Working Paper and Preprint

93/08

ISSUES OF DEVELOPMENT

AND

MANAGEMENT OF DATABASES

L.R. Oldeman

Prepared for the ISNAR/ICASA workshop on "Opportunities for use of systems research methods in agriculture in developing countries" The Hague, 22-24 November 1993

ISRIC

INTERNATIONAL SOIL REFERENCE AND INFORMATION CENTRE

Issues of Development and Management of Databases[1]

by
L.R. Oldeman[2]

1 INTRODUCTION

The intense and increased pressure on land and water resources, leading to degradation and pollution of those resources, and leading to a partial or complete loss of the productive capacity of the soil calls for an approach that strengthens the awareness of users of these resources on the dangers of inappropriate management and at the same time strengthens the capability of national soil/land resources institutions to deliver reliable, up-to-date information on land resources in an accessible format to a wide audience.

The development of an internationally accepted georeferenced system capable of providing accurate, useful and timely information on soil and terrain resources is a prerequisite for policy-makers, decision-makers, resource managers and for the scientific community at large in their assessment of the productive capacity of soils, of the status, risks and rates of soil degradation, and of global change.

The environmental links between soils, food crops and water supplies mean that if soils are degraded or polluted, there is every likelihood that food chains and drinking water will also be affected. Quantification of processes of soil degradation (and pollution) requires that researchers have access to compatible and uniform databases linked to a Geographic Information System, which include data layers on climate, landform, soils and land cover/land use.

The link between research and development can be established through a package of technology that consists of two major elements:

- The establishment of a detailed resource database on natural and human resources in such a way that this information can be immediately and easily accessed and combined.
- The establishment of scientific valid methods, that can analyze this land, water, climate, vegetation, land use and socio-economic information from the point of view of potential land use, both in relation to food an other human requirements and for assessing the environmental impact.

The International Soil Reference and Information Centre (ISRIC) in Wageningen, the Netherlands, has developed in close cooperation with the International Society of Soil Science (ISSS), the Food and Agricultural Organization of the United Nations (FAO), and the United Nations Environment Programme (UNEP), an internationally endorsed land resources information system, called World Soil and Terrain Digital Database (SOTER). The issues related to the development and management of this georeferenced agro-ecological database are discussed in this paper.

---

[1] Prepared for the ISNAR/ICASA workshop on "Opportunities for use of systems research methods in agriculture in developing countries", 22-24 November, 1993, the Hague, the Netherlands.

[2] International Soil Reference and Information Centre, P.O.Box 353, 6700 AJ Wageningen, the Netherlands.

2 THE NEED FOR A NATURAL RESOURCES DATABASE

Throughout history men has been engaged in collecting information on natural resources. The wealth of information generated by national and international institutions all over the world is stored in different forms: maps, tabular data, descriptive reports. Storage of this information is in first instance geared towards the need of the individual institute or country. As a consequence the stored material in one place is often not compatible with information stored elsewhere. The large amounts of information, whether on soils, on climate, on land cover, was generally assembled and then ordered, grouped or classified, and given symbols or names, so that scientists within a certain country or within a certain discipline could comprehend what characteristics are common to a certain soil or climate entity. Since no systems in the past were available to store information on natural resources in a standardized systematic way, scientists had no other options than to classify the natural resources. In national soil classification systems soils were given local names, which were perhaps meaningful to local soil scientists, but did not help a wide audience of non-soil scientists to better understand the characteristics of that soil. It became extremely difficult to extrapolate research results obtained at one site to larger areas; it was virtually impossible to exchange research results to other countries.

To solve the problems of international barriers to exchange natural resource information, international classification schemes were developed. In soil science, two major schools of classification emerged: the first one, the Soil Taxonomy System, developed by the Soil Conservation Service in the United States. The other international system was developed by FAO's Land and Water Division in the framework of their mapping activity for a Soil Map of the World. Both systems focused their classification systems on attributes of the soils which could be measured in the field or analyzed in the laboratory. In contrast the French soil classification system focused on soils processes and soil formation.

