 7.2: Confusion matrix: Land cover types (fictive example)

The producer accuracy as a percentage relates the number of points **correctly classified** in the image or map to the **total number of ground control point** of that land cover type. The user accuracy as a percentage relates the number of points correctly classified to the **total number of points classified** for a certain land cover type and takes into account also the wrongly classified points in the map and allows a better assessment of the quality of the classification.

Global accuracy may be assessed by averaging user and producer accuracy. Another global accuracy measure is the Kappa coefficient that is "a statistical measure of the agreement, beyond chance, between two maps (e.g. output map of classification and ground-truthed map)" [Canadian Centre for Remote Sensing: Glossary].

8. TECHNICAL DEFINITIONS – SOME BASICS

8. TECHNICAL DEFINITIONS – SOME BASICS

8.1. Maps

"A map is a 2-dimensional graphic representation of the physical features (natural, artificial, or both) of a part or the whole of the Earth's surface, by means of signs and symbols or photographic imagery, at an established scale, on a specified projection, and with the means of orientation indicated." (DRURY 1998).

The definitions of (geographic) maps almost always includes the term representation:

1.a: a representation on a flat surface of the whole or a part of an area

1.b: a representation of the celestial sphere or a part of it

2: something that represents with clarity suggestive of a map...

8.1.1. Categorisation of maps

Maps can be categorised according to different aspects (HAKE 1982):

- Content
- Origin
- Process of development
- Scale

8.1.2. Categorisation of maps regarding contents

The most common criterion to categorise maps is the content.

Topographic maps describe the precise and accurate geometric location (= situation) of geographic objects e.g. settlements, infrastructure (road-, rail and waterways, electricity lines etc.), water bodies, limited aspects of land cover and use (wooded areas, vineyards, grassland, orchards), certain administrative boundaries, toponomy etc. Topographic Maps are commonly used as a basis for thematic maps.

Thematic maps present information on specific subject matters on the basis of topographic maps used as the geographic framework, for example Land Cover maps (based on topographic maps), Land use master plans (based on cadastre), statistical maps (based on the administrative boundaries) e.g. population density etc.

8.1.3. Categorisation of maps regarding origin

Distinction is made between maps produced by the **official institution** responsible for surveying in a country ("official" maps from the 'Ordinance Survey' in UK, the 'Vermessungsämter' in Germany, 'Institut Geographique National' (IGN) in France or Spain, the Military Geographic Institute in Italy etc.), e.g. topographic maps, and maps produced by **private organisations** or companies ("private" maps) e.g. tourist maps. Official maps serve as references and have a certain legal character e.g. for the delineation of the national territory, as basis of land use planning maps, delineation of nature conservation sites etc.

8.1.4. Categorisation of maps regarding the process of development

Separation is made into "Base-maps" and "derived maps". Base maps are usually the large scale maps produced by (or under contract for) the official surveying administrations by **field campaigns**, e.g. cadastre maps, Topographic Base Maps, while derived maps are **generalised** according to pre-defined criteria to smaller scales from the large scale base maps e.g. Topographic maps of medium scales, Tourist maps for towns. Ideally base maps are cadastre maps on a very large scale e.g. 1:1.000 from which town planning maps are derived, or topographic maps on a large scale (1:5.000) from which other smaller scale maps are derived. Topographic base maps on medium scales (1:10.000 - 1:300.000) may be elaborated based on aerial photographs and using photogrammetry and photo-interpretation as processing technique.

Generalisation is "the reduction of detail or simplification of reality". It is a cartographic processing technique to reduce the amount of information presented on a paper map according to the physical space available on the media paper in agreement with the scale chosen.

The same concept is applied to thematic maps. For example mapping of a population can be carried out using a topographic base map or a topographic derived map as basis for the mapping. A generalised vegetation map is a **derived** map (see examples Fig. 8.6 *(pop-map Hanover)*, Fig. 8.8 *(NUTS level pop-map)* and Fig. 8.7 *(Vegetation map of Africa)*).

8.1.5. Categorisation of maps regarding scale

Maps can be grouped according to their scale.

General scale categories* are:

Large scales = > 1:10.000

Medium scales = 1:10.000 to 1:300.000

Small scales = < 1:300.000

* There are in fact no fixed limits for this grouping.

8.1.6. Examples

Figure 8.1: Topographic Map at scale 1:200.000 Sheet CC 6302 Trier, D (HAKE 1982)

Content	*Origin*	*Creation Process*	*Scale*
Topographic Map	Official map (Landesvermessungsamt Rheinland-Pfalz, D)	Derived map	Medium Scale map 1 : 200.000

Figure 8.2: Topographic Map at scale 1:50.000 Sheet 6108 Bernkastel-Kues, D (HAKE 1982)

Content	*Origin*	*Creation Process*	*Scale*
Topographic Map	Official map (Landesvermessungsamt Rheinland-Pfalz, D)	Derived map	Medium Scale map 1:50.000

Figure 8.3: Topographic Map at scale 1:25.000 Sheet 6008 Bernkastel-Kues, D (HAKE 1982)

Content	*Origin*	*Creation Process*	*Scale*
Topographic Map	Official map (Landesvermessungsamt Rheinland-Pfalz, D)	Derived map	Medium Scale map 1:25.000

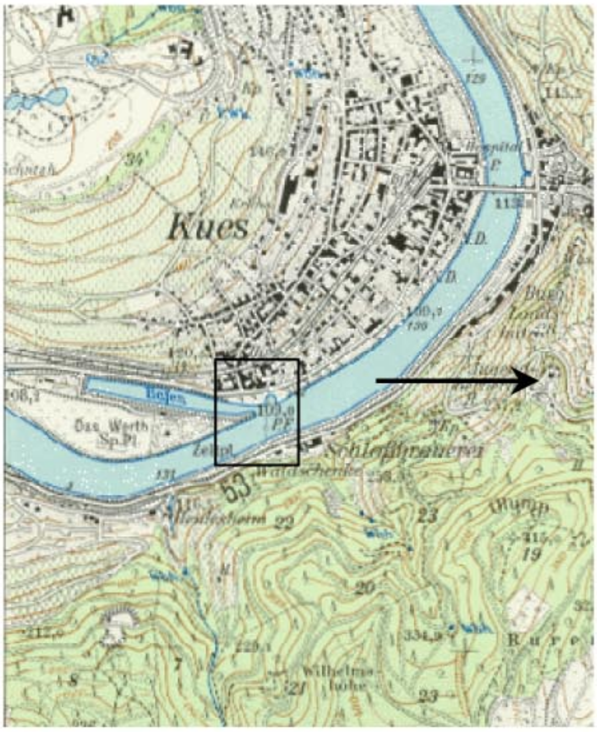


Figure 8.4: Cut-out of the German Topographic Base Map 1:5.000 Bernkastel-Kues, D (HAKE 1982)

Content	*Origin*	*Creation Process*	*Scale*
Topographic Map	Official map (Landesvermessungsamt Rheinland-Pfalz, D)	Base map	Large Scale map 1:5.000

Figure 8.5: Thematic Map - Land Use Plan at scale 1:1.000 Old People's Home, Hamburg D (HAKE 1982)

Content	*Origin*	*Creation Process*	*Scale*
Thematic Map (Planned Land Use)	Official map (Administration for Construction Hamburg and Surveying Administration Hamburg, D)	Base map	Large Scale map 1:1.000

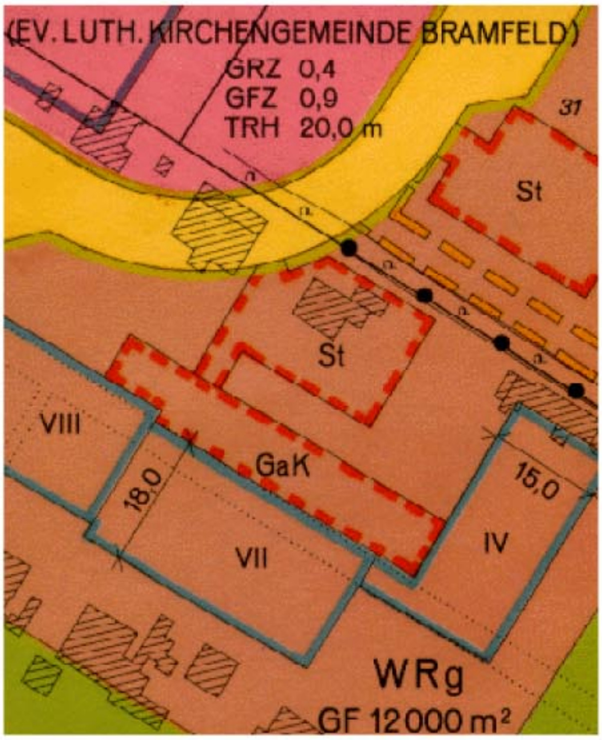


Figure 8.6: Thematic Map - Population Map at scale 1:5.000 Hanover, D (HAKE 1982)

