5.9 INFORMATION EXTRACTION FROM A SOILS AND TERRAIN DIGITAL DATABASE

Carlos R. Valenzuela
International Institute for Aerospace Survey and Earth Sciences (ITC)
and Marion F. Baumgardner
Purdue University

5.9.1. Introduction.

The complexity and increasing volumes of available information, and the demand for the storage, analysis and display of large quantities of environmental data, has led in recent years to rapid development in the application of computers to environmental and natural resources data handling, and the creation of sophisticated information systems (Tomlinson et al., 1976). Increasingly data of all types are being collected and converted to digital format. Extensive digital geographically oriented databases are being developed, and automated spatial information systems are used for storage, retrieval, manipulation, analysis and display of information (Tom and Miller, 1975; Knapp, 1978; Jerie et al., 1980; Marble and Peuquet, 1983). Effective utilization of large spatial data volumes is dependant upon the existence of an efficient geographic handling and processing system that will transform these data into usable information. The major tool for handling spatial data is the geographic information system (Marble and Peuquet, 1983).

A digital geographic information system (GIS), is a computerized system designed to store, process and analyse spatial data and their corresponding attribute information. Advances in computer technology and techniques have made it possible to integrate a wide range of information (Gribbs, 1984). Technological advances have increased input techniques, storage, analysis and retrieval capabilities. Furthermore, there has been a reduction in costs and an increase in accessibility, so that a larger user community has developed (Moellering, 1982). Geographic information systems have provided planners with a readily accessible source of objective earth-science-related facts and an inexpensive, rapid and flexible tool for combining these facts with various other products to create decision alternatives (van Driel, 1975; Stow and Estes, 1981; Stoner, 1982).

Basic information on the location, quantity and availability of natural resources is indispensable for planning more rationally their development, use and/or conservation. The demand for specific, accurate, and rapid soil information is growing in our modern society. Soils, because of their importance in agricultural and non-agricultural matters, and their inherent relationships with other environmental resources are a basic and fundamental component of any complete geographic information system. Johnson (1975) points out that the conventional preparation of soil interpretive maps combining information of the soil resource with other resource information are excessively expensive, especially if various source maps have to be converted to a common scale, and if the interpretive requirements are complex. Automatic data processing systems have created immense opportunities for storing and disseminating soil data (Bertelli, 1979). As the demand for interpretive maps increases, computers are used to speed up and cut down costs of the processes (Bertelli, 1966; Shields, 1976; Bertelli, 1979; Bie, 1980; Valenzuela, 1985). The national Soil Handbook of the USDA Soil Conservation Service (1983), indicates that computer generated interpretive maps are encouraged where the soil survey has been digitized, because they cost less than maps perpared by other means.

5.9.2 Methodology.

The digital geographic information system used for this study basically consists of five major subsystems: 1.- Input subsystem; 2.- Database subsystem; 3.- Management subsystem; 4.- Modelling and analysis subsystem, and 5.- Output subsystem.

The soil association map of Indiana, U.S.A., was prepared by the Indiana soil survey staff of the United States Department of Agriculture Soil Conservation Service and Purdue University Agricultural Experiment Station and made available to users as publication AY 209, at a scale of 1:500,000 in 1980.

The soil association map was digitized using the Purdue/LARS digitizing system. The system is composed of a Talos table digitizer and an APPLE II Plus microcomputer. A complete documentation of this menu-driven system was prepared by Phillips (1983). The data capture consisted of the transformation of three (3) map primitives, i.e., control points, boundaries (limits of soil units), and centroids, into a format compatible with digital computers. After the process of data capture was completed, the computer compatible data were transferred from the APPLE II Plus microprocessor to the host (main) computer (IBM 360/158), where the data were stored and the activities of editing, coordinate transformation and rasterization were performed. Editing the data was accomplished by manual and automatic editing routines using a Tektronics 4054 graphics terminal. Twelve (12) control points were used to derive statistically a biquadratic regression model required to transform the digitized X and Y values into longitude and latitude geographic coordinates. These data were subsequently transformed into an Albers equal-area cartographic projection.