These international classification systems, now universally adopted were - and still are - an essential and comprehensive methodology to characterize the soils across national boundaries, but they have an important consequence. Only very experienced soil scientists understand the system. For the non-soil science audience, the system is incomprehensible.

Every new soil survey report published in a country is more advanced, more complex and technically more difficult than previously published reports. However, the majority of potential users of soil and terrain information are not technically trained in soils. The consequences are that these users either will not or cannot extract the specific information they need from a modern soil survey report. Purnell (1992) stated that there is a need for more reliable and more timely information about soils. This information must be more accessible and intelligible to non-soil scientists: "Clear language, free from jargon does not come easily to soil surveyors, but soil survey reports are of no value if not read".

The urgent need for a global inventory of the Earth's natural resources has been expressed by many international research organisation and policy-makers (UNEP, 1992; UNSO/UNDP, 1992; WRI, Call.Tech 1992; ISNAR, 1992). While there have been substantial advances in our understanding of environmental processes, and also a consensus has emerged that all forms of development must be sustainable, there are still important gaps in our understanding of how to assess adequately the impact of environmental damage. The many international initiatives of the modelling of the effects of an environmental change in combination with population growth in the developing countries are in dire need of sound ground-truthed and up-to-date spatial databases on agroclimatic conditions; on basic landform-soil relationships; on water resources and agro-hydrology; on present-day land cover and land use; on status and hazard of land degradation (Sombroek, 1992).

There is no doubt that what is needed at global level also applies at national and regional level. Rural development planning in developing countries is increasingly dependent upon reliable natural resource information. The history of agricultural development projects during the past decades is replete with examples of how the results might have been more positive if accurate and timely information about land and water resources had been available to the planners (ISSS, 1986).

For and improved assessment of land resources for the development of a sustained use of the land the following is necessary (Stoops and Cheverry, 1992):

- Database development
  Such a database should contain all available information on topography, soil attributes, climatic parameters, vegetation, land use, population (both human and animal population), infrastructure, as well as socio-economic factors such as food requirement attitudes, skills, costs of input (including labour), market availability, and stability.
- Geographic referencing
  Each item of information should be linked to its precise geographic position, using tools such as Geographic Information Systems (GIS) and remote sensing.
- Estimation of biomass
  Relevant modeling approaches can be used as a tool to make assessment of biomass from each land unit to evaluate levels of production at different levels of input.
- Estimation of status and risk of land degradation
  These risks of land degradation (e.g. loss of soils by water and wind erosion; nutrient decline; salinization; acidification; soil, water, and air pollution; soil physical degradation; vegetation degradation) can be assessed for given land units with specific land use and production systems to evaluate environmental impacts.
- Improved access to existing knowledge
  This will require the development of a documentation information system on development oriented research activities such as those reported in national and international journals.

The development of a georeferenced natural resources information system with internationally accepted standards will provide policy-makers, decision-makers and resource managers important and indispensable tools to reverse the current trends of soil degradation in developing and industrialized countries and to implement a programme for soil conservation and sustainable use of the land.

3 ISSUES OF DEVELOPMENT OF DATABASES

The objective of a natural resources information system is to utilize emerging information technology to produce an internationally accepted, standardized database, containing digitized map unit boundaries in a GIS and their attribute database in a relational database management system (RDBMS). The World Soils and Terrain Database (SOTER) is an example of such an information system, which will hold the data necessary for improved mapping and characterization of world soil and terrain resources, and for monitoring changes therein. This database is organized in such a way that it provides a comprehensive framework for the storage and retrieval of uniform soil and terrain data, that can be used for a wide range of applications at different scales.

3.1 Organisation of databases for agro-ecological characterization

Two types of database are required in every discipline engaged in mapping of spatial phenomena: 1) geometric data, i.e. the location and extent of an object represented by a point, a line or a surface, including its topology (shapes, neighbours and hierarchy of delineations) and 2) attribute data, i.e. the characteristics of the object. The geometry is stored in that part of the database that is handled by GIS software, while the attribute data is stored in a separate set of attribute files, manipulated by a RDBMS. A unique label attached to both the geometric and the attribute database connect these two types of information for each mapped unit (see figure 1). The overall systems (GIS plus RDBMS) stores and handles both the geometric and attribute database.