Content	*Origin*	*Creation Process*	*Scale*
Thematic Map (Population)	Official map (Statistical Office Hanover and Surveying Administration of Hanover)	Base map	Large Scale map 1:5.000

Figure 8.7: Thematic Map - Vegetation and Climate at scale 1:60.000.000 West Africa (HAKE 1982)

Content	*Origin*	*Creation Process*	*Scale*
Thematic Map (Vegetation and Climate)	Private map (Westermann Verlag)	Derived map	Small Scale map 1:60.000.000

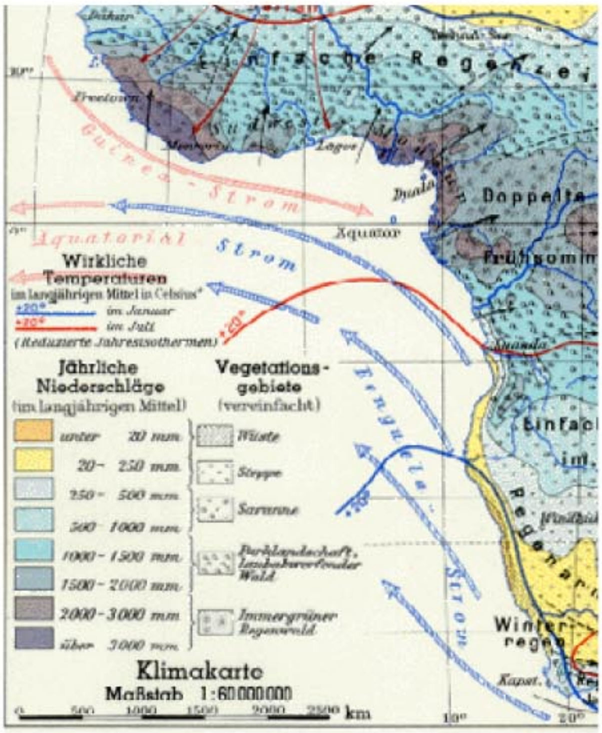
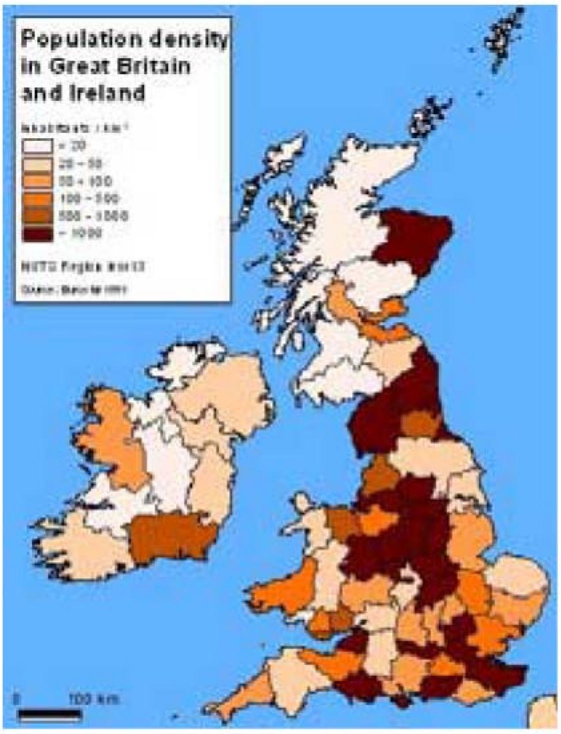


Figure 8.8: Thematic Map - Population Density in great Britain and Ireland 1991, at scale approx. 1:25.000.000

Content	*Origin*	*Creation Process*	*Scale*
Thematic Map (Population Density)	Private map (CESD Communautaire, data source: CEC-Eurostat GISCO)	Derived map	Small Scale map 1:10.000.000



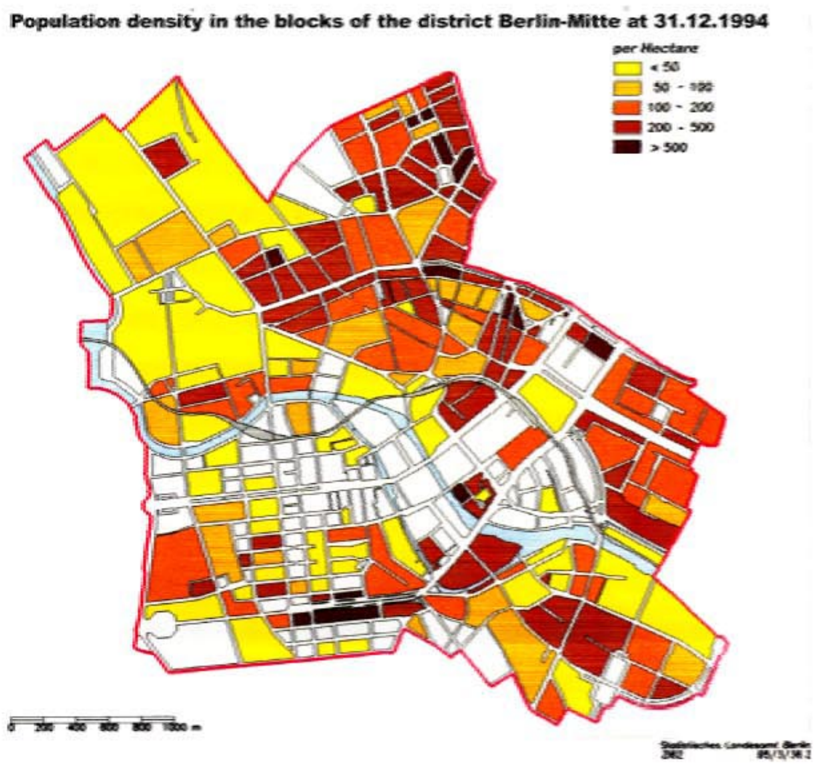
8.2. Maps in statistics

Regarding map use in statistics, a distinction should be made between i) mapping of statistics for presentation purposes (= thematic mapping) and ii) the use of analogue maps and/or digital geographic data as input to the statistical data collection process e.g. as information source, design of stratification, sampling or the preparation of census.

8.2.1. Mapping of statistics

Thematic mapping is the 'simple' use of maps to present statistics or results of statistical analysis. Depending on the subject, the geometric accuracy and the appearance of geographic features is of minor importance. The required scale of the map depends on the geographic extent and the Regional Reference System of the statistics to be presented. If the geographic extent is the territory of the European Union and the regional reference system is the NUTS (Nomenclature des Unités Territioire Statistique) at level II the required scale to present statistics (e.g. population density) is different from the scale to present statistics on an urban agglomeration, e.g. Berlin, with the regional reference system of statistical blocks (enumeration districts). In the figure 8.8 above the required geographic data are the boundaries of the NUTS regions including the coastlines at very small scale, e.g. 1:30.000.000. For a map with the same content - statistics on population density - on Berlin (see Figure 8.9) with the regional reference system of the statistical blocks, the scale is about 1:5.000 and the required geographic data are the boundaries of the statistical blocks.

Figure 8.9: Population density in the Berlin district 'Berlin-Mitte' at statistical block level (source Statistisches Landesamt Berlin)



8.2.2. Maps or geographic data as input to statistics

The use of maps or digital geographic data as a basic input information for statistical work is more complex. In such applications, the geographic information is part of the resulting statistics and influences very much the result. The map or data to be used must correspond in terms of thematic content and scale (that includes precision and accuracy) to the geographic aspect of the required statistics.

The calculation of the total land area for example depends first on the statistical concept of the land area that may include water bodies, surfaces under sea level, remote and inaccessible areas etc. The precision and accuracy of geographic data is the second important issue concerning calculation of areas from digital geographic databases. If the geographic data was acquired by field surveying, the accuracy and the precision are normally sufficient to measure areas. Field surveying is normally carried out for mapping at large scales. If existing maps at smaller scales have been digitised, the scale of the source map should not be smaller than 1:10.000 for area measurements.

Another example for map use as basic data is in census preparation, where the statistical blocks and some data on infrastructure (road network with street names and house numbering etc.) at the adequate large scale is required. If the map is not up-to-date the survey will be incomplete. To set-up an area frame sampling or stratification for example in the design of an agricultural survey, the basic geographic data is crucial because the whole statistical work is based on the quality of the underlying map.

Accuracy of basic geographic data is also crucial for spatial statistics, not only in terms of spatial accuracy but also in terms of accuracy in time: the geographic boundaries of statistical units should correspond to the date of the statistics. For example a re-structuring of administrative units may lead to changes in boundaries (2 units getting one or vice-versa, position of boundaries change etc.).