The final step in the map input procedure was the rasterization process. During this process, the boundary and centroid files stored in addresses corresponding to the Albers cartographic projection, were converted into an image file. The map units were filled-in cells according to a predefined grid (500 m x 500 m on the ground, or, 1mm x 1 mm on the map), and subsequently each cell was assigned with a class code (0 to 255) associated with the centroid file. The codes (fill characters) assigned to each of the 55 soil associations present in the map and to the portion of Lake Michigan in Indiana are shown in Table 5.9.1.

For the construction of the attribute database (hierarchical), extensive use was made of the available information generated for the state soil associations of Indiana (Galloway et al., 1975). Other information not readily available in tables or maps, were obtained by generalization of the information present in the description of the soil series forming each soil association (Galloway and Steinhardt, 1981; Franzmeier and Sinclair, 1982). Information generated by visual and digital interpretations of LANDSAT data from Indiana were also used. Figure 5.9.1 illustrates a LANDSAT mosaic covering the whole state.

For display purposes and generation of color outputs of the computer generated interpretive soil maps, the rasterized image file was transfered to the image processing device IBM 7350 (HACIENDA).

5.9.3 Results.

Once the input of the data is completed and the rasterized data set and their attribute information are stored in the database, this spatial information can be easily retrieved, handled, analysed and displayed. The degree of the analytical capabilities implemented in a system depends on the

nature, purpose and objectives of the principal user. However, a well thought-out system will be one that is flexible enough to respond to the needs for input, analysis and display of different kinds of data required by the user.

In almost every digital geographic information system currently operational, one element appears to be present to satisfy the requirements of the main user. It is the element soils, depicting soil types as obtained from soil surveys. This is the result of its relation to the fauna, vegetation and climate, and its strong interaction with other natural resources elements. The nature and type of information available from a soil survey enables the generation of several interpretive soil maps. These maps can be used as new variables for analysis or modelling of resources to predict changes that may occur through time.

The soil association of Indiana in digital format displayed in the High Level Image Processing System (HLIPS) device IBM 7350 are shown in Figure 5.9.2. The area estimates and percentage of occurance of each soil association in Indiana are presented in Table 5.9.1. Soil association Crosby-Brookston present on nearly level surfaces of Wisconsin age glacial till plains in Central Indiana, constitutes the largest association covering an area of approximately 703,050 ha or 7.4% of the state. Figure 5.9.3 illustrates the parent materials from which Indiana soils were developed. It depicts the various kinds of materials, including old sedimentary rocks in the southern part of the state; different thickness of loess deposits over glacial till; alluvial, lacustrine and eolian deposits from which the soils were developed.

The potential soil erosion was calculated using the Universal Soil Loss Equation (USLE). The factors of the USLE for each soil association were estimated by Brentlinger et al. (1979). This information was used to reclassify the digital soil association map into four (4) potential soil erosion groups; Low, Medium, High and Very High. The potential soil erosion map is illustrated in Figure 5.9.4. This interpretive information can be used in conjunction with landuse/landcover data to predict the erosion hazard or gross erosion in the state. It can also be related to slope, landuse and proximity to streams to determine agricultural pollution due to erosion and to estimate sedimentation hazards, and the related dangers of floodings. The dominant drainage, as determined by the characteristics of the soil associations, is shown in Figure 5.9.5. This information is very important for the planning, design and construction of septic systems in the state.

Soil maps in Indiana are used in the assessment of agricultural potential. The basic aim of any assessment activity is the equal treatment of all individual land owners. Yahner (1979) described the procedures employed in agricultural land reassessment using stimates of corn yields. Each soil association has been assigned an estimated corn yield value. Figure 5.9.6 illustrates the corn yield estimate map of Indiana. Table 5.9.2 presents the corn yield values for each soil association. Because of the resolution (scale) of the data and the generalization involved in the creation of the soil associations, some problems and difficulties exist in the assessment of individual farms. However, it can be used to obtain rapid information on the approximate value of agricultural land. Figure 5.9.7 illustrates the soil associations depicting organic matter levels for the state of Indiana. This figure shows weighted averages of organic matter content, which were calculated taking into consideration the relative occurrence of soil series forming the soil associations, as tabulated by Galloway et al. (1985).