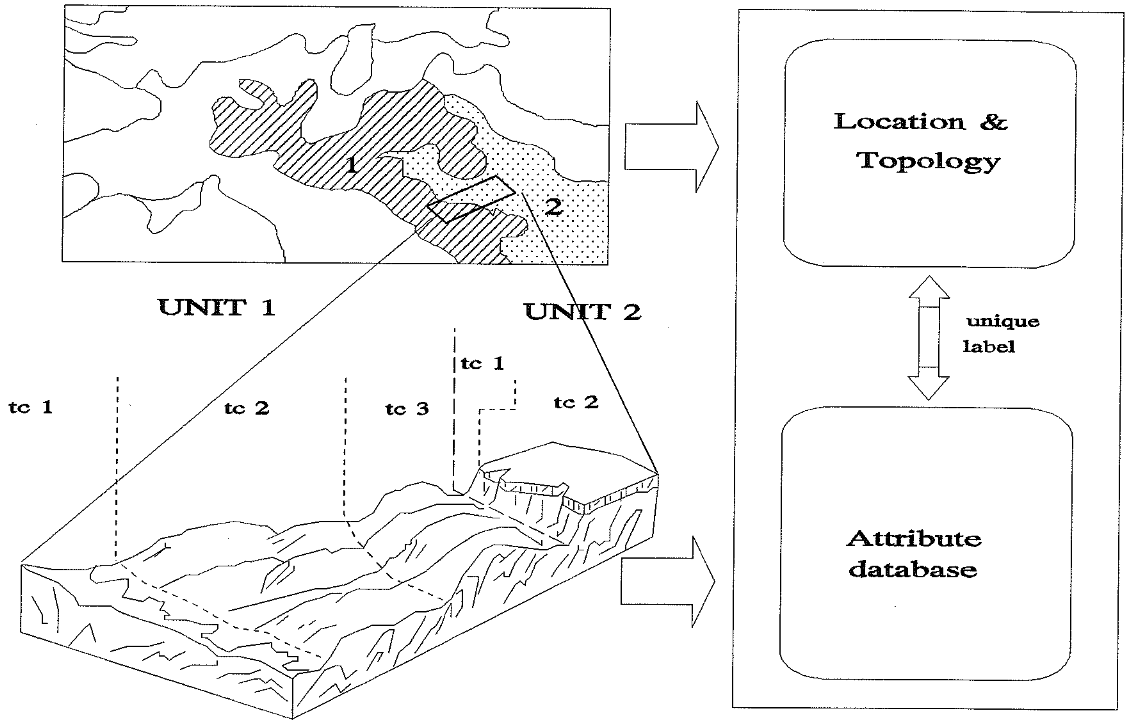

Figure 1 SOTER units, their terrain components (tc), attributes, and location.

3.1.1 The geometric database

This database contains information on the delineations of the mapped unit. It also includes data on features of the base topographic map, such as roads and towns, the hydrological network and administrative boundaries. In order to enhance the usefulness of the database, it will be possible to include additional overlays for boundaries outside the physiographic unit mosaic. Examples of such overlays could be socio-economic areas (population densities), hydrological units (watersheds) or other natural resource patterns (vegetation, agro-ecological zones).

3.1.2 The attribute database

A relational database is considered to be one of the most effective and flexible tools for storing and managing non-spatial attributes (Pulles, 1988). Under such a system the data is stored in tables, whose records are related to each other through specific identification codes (primary keys). These codes form the link between the various subsections of the database (e.g. the terrain table, the terrain component table and the soil component tables in SOTER). Another characteristic of the relational database is that when two or more components are similar, their attribute data need only to be entered once. Figure 2 gives a schematic representation of the organisation of the attribute database. The blocks represent tables in the SOTER database and the solid lines between the blocks indicate the links between the tables.