If the statistical concept is defined, the figure for the area can be calculated in different ways: using the administrative data from the ownership registers or the cadastre maps or by the surveying agencies using large-scale base maps (1:1.000). The statistical offices normally rely on the data from the administration or surveying agencies. The following figure illustrates the problem related to the statistical concept of land area in Finland.

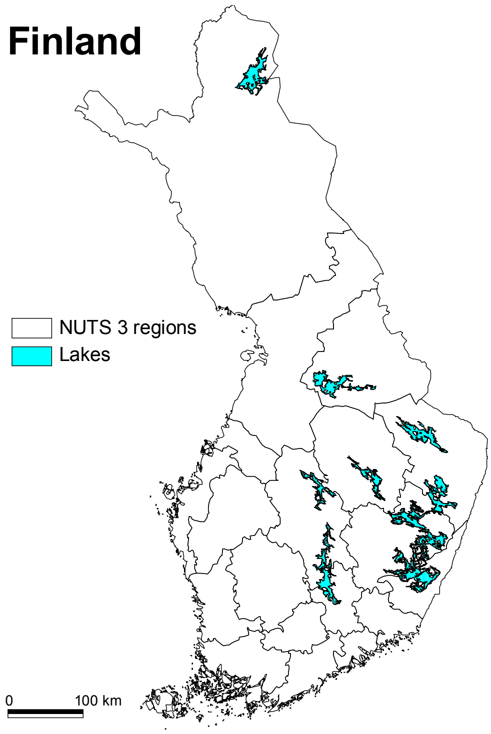


Figure 8.10: Finland's Lakes[6]

The water area is counted up to 33.560 km^2, app. 9.9% of the total area according to FAO land use statistics.

A more demanding example is the use of maps in preparation of a statistical survey, e.g. a population census. Therefore, precise (=detailed) data in terms of geometry (buildings must be distinguishable, so minimum scale 1:5.000, see figure 8.6 *Hanover pop-map*) is required. Detailed knowledge of the use of buildings would be very helpful to exclude e.g. warehouses or industrial zones. Regarding the design of stratification for e.g. crop yield estimation, parcel boundaries with a high geometric precision and accuracy are required to allow area estimates.

Spatial analysis needs in most cases the information on neighbourhood, so-called topology. For those purposes, the requirement for geographic data is much more demanding in terms of detail and accuracy so the quality aspect plays a crucial role in selection of source data.

Meta-data, information about the geographic data itself, is necessary to fully understand the data and enable meaningful application for specific analysis and data integration.

[6] The source data for this map is at scale 1:1.000.000. Many of the small lakes disappear at that scale because of the generalisation process. Calculating the water area from this basic geographic data, it counts only to some 3% of the total area.

8.3. Geodetic Reference System and Regional Reference System

8.3.1. Geodetic Reference System

The geodetic reference system is defined as 'a complete reference system for positioning a point on the earth, including datum, co-ordinate description, co-ordinate system and possibly a projection' (CEN 1998). Any type of map is based on such a positional reference system: each point on the Earth's surface corresponds to a point referenced on the map. In other words, it allows an exhaustive description of the earth's surface.

The crucial problem of this mapping exercise is the representation of three dimensional surface of the body of the Earth onto a two dimensional plane (the paper map). Only a three-dimensional model of the Earth - a globe - can avoid distortions of distances, areas and shape of geographic objects simultaneously. Two-dimensional maps are always a compromise according to the purpose of the map.

Geodetic Datum

To enable the representation of the Earth's surface on a piece of paper, parameters of the size and the shape of the Earth and the origin and orientation of a co-ordinate system need to be defined. Such parameters are described by so-called Geodetic datums.

These descriptions of size and shape of the Earth may be simple two-dimensional systems just taking into account the plane area. Such flat systems are sufficient for large scale maps for distance measuring up to 10 km avoiding problems with the earth curvature. The other extreme - describing the Earth as a sphere - is described by spherical models with a specific radius for global distance approximations at very small scale. For accurate measurements e.g. for navigation of ships or aeroplanes more complex systems taking into account the true ellipsoidal form of the Earth (poles are flattened), the topography meaning the relief and the gravity have been defined. Figure 8.11 shows a simple model of the Earth size.

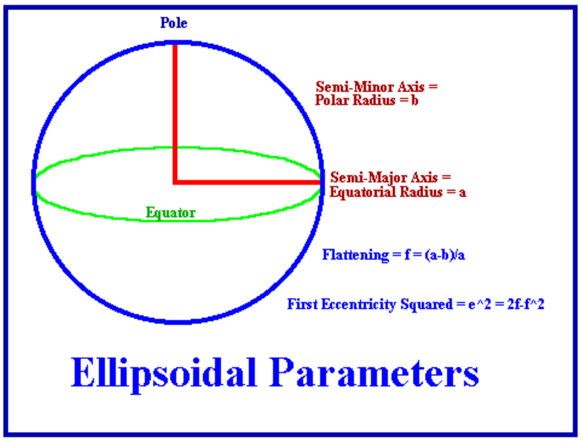


Figure 8.11: The Geodetic Datum describes the size and the shape of the earth. Ellipsoidal parameters (adapted from DANA 1999)

The topographical surface is very irregular and changing over time: Long-term geological movements of the tectonic plates and mid- to short-term geomorphologic movements change the relief. Gravity differences and tidal forces influence the sea level daily from location to location by hundreds of meters. A type of average surface of the Earth - the Geoid surface - has been developed that takes into account the above mentioned factors.

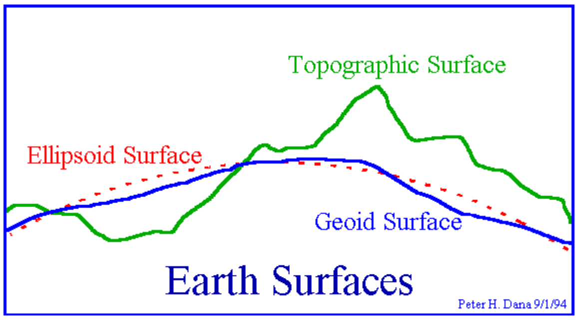


Figure 8.12: The ellipsoid, topographic and the geoid surfaces (DANA 1999)

As was already said, different datums are defined for various purposes and in different countries. Many different datums are used by different countries and agencies throughout the world according to the best fitting ellipsoid for the purpose and the country.

Using a wrong datum when referring geodetic coordinates in a coordinate system - meaning not the one that is the basis of the map - can result in position errors of hundreds of meters. Trying to match maps based on different datums results in distortions, gaps or overlaps in the spatial coverage.

The technological progress achieved in global positioning measurements with sub-meter accuracies requires careful datum selection and careful conversion between coordinates in different datums.

A unified world geodetic system is essential for intercontinental accurate geodetic information. In the 1950s, development of such a system started in the US and resulted in the World Geodetic System 1960 (WGS1960) developed by the Department of Defense (DoD). A major improvement has been achieved with the WGS 72 datum resulting from a extensive effort over 3 years to collect selected satellite, surface gravity, and astrogeodetic data available throughout 1972. The WGS 84 datum replaced the WGS 72 using new and more accurate instrumentation and a more comprehensive control network of ground stations. The WGS 84 is geocentric, the centre of mass being defined for the whole Earth including oceans and atmosphere. Its scale is that of the local Earth frame, in the meaning of a relativistic theory of gravity, its orientation was initially given by the Bureau International de 'Heure (BIH) orientation of 1984.0 and its time evolution in orientation will create no residual global rotation with regards to the crust (McCARTHY, D, 1996).

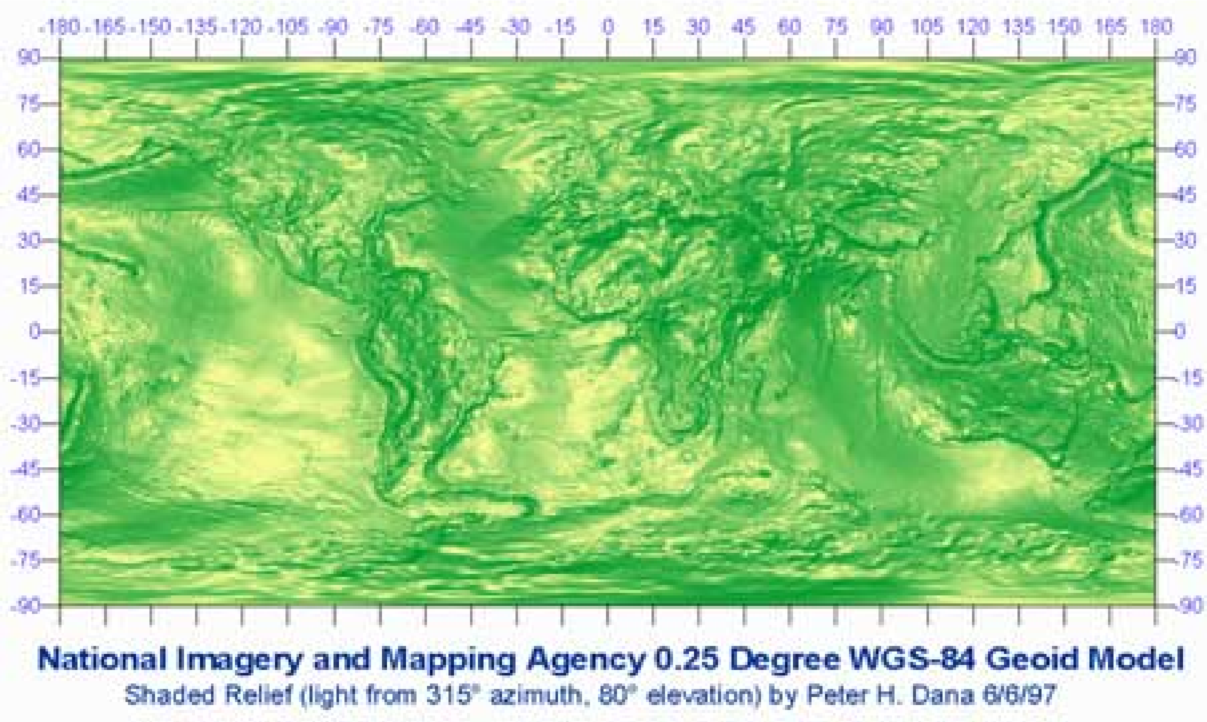


Figure 8.13 Shaded relief of WGS-84 Geoid model (DANA 1999)

Today, the 'World Geodetic System 1984' is the basis for the 'Global Positional System', the satellite navigation system developed and controlled by the US Department of Defense. A complete description of WGS84 is given in National Imagery and Mapping Agency (2000): Technical Report 8350.2 3rd edition: World Geodetic System 1984.3.

Map projections and co-ordinate systems

With the geodetic datum, the 'anchor points' are given for the absolute positioning of locations on a map.

The next step is to define a system to identify a point on Earth on a piece of paper using coordinates. Coordinate systems maybe based on different **map projections**. Points on the Earth's surface (or the parallels and meridians) are projected onto a plane. One should imagine a sphere painted with parallels and meridians and a light source illuminating it. The source of illumination maybe (as illustrated in figure 8.14) internal the sphere (centre projection) or on a point on the sphere (stereographic) or maybe very far from it - like the sun from the Earth- sending parallel radiation (orthographic projection).

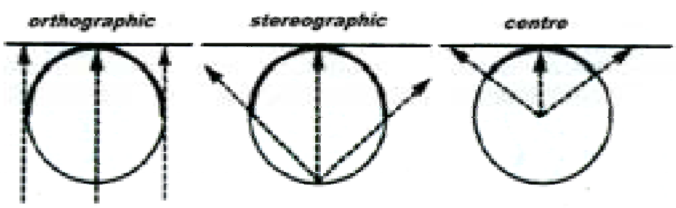


Figure 8.14 Orthographic, stereographic and centre projection (adapted from BRUNET 1987)

The parallels and meridians are being projected onto a plane. This plane - our paper map - maybe taken as is - planar - and the projection is called azimuthal. The plane may touch the sphere (tangent) on one point e.g. the North pole, or cut through it (secant). Taking the paper and rolling it around the globe results in a cylindrical or a conical projection. There are many possibilities as illustrated in Fig. 8.15.

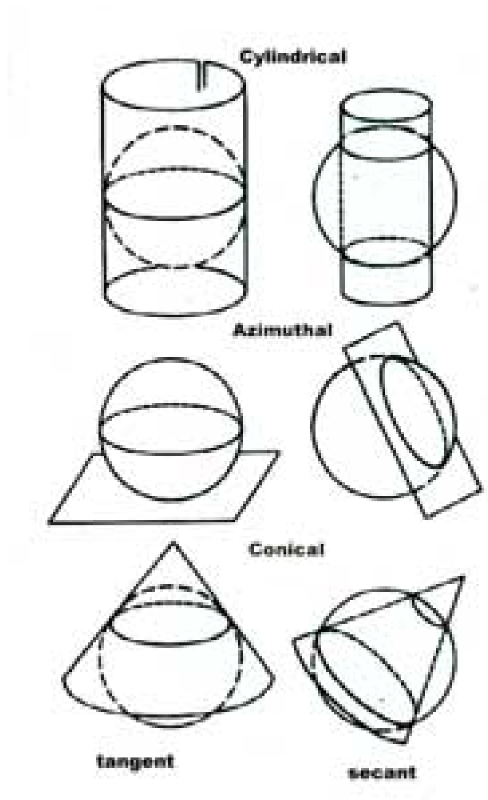


Figure 8.15: Principles of projection (BRUNET 1987)

The projected grid system of the parallels and meridians will be geometrically distorted. Depending on the type of projection and its parameters a straight line, for example a road, will not be straight on the paper plane and its length will not be correct; a circle may be an ellipsoid and the shape of a square on Earth may be rectangle on the plane, and so on. Again depending on the projection system and its parameters applied, one of the following conditions maybe fulfilled:

<u>Equidistance</u> meaning true distance measuring over the entire map. This characteristic is required for navigation purposes.

<u>Equivalence</u> that means true area calculations are possible over the entire map. This property is important for most statistical mapping for comparing data of spatial statistical units.

<u>Conformality</u> means that shapes (so angles) of geographical areas are preserved.

In general the selection of a projection system for a map dedicated to a specific purpose is a compromise regarding the issues mentioned above.

Based on the selected map projection, a co-ordinate system allowing to identify a point on Earth on a map needs to be developed. The most commonly known co-ordinates are the geographic latitude, longitude.

Figure 8. 16: Geographic co-ordinate system (DANA 1999)

As there are many projection systems, there are also many co-ordinate systems. In Germany for example the so-called Gauss-Krüger geodetic co-ordinate system (named after its developers) is based on a conform projection - meaning the shape of areas is stable - and referenced using the Bessel ellipsoid that is the name of the geodetic datum. Similar to that is the Universal Transverse Mercator (UTM) projection used to define horizontal, positions world-wide by dividing the surface of the Earth into 6 degree zones, each mapped by the Transverse Mercator projection with a central meridian in the centre of the zone. UTM zone numbers designate 6 degree longitudinal strips extending from 80 degrees South latitude to 84 degrees North latitude. UTM zone characters designate 8 degree zones extending north and south from the equator. (DANA 1999)

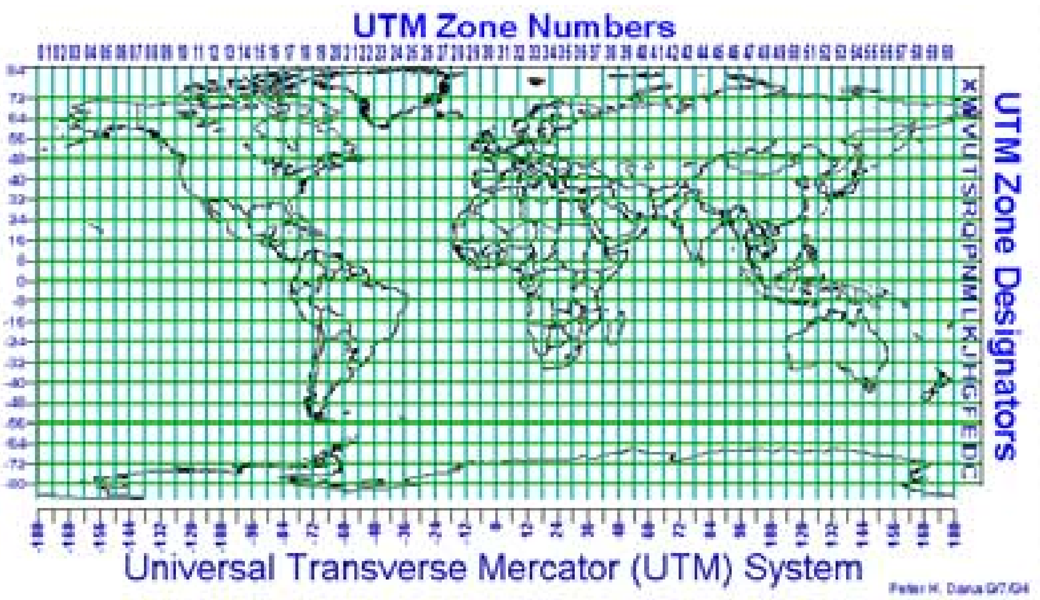


Figure 8.17: UTM Coordinate system (DANA 1999)

8.3.2. Regional Reference Systems

Statistical data is mostly related (or 'geo-coded') to a geographic extent, e.g. a specific town or a country or a group of countries like the European Union. It is anticipated that the statistical information **is harmonised and comparable** across the **whole geographic extension** of this so-called **Regional Reference System**. In the geographer's vocabulary a similar concept is defined as a 'Geographic Identifiers System' that is a "structured collection of geographic identifiers with a common theme and format', e.g. a post code (CEN 1998).