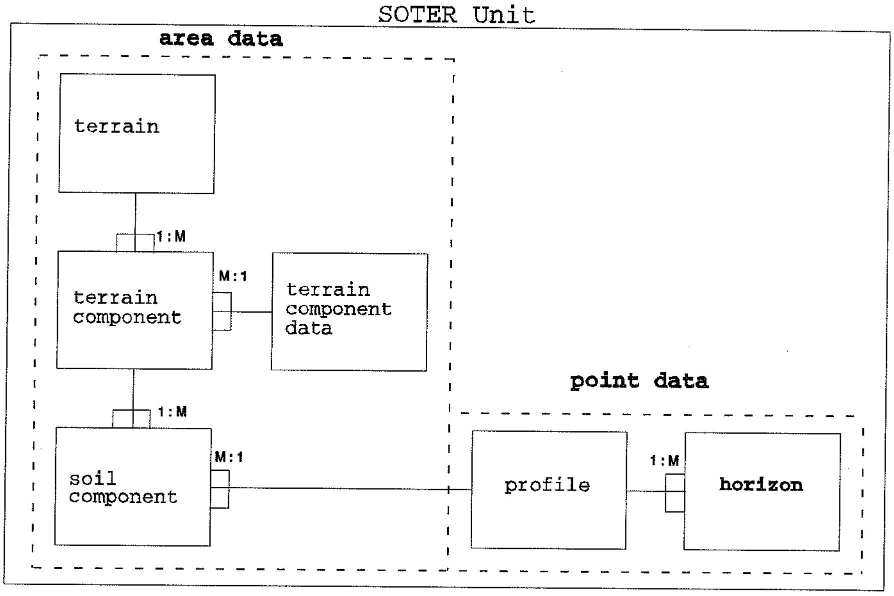

Figure 2 SOTER attribute database structure with area and point data (1:M = one to many, M:1 = many to one relations).

3.2 Characteristics of an agro-ecological database

An agro-ecological database should provide a comprehensive framework for the storage and retrieval of uniform data on soil and terrain data, on climatological, land cover and land use data, that can be used for a wide range of applications at different scales. The SOTER database holds soil and terrain information and will contain sufficient information extraction at a resolution of 1:1 Million, both in forms of maps and as tables. However the hierarchical structure of components of the database will allow the system to become also useful for continental databases at smaller scales, and for national database development at larger scales from 1:1 Million to 1:100.000.

The system will be amenable to updating, and completing. This implies that the system can be filled immediately with the wealth of information that is available now, but when new information becomes available, it can be entered at any time. Old data sets can be saved for monitoring purposes. The orderly arrangements of attributes in the database also helps to identify possible gaps in information.

A significant advantage of a georeferenced database is that it is user-oriented. While conventional natural resource mapping is a once-in-a-time exercise, directed to a few selected users, a georeferenced database is providing information to a broad array of international, regional, national environmental specialists and policy-makers through the provision of standardized resource maps, interpretative maps and tabular information at the specific request of the user.

3.3 Mapping issues

The concept of a georeferenced database implies the mapping or delineation of areas. Being an agro-ecological database the basic approach in mapping is the delineation of areas of land with a distinctive pattern of landform, surface form, slope, parent material and soils. Differentiating criteria have to be developed. These are then applied in a stepwise manner, following a hierarchical structure.

At the highest level of differentiation major land forms are identified and quantified (dominant slope gradient and relief intensity). Areas in a similar landform can be further segregated according to their lithology (or parent material). Differentiating criteria for dividing these terrain units into terrain components are surface form, meso relief, aspects of the parent material.

At this stage the complexity of terrain components does not always allow individual mapping at the 1:1 M reference scale. In such cases the percentage of occurrence of these non-mappable terrain components in the mapped unit is indicated in the terrain component table, while the attributes of these non-mappable terrain components is stored in the terrain component data tables (see figure 2).

The final step in the differentiation of the terrain is the identification of soil components within the terrain component. Differentiating criteria are based on diagnostic horizons and properties as formulated by FAO (FAO, 1988). As with terrain components, these soil components can be mappable or non-mappable. Most likely these terrain components comprise (at the reference scale of 1:1 M) a number of soil associations or soil complexes. The percentage of occurrence within the mapped unit is indicated in the soil component table, while the attributes are stored in the soil profile and soil horizon data tables.