In general statistics are **spatially differentiated within** the geographic extent, meaning spatially referenced to smaller parts of the geographic extent e.g. statistical regions, administrative areas or other functionally defined regions (e.g. travel to work areas). If these spatial units, as it is mostly the case, are defined in a hierarchical system, meaning that smaller regions may be aggregated to regenerate a statistic for a larger region, one can talk about different levels of the Regional Reference System, e.g. NUTS III, II and I. The finer the regional reference system, the higher the geometric resolution (and the number of units) and the smaller the restitution unit.

Examples of Regional Reference Systems:

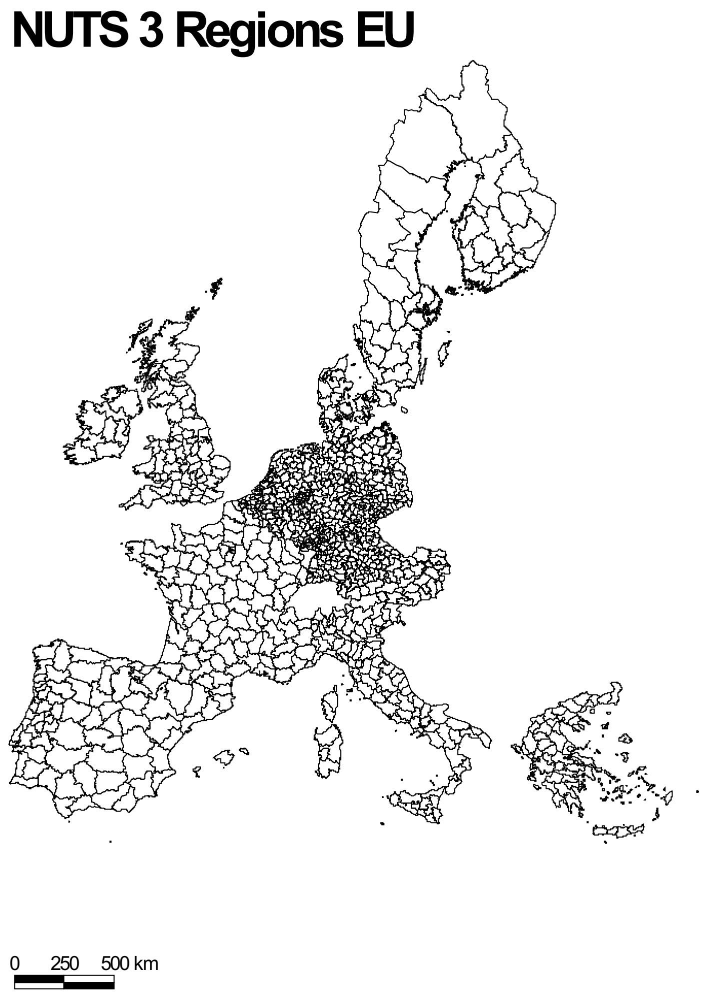


Figure 8.18: NUTS Regions (Eurostat 1999)

Figure 8.19: Regionales Bezugssystem Berlin, Statistisches Landesamt Berlin

Spatial aggregation does not necessarily mean that the thematic content of the data is also required to be aggregated or 'generalised'. For example, statistics per country (EU Member State) on application of fertilisers to agricultural land, split into nitrogen fertilisers and phosphate fertilisers, may be aggregated (= summed up) to the European level with aggregation to one figure on fertilisers (nitrogen + phosphate) and without aggregating the two classes. If data is collected on a sampling for all Europe (European level) and for fertilisers in general, it is not possible to segregate i) from European to Member State level and ii) from one class (nitrogen + phosphate) to 2 classes (nitrogen and phosphate).

Non-cartographers mostly use the notion 'scale' in a wider sense. When talking about scale statisticians for example mean the geographic extent of their area of interest or the level of their regional reference system e.g. '… European scale at NUTS 2 regions level'. Biologists and ecologists may think in terms of habitats of species, which can be a pond or a forest or a whole mountain area, geographers think in terms of patches of a landscape.

To avoid confusion in interdisciplinary collaboration care needs to be taken because in the cartographic sense, the small-scale map is highly generalised in terms of geometry **and** of thematic content (simply because of the size of the Minimum Legible Delineation) what does not necessarily apply to statistical information over Europe which can be very detailed in terms of content (because of sampling) but not regarding the restitution unit, e.g. national level.

8.4. Scale

In discussions of experts coming from different disciplines, the notion 'scale' sometimes causes confusion because of the different concepts behind it.

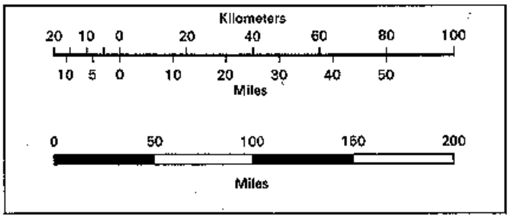
Generally, statisticians use the term scale synonymous with the level of the regional reference system. Ecologists use the term 'spatial scale' to describe 2 characteristics of data collection; **grain** that is the finest spatial or geometric resolution within which data are collected, and **extent** that is the size of the area of interest or the geographic extent of the statistician's regional reference system. The adjectives 'small' and 'large' used in conjunction with scale cause also sometimes misunderstanding. In a common sense, a large scale application refers to a project covering a large area and conversely covers a small scale analysis a little study area. In the following the cartographic definition of scale is described.

Scale is the linear reduction relation of a map on reality. To calculate the distance between two points in the real world, the distance measured on the map needs to be multiplied by the scale factor, e.g. scale 1:1.000.000 means a distance of 1 cm in the map corresponds to a distance of 1.000.000 cm = 10.000 m = 10 Km in the real world. Without a scale, a map is not a map, it is a diagram. As already discussed in previous chapters a map is always a geometrically distorted representation in two dimensions of a three dimensional world. Distortions may influence distance and area measurements as well as the shape of delineated areas depending on the projection and the coordinate system used on the map. In the narrow sense, the given scale of a map is an approximation over the whole area shown on the map because for each point true scale may be calculated using the exact parameters of the coordinates, the projection and the geodetic datum at that point on the Earth's surface.

The scale can be given in one of 3 ways:

1. Verbal statement, e.g. 1cm to 10km, 1" to 1 mile; simple to understand, but requires a 'ruler';

2. Scale bar: km or miles - most useful for graphic reproduction as bar changes with printing;



Often the left end of the bar is subdivided in smaller units in order to provide easier estimation of precise distances

Figure 8.20: Examples of bar scales (University of Texas at Austin - Department of Geography 1999):

3. Representative fraction (RF) states the amount of reduction as a fraction / ratio e.g. 1/100,000 or 1:100,000 (free of units), used to describe map series, but can be confusing some as right side of ratio 'increases' with decreasing scale:

   1:1.000 larger scale than 1:10.000 larger scale than 1:100.000 larger than 1:1.000.000 etc.

Going from a large scale to a small scale implies *generalisation* that means reduction of the amount of information presented on the map simply because the space available on the paper is also reduced. Polygons (areas) become points, two sides of a river become one line, some detailed information is dropped.

Scale in Spatial Databases

Strictly speaking the notion of scale is a cartographic concept and **does not apply** to abstract representations stored in a geographic database. In an ideal geographic database, the information is scale-less meaning that the co-ordinates of a point (line, polygon) correspond to co-ordinates in the real world. In fact, most data is digitised from maps, which have been created on a defined scale, and the error in geometry is propagated into the digital system.

In *SPATIAL DATABASES*, the concepts are:

- Precision: is the degree of detail in the reporting of a measurement.
- Accuracy: is the relation between a measurement and the reality, which it purports to represent.
- Resolution: the smallest object which can be represented

In *REMOTE SENSING* terminology 'the **Spatial Resolution** is the smallest area (spatial unit) on the ground over which the radiometric signal captured by a sensor is integrated'. The size of that unit depends on the characteristics of the sensor and the altitude of the platform. In discussions of statisticians and remote sensing specialist, this term is sometimes confused with *'observation unit'* or *'Mapping unit'.*

GOODCHILD, M (1991): Issue of quality and uncertainty. In: Advances in Cartography.- London (ed. MÜLLER) pp 113-139

Regarding generalisation, the content of a geographic database is not required to be generalised, it is just required for presentation purposes (on screen or on paper).

8.5. Units

8.5.1 Map Units

Map units are the metric measurement units in which the geographic data are stored (in any format: analogue or digital), such as inches, feet, or meters or degrees, minutes and seconds.

8.5.2 Mapping Unit

The mapping unit is the smallest area measured and mapped during the surveying. The size of this unit is first of all determined by the user requirement, in our case the scale of the map to be produced, or better to say the precision and accuracy required to identify the smallest objects to be mapped. Secondly, the size of the unit depends on geometric precision or the spatial resolution determined by the data collection technique or the tool, which should be selected accordingly.

The precision and the accuracy of terrestrial measurements may be several Millimetres or Centimetres e.g. for the network of trigonometric points, facility mapping, ownership maps. Other techniques of e.g. photogrammetric measurements from aerial photographs or satellite imagery result in mapping units in the range of less than a metre or several metres, depending on the spatial resolution. Objects smaller or with equal size of the spatial resolution may in general not be identified in optical remote sensing imagery. The assignment of a pixel to e.g. a land cover class requires the analysis of the neighbouring pixels, at minimum a 3x3 pixels window. In practice following an automatic or semi-classification procedure (e.g. cluster-analysis, Maximum-Likelihood classification etc.) a filtering process is carried out to eliminate single 'noise' pixels. Recent developments in classification of remote sensing imagery firstly identify segments of homogeneous spectral values and then classify these segments. The mapping unit would then be the minimum size of such segments to be classified.

8.5.3. Minimum Legible Delineation

The Minimum Legible Delineation (MLD) of a map is the minimum legible size of a line or a polygon on a paper map. A black line should be 0.05 mm thick, a coloured line 0.08-0.1 mm. Concerning the spatial dimensions of polygons on the map the width should be 0.3 mm, for coloured areas 1 mm².

Thickness of a black line: 0.05 mm

Thickness of a coloured line: 0.08 - 0,1 mm

Distance between lines: 0.25 mm

Dimensions of polygons 0.3 mm

Space in between filled polygons: 0.2 mm

The figures are oversized.

Depending on the scale of the map, this unit represents variable surfaces in reality: at a scale of 1:1.000 the smallest surface to be represented is 1 m² (1 mm * 1 mm on the map), at a scale of 1:100.000 the same surface on the map represents 10.000 m² (=1ha). That means that precision and accuracy of terrestrial measurement of centimetres will be blurred if represented at a scale of 1:100.000 on a paper map: 1 mm on the map corresponds to 100 metres in reality. The same concept applies when surveying in the field: a line of a piece of paper represents a certain area depending on the scale of the map to be drawn.

Sometimes the MLD is confused with 'Mapping Unit' of a map.

8.5.4. Observation Unit

In statistical investigations, observation units are the subjects of study, e.g. depending on the question, human beings, families, households, plants, animals, climate stations, agricultural parcels, administrative areas or other spatial units etc. The properties of investigation, the variables, maybe of various type: age, number of children, income of household, degree of damage (of trees for examples), sunshine duration, land cover or land use etc.

8.5.5. Restitution Unit

In general, the restitution unit is the unit in which statistical data are represented. These units may have various dimensions of time (interval) and space (regions).

The data collected during a population census in a whole country may be presented for each household. In this case the observation unit is the restitution unit. This is the case in several Nordic countries. The results may also be aggregated in space to an upper level of the regional reference system, e.g. the NUTS 4 that becomes then the restitution unit. Data resulting from an area frame sample is representative for a certain regional reference unit depending on the sample design.

One of the most crucial problems in interdisciplinary collaboration is the misunderstanding or misinterpretation of terms that are precisely defined but based on different sectoral concepts. This is especially the case when similar or even the same words are used in the terminology.

Misunderstandings or confusion when using the terms 'mapping unit', and 'observation unit', 'minimum legible delineation', 'spatial resolution' and 'restitution unit' may be recognised in discussions between statisticians and geographers (cartographers, remote sensers).

As described above, the mapping unit in the cartographic terminology is always a surface. The required size of the mapping unit depends on the scale of the map to be created (that depends on the size of the objects to be mapped) which also determines the technique for data collection e.g. terrestrial survey or aerial photogrammetry. If information needs require very high geometric precision and accuracy, for example a facility map showing the water or electricity supply lines to enable the planning of construction sites, the scale of the resulting map is required to be very large, e.g. 1:100 or 1:500. If ownership of land parcels is the subject of the map, the scale required may be 1:1.000 or 1:5.000. The geographic data is 'only' used as background for thematic mapping, e.g. showing the population density of the EU Member States referenced to NUTS II regions, the geographic extent to be covered, which is the whole EU territory, and the space available to print the map determines the scale to be used (e.g. 1:30.000.000 for A4 format). The geometric precision and accuracy in that case is not an important issue.

If for example the restitution unit for the same map is the Enumeration District (corresponding to the finest level of the regional reference system), the minimum legible delineation will play a role in the determination of the scale: an enumeration district of a size of 100 ha will be represented on a map at a 1:10.000.000 scale with an area of 1 mm², which corresponds to the minimum legible delineation of a coloured spot on a paper map.

As mentioned above, the observation units of statistical investigations **may be** delimited areas, e.g. parcels, habitats, urban agglomerations, administrative regions such as communes, NUTS regions or countries. A distinction is to be made between the regional reference of an observation unit and an observation unit that **is** a region.

The observation unit in the statistical terminology has nearly always a regional reference: statistics normally refer to a geographic extent: a whole country e.g. population number of France or a smaller precisely delimited area like unemployment in Dublin. In addition to the geographic extent statistics may have a finer geographic reference. A regional reference system allows comparison of properties of observation units in different regions. Normally such a system is organised in a hierarchical order and consists of addresses at the smallest regional reference, statistical blocks (or enumeration districts), districts, communes, departments, provinces up to the national level and international levels (EU, OECD, WTO Member States etc.). Of course it is a prerequisite for comparison that the statistics are harmonised or even standardised across the geographic extent.

8.6. Geocoded statistics and geographic information

In general all statistics are somehow *geocoded* meaning that they are referred to a regional reference system using alphanumerical codes of areas or locations (e.g. names, numbers): National Data refer to a whole country, regional data refer to smaller areas (mostly administrative regions, like the European NUTS regions, divided in 3 levels).

Geographic information is:

- information about places on the Earth's surface
- knowledge about where something is
- knowledge about what is at a given location

(GOODCHILD 1997)

Geographic information is traditionally 'stored' in analogue format on paper maps. This format has several constraints concerned with the presentation and the content of the information, depending on the scale, the grade of generalisation and the type of the map. Today, geographic information is handled like many other types of information in computerised systems.

Ideal geographic objects have spatial boundaries and a well defined set of attributes, for example land parcels with accurate and precise boundaries surveyed in the field and attributes like ownership, actual use, permitted use, tax value and so on, which apply uniformly to the whole object[7].

Points, lines and areas delineate these geographic objects in a defined and absolute geographic reference system. Lines are composed by points with exact co-ordinates of the absolute reference system, areas (polygons) are composed by lines. The information on the spatial delineation of such geographic phenomena is called geographic data.

[7] "… there are [other] geographic phenomena that are more often thought of as continuous fields - air pressure, elevation as represented by the hypsometric surface, hydraulic heads or pollution plumes. These are usually represented by smooth mathematical surfaces (often polynomial functions) that vary continuously and smoothly over space-time." (BURROUGH & FRANK 1996 p.4). The delineation of such phenomena is another problem.

Analogue geographic information systems (=Maps) present those points, lines and areas (areas=polygons) at a specified scale on a piece of paper. Digital geographic information systems store the real world co-ordinates of the geographic objects in numerical format.

A Geographic information system (GIS) is

- a system for input, storage, manipulation, and output of geographic information
- a class of software
- a practical instance of a GIS combines software with hardware, data, a user, etc., to solve a problem, support a decision, help to plan

  (GOODCHILD 1997)

8.7. Georeferencing, Geocoding

8.7.1 Georeferencing

As already mentioned, geographic objects are delineated by points, lines (which are linked points) and areas (which are linked lines). The assignment of co-ordinates of an absolute geographic reference system to those points (lines, polygons) is called *geo-referencing*. For example the geographic co-ordinates of the airport of Lisbon are (approx.) 38°45' North (of the equator) and 9°19' West (of Greenwich).