3.4 Content of the database

The attributes of the database are elements that can be quantified, either through visual observations in the field or by measuring in the laboratory. The general approach adopted by SOTER is to screen all existing soil and terrain data in a georeferenced area and to complement the terrain information with remote sensing data where necessary. The various attributes of the terrain, terrain component and soil component tables are listed in Table 1. Each attribute is described in detail in the SOTER Procedures Manual (Van Engelen and Wen Tin Tiang, 1993). Some issues related to the content of the database are:

1. Descriptive attributes (land forms, lithology, surface forms).
   Description are foremost by their morphology, and not by genetic origin, or by processes responsible for their shape.
2. Class values versus numeric values.
   Numeric values are needed for any algorithms, for which attributes values are needed. Class values are nevertheless also important for land evaluation purposes.
3. Mandatory versus optional values.
   The attributes for terrain and terrain components are either directly available or can be derived from other parameters during the compilation of the database. Many of the soil parameters consist of measured values of which the availability varies considerably. A minimum set of soil attributes is generally needed for any realistic interpretation. Their presence in the database is mandatory. In case data is not available for some of these mandatory attributes, the database management allows expert estimates to be used. Measured and estimated values will be stored separately.
4. Spatial variability.
   A major problem in any survey is the reconciliation of point data and area data. In SOTER for every soil component at least one, but preferably more, fully described and analyzed reference profiles should be available from existing soil information sources. Following judicious selection by the local soil survey team, one of these reference profiles will be designated as the representative profile for the soil component. In order to give some degree of variability for soil profile data single values taken from the representative profile are complemented with maximum values and minimum values taken from all available profiles within the soil component. These values are an indication of the range of variation that exist within the component. Although statistical parameters like standard deviation and means could give a better idea of the variability this is in many cases not realistic considering the small number of profiles generally available for the compilation of the soil component.

Table 1. Non-spatial attributes of a SOTER unit.

TERRAIN

1 SOTER unit_ID
2 year of data collection
3 map_ID
4 minimum elevation
5 maximum elevation
6 slope gradient
7 relief intensity
8 major landform
9 regional slope
10 hypsometry
11 dissection
12 general lithology
13 permanent water surface

TERRAIN COMPONENT

14 SOTER unit_ID
15 terrain component number
16 proportion of SOTER unit
17 terrain component data_ID

TERRAIN COMPONENT DATA

18 terrain component data_ID
19 dominant slope
20 length of slope
21 form of slope
22 local surface form
23 average height
24 coverage
25 surface lithology
26 texture group non-consolidated parent material
27 depth to bedrock
28 surface drainage
29 depth to groundwater
30 frequency of flooding
31 duration of flooding
32 start of flooding

SOIL COMPONENT

33 SOTER unit_ID
34 terrain component number
35 soil component number
36 proportion of SOTER unit
37 profile_ID
38 number of reference profiles
39 position in terrain component
40 surface rockiness
41 surface stoniness
42 types of erosion/deposition
43 area affected
44 degree of erosion
45 sensitivity to capping
46 rootable depth
47 relation with other soil components

PROFILE

48 profile_ID
49 profile database_ID
50 latitude
51 longitude
52 elevation
53 sampling date
54 lab_ID
55 drainage
56 infiltration rate
57 surface organic matter
58 classification FAO
59 classification version
60 national classification
61 Soil Taxonomy
62 phase

HORIZON (* = mandatory)

63 profile_ID*
64 horizon number*
65 diagnostic horizon*
66 diagnostic property*
67 horizon designation
68 lower depth*
69 distinctness of transition
70 moist colour*
71 dry colour
72 grade of structure
73 size of structure elements
74 type of structure*
75 abundance of coarse fragments*
76 size of coarse fragments
77 very coarse sand
78 coarse sand
79 medium sand
80 fine sand
81 very fine sand
82 total sand*
83 silt*
84 clay*
85 particle size class
86 bulk density*
87 moisture content at various tensions
88 hydraulic conductivity
89 infiltration rate
90 pH H_2O*
91 pH KCl
92 electrical conductivity
93 exchangeable Ca^{++}
94 exchangeable Mg^{++}
95 exchangeable Na^{+}
96 exchangeable K^{+}
97 exchangeable Al^{+++}
98 exchangeable acidity
99 CEC soil*
100 total carbonate equivalent
101 gypsum
102 total carbon*
103 total nitrogen
104 P_2O_5
105 phosphate retention
106 Fe dithionite
107 Al dithionite
108 Fe pyrophosphate
109 Al pyrophosphate
110 clay mineralogy