- In the *REMOTE SENSING* terminology georeferencing concerns image pre-processing regarding geometric correction:
- *'Geometric Correction'* is the processing procedure that corrects spatial distortions in an image.
- *'Registration'* is the process of superposing two or more images or photographs so that equivalent geographic points coincide.
- *'Georeferencing'* is the transformation of the image (each pixel) according to a cartographic reference system.

Those processing techniques require assigning, calculating or 'resampling' a new value for each pixel to be moved to another location within the raster grid of the image: with the help of for example Ground Control Points, which co-ordinates are known from a map, and the corresponding Image Control Points, the geographic co-ordinates that means the precise and accurate locations of the pixel can be calculated in a new geo-referenced image matrix. The process of recalculation and assigning a new spectral value for the pixel location in the geo-referenced image is called *'Resampling'.* It is based on the values in the local area around the uncorrected pixels in the original image matrix. Resampling may be carried out by using simply the value of the nearest neighbour pixel or by interpolation using a window around the old pixel position.

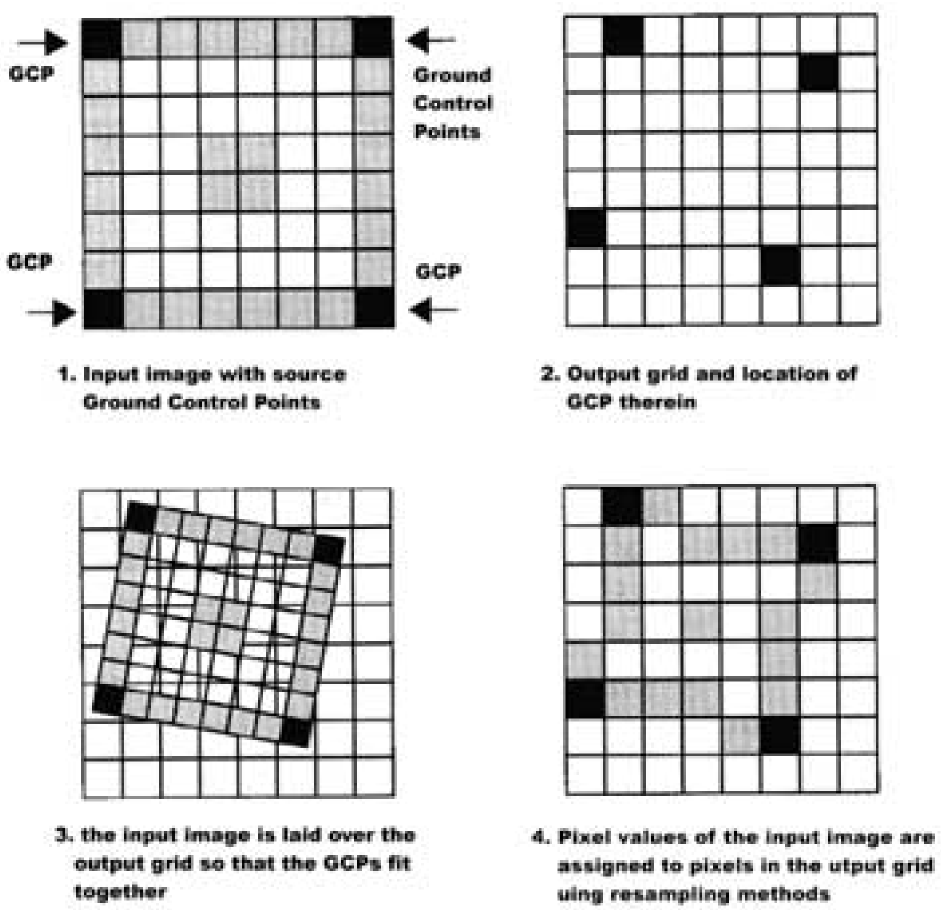


Figure 8.21: Principle of georeferencing (ERDAS 1994)

8.7.2. Geocoding

The process of assigning a code (e.g. name, number) to a geographic feature is called *geocoding*.

In *STATISTICS* the terms 'geocoding' or 'geo-referencing' are used in a wider sense, meaning that a statistical figure is referenced or coded to a regional reference, mostly a statistical territorial unit e.g. a statistical block, a NUTS region (see: Regional Reference System). The Lisbon airport is geocoded with RC132 'Grande Lisboa' at NUTS 3 level.

In a population census, data corresponding to observation units - individuals - may be referred to (= geo-coded) the address (very precise) or the statistical block (spatially aggregated). With this specification a geo-reference in the narrow sense - assignment of geographic co-ordinates - is not yet given because the data may be stored in tables with the addresses or the codes of enumeration districts. The geographic data require for georeferencing the point consists of the boundaries of the reference units in geographic co-ordinates (e.g. degrees, minutes, seconds or meters using a defined co-ordinate system). The geo-reference allows for the building up of topology that means neighbourhood relationships between the reference regions, which is crucial for certain types of spatial analysis.

In the case of an area sampling, the information collected on a part of the area is representative for - and to be assigned (geo-coded) to - the whole area[8]. If statistics are to be geo-coded more precisely, at a lower level of the regional reference system, a representative sample needs to be defined according to the size (in terms of area) and the number of reference units of the different levels of that regional reference system. The sample design determines the minimum spatial restitution unit.

[8] For example, the MARS area frame sample is designed to allow estimates at the European level. The samples are not representative on the national level (meaning for single EU Member States).

9. REFERENCES

9. REFERENCES

ABEL, D. J., BENG CHIN O., KIAN-LEE T. & SOON HUAT T. (1998): Towards integrated geographical information processing. - In: International Journal on Geographic Information Science, vol. 12, N°4 1998, pp 353-371

ALBERTZ, J. (1991): Grundlagen der Interpretation von Luft- und Satellitenbildern : eine Einführung in die Fernerkundung. Wissenschaftliche Buchgesellschaft Darmstadt

ALUN, J., CLARK, J. (1997): Driving forces behind European land use change: an overview. CLAUDE Resource Paper No.1

BÄHR, H.-P. (1985): Digitale Bildverarbeitung: Anwendung in der Photogrammetrie und Fernerkundung.- Karlsruhe

BAULIES, X.I., SZEJWACH, G. (ed.) (1997): Survey of needs, gaps and priorities on data for land use and land cover change research. LUCC Data requirements workshop, Barcelona 11-14 November 1997, LUCC report series 3

BRUNET, R. (1987): La Carte Mode d'Emploi.- Paris

BUNGE, M. (1983): Epistemology and methodology I: exploring the world, vol. 5 of Treatise on basic philosophy, Dordrecht and Boston, Reidel

BURROUGH, P.A. & FRANK, A.U. (1996): Geographic objects with indeterminate boundaries.- GISDATA series European Science Foundation, London, Bristol

Canadian Centre for Remote Sensing (1999): Fundamentals of Remote Sensing Tutorial

Canadian Cente for Remote Sensing (1999): Glossary - http://www.ccrs.nrcan.gc.ca/ccrs/eduref/ref/glosndxe.html

CEN (1998): Geographic information - Vocabulary. - CEN Report CR 13436:1998, Paris

CHRISMAN, N R. (1998): Rethinking levels of measurements for cartography.- In Cartography and Geographic Information Systems, Vol. 25 N°4, 1998, pp. 231-242

COCHRAN, W.G. (1977): Sampling Techniques. Wiley, New York, 3rd edition

CONGALTON, R. (1991): A review of assessing the accuracy of classifications of remotely sensed data.- Remote Sensing of Environment, Vol. 37, pp. 35-46

COTTER, J.; NEALON, J. (1987): Area Frame Design for Agricultural Surveys. United States Department of agriculture, National Agricultural Statistics Service, Research and Applications Division

COWEN, D. J. (1997): NCGIA Core Curriculum in GIScience

CROI, W. (1999): European Land Cover and Land Use Information Database.- Working Party Land Use Statistics May 1999 (Doc. LAND/19)

DANA, H. P. (1999): The Geographer's Craft Project, Department of Geography, The University of Texas at Austin http://www.utexas.edu/depts/grg/gcraft/notes/datum/datum_ftoc.html#Contents

DI GREGORIO, A. & JANSEN, LJM. (1997): A new concept for a Land Cover Classification System. Earth observation and evolution classification, 1997, Conference Proceedings. 13-16 October 1997, Alexandria, Egypt, 10p.p.

DEVOGELE, T., PARENT, C. & SPACCAPIETRA, S. (1998): On spatial database integration.- In: International Journal on Geographic Information Science, Vol. 12, N°4 1998, pp 335-352

DRURY, S.A. (1990):: A Guide to Remote Sensing. -Oxford, 199 pp.

DUHAMEL, C. & CROI, W. (1998): LANES - Development of a harmonised framework for multi-purpose lan cover / land use information systems derived from earth observation data.- Final Report (unpubl.)

DUHAMEL, C. (1998): Cesd-Communautaire: Background paper on nomenclatures, LANES concerted action, Research program on the Environment and climate, European commission

DUHAMEL, C. & VIDAL, C. (1999): Objectives, tools and nomenclatures. In Eurostat (1999) : Land Cover and Land Use information systems for European Union Policy needs. Proceedings of the seminar Luxembourg 21-23 January 1998

DUPRAT, H. (1972): Les conditions linguistiques du transfert technologique in Revue Automatisme Tome XVII n°11 Paris

ECO, U. (1988): Le Signe, Editions Labor Bruxelles

EKHOLM, A. (1996): A conceptual framework for classification of construction works

EUROSTAT (1992): Task force statistiques spatiales, rapport final Statistiques d’utilisation des sols

EUROSTAT (1995): Régions - Nomenclature des Unités Territoriales Statistiques NUTS.- Luxembourg unpubl.

EUROSTAT (1996): CLUSTERS (Classification of Land Use STatistics - Eurostat Remote Sensing): a pilot nomenclature.- In: A statistical framework for remote sensing applications in statistics.- unpubl.

EUROSTAT (1999): Nomenclature of Territorial Units for Statistics NUTS.- Luxembourg

Eurostat (1998): The GISCO Database Manual.- unpubli.

Erdas (1994): Field Guide. - Atlanta, GA, USA

FAO (1996): Multiple frame agricultural surveys. Current surveys based on area and list frame sampling methods. FAO Statistical Development Series, Vol. 1

FAO (1998): Multiple frame agricultural surveys. Agricultural survey programs based on area frame or dual frame (area and list) sample designs. FAO Statistical Development Series Vol. 2

FRITSCH, D., GLEMSER, M., KLEIN, U., SESTER, M., STRUNZ, G. (1998): Zur Integration von Unsicherheit bei Vektor- und Rasterdaten.- In: Geographische Informationssystems 4 / 1998, pp 26-35

GALLEGO, F.J., CARFAGNA, E. (1998): Research project on Geographic sampling strategies and remote sensing - Final Report.- unpubl.

GALLEGO, F.J. (1995): Sampling Frames of Square Segments. Office for official Publications of the European Communities, Luxembourg

GÉOIMAGE (1995): Utilisation de la télédétection pour l'inventaire, le suivi et la gestion du couloir rhodanien^- Rapport final.- Eurostat project, unpubl.

GOODCHILD, M. F. (1997): What is Geographic Information Science? - NCGIA Core Curriculum in GIScience, http://www.ncgia.ucsb.edu /giscc/units/u002/u002.html, posted October 7, 1997.

GOODCHILD; M. F.; LONGLEY, P. A.; MAGUIRE, D. J.; RHIND, D. W. (1999): Geographic Information Systems. - 2nd edition, New York

HAKE, G. (1982): Kartographie.- Berlin, New York

HJELMSLEV, L. (1959): Pour une sémantique structurale in Essais linguistiques Copenhagen, Nordisk Sproog-OG Kulturforlag

HUNTING Technical Service & The Centre for Agricultural Strategy (1997): Quality Control Procedures for Land Use Data derived using Remote Sensing Technology - Final Report.- Eurostat Research Project (unpubl.)

JAAKKOLA, O. (1998): Multi-scale categorical data bases with automatic generalisation transformations based on map algebra.- In: Cartography and geographic information systems, Vol.25, No. 4, pp. 195-207

JANSEN, L.J.M.; DI GREGORIO A. (1999): The Problems of Current Land Cover classifications: Development of a new approach. In Eurostat (1999): Land Cover and Land Use information systems for European Union Policy Needs. Proceedings of the seminar Luxembourg 21.-23. January 1998

LAURINI, R. (1998): Spatial multi-database topological continuity and indexing: a step towards seamless GIS data interoperability.- In: International Journal on Geographic Information Science, Vol. 12, N°4 1998, pp 373-402

LILLESAND, T.M.; KIEFER, R.W. (1994): Remote sensing and image interpretation. 3rd ed. Wiley, New York

LUND, HG. (1998): A “Forest” by any other name. Paper in Land Use, Land Use change and forestry in the Kyoto protocol. Environmental Science & Policy, Elsevier Science publications

McCARTHY, D, (1996): International Earth Rotation Service (IERS) Technical Note 21.- IERS Conventions, Observatoire de Paris

MÜCHER, C.A., STOMPH, T.J., Fresco L.O. (1993): Proposal for a global land use classification FAO, Rome 37 p.p.

MEGRIN (1999): Geographical Data Description Directory Version 2.1 - GDDD Data Model.- http://www.megrin.org/GDDD/Overview.html

NASA Goddard Space Flight Applied Information Sciences Branch Center (1999): Remote Sensing and Photo Interpretation Tutorial, available at http://rst.gsfc.nasa.gov/TofC/table.html

National Imagery and Mapping Agency (2000): Technical Report 8350.2 3rd edition: World Geodetic System 1984. - Bethesda

OLSON, J. M. (1998): Maps as Representations of the World, NCGIA Core Curriculum in GIS, National Center for Geographic Information and Analysis, University of California, Santa Barbara, Unit 020, http://www.ncgia.ucsb.edu/giscc/units/u020/u020.html, posted December 23, 1998.

POIRET, M. (1997): Groupe de travail sur Les Statistiques d’Utilisation des Sols -

QUATTROCHI, D. A., SIU-NGAN LAM, Nina (1991): Perspectives on Integrating Multiscale. Multitemporal Remote Sensing Data with Geographic Information Systems.- In: The Integration of Remote Sensing and Geographic Information Systems.- Proceedings of a special session held at the Baltimore 1991 ACSM-ASPRS Annual Convention, ASPRS Bethesda Maryland USA

RADERMACHER, W. (1988): Gedanken zu einer Basisklassifikation der Bodennutzungen

ROBINSON, A.; MORRISON, J.; MUEHRCKE, P. C.; KIMERLING, J. and GUPTILL, S. C. (1995): Elements of Cartography.- 6th Edition. New York: John Wiley and Sons, Inc.

SABINS, F.F. (1987): Remote Sensing Principles and Interpretation - 2nd Edition, (W.H. Freeman & Co.), 449 pp.

SIMPSON, G.G. (1961): Principles of Animal Taxonomy, Columbia University Press, New York

SOKAL, R. (1974): Classification: purposes, principles, progress, prospects in Science vol. 185, number 4197, pp 111-123

SNEATH, PHA. and SOKAL R. (1973): Numerical taxonomy, Freeman and Co.

STATISTICS NORWAY (2000): Land use statistics for urban settlements.- Final Report for Eurostat

STEFANAKIS, E. SELLIS, T. (1998): Enhancing operations with spatial access methods in a database management system for GIS.- In Cartography and Geographic Information Systems, Vol. 25 N°4, 1998, pp. 16-32

STOTT, A.; HAINES-YOUNG, R. (1996): Linking land cover, intensity of use and botanical diversity in an accounting framework in the UK.

SUTCLIFFE, J.P. (1993): Concept, class and category in the tradition of Aristotle. Chapter 3 in Van Mecheler et al. (eds.) Categories and concepts: theoretical views and inductive data analysis, Academic press, London

SWISS FEDERAL STATISTICAL OFFICE (1996): Arealstatistik Schweiz: Die Bodennutzung in den Kantonen. Municipal results 1979/85 and 19992/97. This series (in French and German) gives the latest results of the follow-up survey as it is completed in each Kanton. The first volumes appeared in 1996, and the last will probably come out in 2001.

SYSAME (1992): Cadastres fonciers nationaux - Synthèse de l'etude comparative des cadastres nationaux réelisée pour le casier viticole communautaire.-

TAYLOR, C.; SANNIER, C.; DELINCÉ, J.; GALLEGO, F.J. (1997): Regional Crop Inventories in Europe assisted by Remote Sensing: 1988-1993. Synthesis Report of the MARS Project - Action 1. Official Publication of the European Communities, Luxembourg, EUR 17318

TOMPPO, E. (1992): Multi-source National Forest Inventory of Finland - Finnish Forest Research Institute - Research Papers 444, Helsinki

University of Texas at Austin - Department of Geography (1999): The Geographer's Craft Project.- Austin (URL: http://www.utexas.edu/depts/grg/gcraft/about/about.html)

WEISSEL, J. et. al (1999): Glossary. -Lamont-Doherty Earth Observatory of Columbia University; http://rst.gsfc.nasa.gov/Glossary/glossary.html

WÜSTER, E. (1971): Les classifications de notions et de thèmes; Différences essentielles et applications vol. 22 n°3 et 4 Bibliothèque Infoterm

WYATT, B. (1997): Guidelines for Land Use and Land Cover description and classification, Institute of Terrestrial Ecology - Report to FAO