3.5 Source materials

A major issue of the development of database is the availability of information. An agro-ecological information system like SOTER is based on existing soil and terrain information and specifically excludes new land resource surveys within its programme. Basic data sources for the construction of SOTER mapping units are existing topographic, morphological, geological and soil maps. All soil maps, that are accompanied by sufficient analytical data for soil characterization can be used and data entered into the database. One of the major advantages of the orderly arrangement of information in a database is that gaps in information can be more easily identified. Some extra analytical work may then have to be carried out to complement existing information.

A database like SOTER should include a reference file which contains information on the source materials used for the compilation of the database: in case of SOTER this would include information of the source of the map, of the laboratories that analyzed the soil samples, of the laboratory methods employed and of the organisations responsible for the national profile database. An ideal situation where all soil samples are analyzed according to standardized analytical methods does not exist.

3.6 Associated Data

The issues of development and management of databases discussed so far are related to soil and terrain databases. Climatic data form an inseparable part of the basic inventory of natural resources. Nevertheless climatic data are treated separately from soil and terrain data since this set of information is not directly linked to the physiographic SOTER areas. Climatic data are based on point observation only. The link with the soil and terrain information exists by means of the geographical location of these points.

Land cover characteristics (vegetation and land use) are also stored in separate files. In contrast with the more stable terrain and soil characteristics, land cover is considered a more dynamic entity which can change quickly in time. Thus there may be a frequent need for addition of more recent data. Moreover, other groups are working on global databases for land use and for vegetation. At present such databases are not available, but they should be included in a database for agro-ecological characterization. In SOTER, the land cover information is given at the level of the mapped unit. Land use classes are defined in a hierarchical system (Remmelzwaal, 1990), while for vegetation the generalized hierarchical description of the physiognomy of the present native vegetation developed by Unesco is followed (Unesco, 1973).

4 CONCLUSION

The rapid growth of the world population leads to the key question whether the supply of the land will accommodate a doubling of global agricultural demands in 40 years time. Although there are estimates (Buringh and Dudal, 1987) that around 1800 million hectares of land not yet used for agricultural food corps is still available and suitable to grow crops, most of this land has a moderate to low productivity, the spatial distribution of available land is not in proportion to the population density, the infrastructure needed to transport the produce to domestic and foreign markets form this potential available land is often very poorly developed or absent, the potential hazards for soil erosion and soil degradation are threatening once these areas -now under forest or grasslands- are transformed into croplands, and conversion of tropical forests into cropland is a hotly debated issue.

Past and present human intervention in the utilization and manipulation of environmental resources are having unanticipated consequences. The often indiscriminate destruction of forests and woodlands, and the spectre of land degradation resulting in decreased productivity with dire social consequences is generally recognized. "There is a growing realization in national and international institutions that not only many forms of economic development erode the environmental resources upon which they are based, but at the same time environmental degradation can undermine economic development" (Brundtland *et al.*, 1987). Soil degradation implies by definition a social problem. Environmental processes such as leaching and erosion occur with or without human interference, but for these processes to be described as "degradation" imply social criteria which relate land to its actual or possible uses (Blaikie and Brookfield, 1987). The World Map of the Status of Human-induced Soil Degradation (Oldeman, Hakkeling and Sombroek, 1991; Oldeman, 1993) has revealed that almost 2000 million hectares of agricultural land, permanent pastures and forest and woodland has been degraded. Worldwide almost 40% of the agricultural land is affected. Some 900 million hectares have a moderate degree of soil degradation, indicating a greatly reduced productivity.

The need for a georeferenced socio-economic database is evident. The issue is, whether a georeferenced socio-economic database can be directly linked to a georeferenced agro-ecological database. It is possible to link socio-economic attributes to the physiographic mapped units of the agro-ecological database. Socio-economic information is generally assembled for administrative areas, which generally do not conform to physiographic delineations. Socio-economic attributes do change over time (as land use/land cover attributes do).

A holistic approach to agro-ecological characterization is obvious. Stoops and Cheverry (1992): "The task of all people concerned with land and water resources is to direct their interest not just to the physical, chemical and biological aspects, but also to those environmental, economic, social, legal and technical aspects that affect soil use". Life, and livelihood on earth depend largely on the capacity of the soil to produce, whether by agricultural, industrial or other practices. The soil is a natural resource, non renewable in the short term or very difficult to renew, and expensive either to reclaim or to improve following erosion, physical degradation or chemical depletion. It is our duty to maintain it for the future, as well as to obtain the best benefit from its use today.

REFERENCES

Blaikie P., and H. Brookfield, 1987. Land Degradation and Society. Methuen, London and New York.

Brundtland, G.H., M. Khalid, a.o. 1987. Our Common Future. Report of World Commission on Environment and Development presented to the chairman of Intergovernmental Intersessional Preparatory Committee, UNEP Governing Council. Oxford University Press, Oxford.

Buringh, P. and Dudal R., 1987. Agricultural Land Use in Space and Time. In: Wolman M.G. and Fournier F.G.A. (Editors), Land Transformation in Agriculture. SCOPE 32, John Wiley & Sons, Chichester, New York, pp. 9-43.

FAO, 1988. Soil map of the world, Revised legend. World Soil Resources Report 60, FAO, Rome.

ISNAR, 1992. Summary of Agricultural Research Policy. International Quantitative Perspectives, ISNAR, The Hague, 24 p.

ISSS, 1986. Proceedings of an international workshop on the structure of a digital international soil resources map annex database. Ed. by M.F. Baumgardner and L.R. Oldeman. SOTER report 1, ISSS, Wageningen, 138p.

Oldeman, L.R., 1993. Global Extent of Soil Degradation. In Bi-Annual Report 1991-1992, International Soil Reference and Information Centre, Wageningen, p. 19-36.

Oldeman, L.R., R.T.A. Hakkeling and W.G. Sombroek, 1990. World Map of the Status of Human-induced Soil Degradation: An explanatory note. Wageningen, International Soil Reference and Information Centre; Nairobi, United Nations Environment Programme. 27 pp + 3 maps.

Pulles, J.H.M., 1988. A Model for a Soils and Terrain Digital Database. Working Paper and Preprint No. 88/5, ISRIC, Wageningen.

Purnell, M.F., 1992. Offer and Demand of Soil Information: International Policies and Stimulation Programmes. In: Soil Survey, Perspectives and Strategies for the 21st Century; Proceedings of an International Workshop for Heads of National Soil Survey Organizations (Keynote speeches). International Institute for Aerospace Survey and Earth Sciences (ITC), Enschede. 9 p.

Remmelzwaal, A., 1990. Classification of land and land use, first approach. FAO, Rome (unpublished).

Sombroek, W.G., 1992. Some thoughts of the new director, AGL, Land and Water, no.36, FAO, Rome, 24 p.

UNEP, 1992. Two decades of achievement and challenge. The United Nations Environment Programme 20th Anniversary. Our Planet, Vol.4, no.5, Nairobi, 23 p.

Unesco, 1973. International classification and mapping of vegetation. Ecology conservation 6. Unesco, Paris.

UNSO/UNDP, UNEP, 1992. Draft report on the Expert Meeting on Desertification, Land Degradation and the Global Environment Facility, Nairobi, 28-30 Nov. 1992.

Unesco, 1973. International classification and mapping of vegetation. Ecology conservation 6. Unesco, Paris.

Van Engelen, V.W.P., and Wen Ting-Tiang, 1993. Global and National Soil and Terrain Digital Databases (SOTER), Procedures Manual. FAO, ISRIC, ISSS, UNEP, Wageningen, 115 p.

WRI, Cal Tech, 1992. Global Environmental Monitoring: Pathways to Responsible Planetary Management. A proposal by the World Resources Institute, Washington DC, and the California Institute of Technology, Pasadena, Washington, 17 p.