- Figure 5.9.8 shows the digital county map of Indiana. This information was combined with the digital soil information and it was used to calculate the areas of soil associations present in each county, which is very important for planning the allocation of resources at county level. Table 5.9.3 includes the soil associations and their corresponding areas for some counties in the state.
- Table 5.9.4 presents all the soil associations in the state with their respective information used to generate the interpretive maps and tabular information. The production of digital interpretive maps does not involve any changes in the original data set. It uses atribute files to regroup the original units into interpretive classes in real time. These new data (interpretive) can be used in analysis and modelling with other data sets available in the database. Figure 5.9.9 illustrates the county boundaries and the soil association original data sets, and the interpretive information generated using the analysis capabilities of the GIS.

5.9.4 References.

- Bertelli, L.J., 1966. General soil maps: a study of landscapes. Journal of soil sciences and water conservation, 21(1).
- Bertelli, J.L., 1979. Interpreting soil data: planning the use of management of land. ASA-CSSA-SSSA, 677 South Segoe Road, Md.
- Bie, S.W., 1980. Computer assisted soil mapping: the computer in contemporary cartography. Edited by D.R.F. Taylor, John Wiley and sons Ltd.
- Brentlinger, J.L., J.B. Hays, R.L. Lauster and C.S. Pierce, 1979.
 Agricultural assessment for the non-designated and designated 208
 planning areas of Indiana. State soil and water conservation committee,
 Indiana Department of Natural Resources.
- Driel van, N., 1975. Geological information in a computer mapping system for landuse planning. Proceedings of the international symposium on computer assisted cartography. Bureau of the census and ACSM, Auto Carto II.
- Franzmeier, D.P., and H.R. Sinclair, 1982. Key to soils in Indiana: showing the relationships of soils to each other and their environment. AY-249, Cooperative Extension Service, Purdue University, West Lafayette, Indiana
- Galloway, H.M., J. Yahner, G. Srinivasan and D. Franzmeier, 1975. User's guide to the general soils map and interpretive data for counties in Indiana. Supplement to AY-50 series. Cooperative Extension Service, Purdue University, West Lafayette, Indiana.
- Galloway, H.M. and G.C. Steinhardt, 1981. Indiana's soil series and their properties. AY-212, Cooperative Extension Service, Purdue University, West Lafayette, Indiana.
- Gribbs, W.J., 1984. The combining of geographic and other information systems to create a new type of information system. Technical papers of the 44th annual meeting of the American congress on surveying and mapping. ASP/ACSM convention, Washington DC.
- Jerie, H.G., J. Kure and H.K. Larsen, 1980. A system approach to improving geo-information systems. ITC journal 1980-4, Enschede, the Netherlands.

- Johnson, C.G., 1975. The role of automated cartography in soil survey. Proceedings of the meeting of the ISSS working group in soil information systems, Wageningen, the Netherlands.
- Knapp, E., 1978. Landsat and ancillary data inputs to an automated geographic information system: application for urbanized area delineation. Goddard Space Flight Center and Computer Science Corporation.
- Marble, D.F., and D.J. Peuquet, 1983. Geographic information systems and remote sensing. Manual of remote sensing, vol 1, chapter 22.
- Moellering, H., 1982. The challenge of developing a set of national digital cartographic data standards for the United States. Proceedings ACSM/ASP, spring meeting, Denver.
- Phillips, T.L., 1983. Operation of the Purdue/Lars digitizing system. Laboratory or Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.
- Shields, R.L., 1976. New generalized soil maps guide for land use in Maryland. Journal of soil and water conservation, vol 32, No.6.
- Stoner. E.R., 1982. Agricultural land cover mapping in the context of a geographically referenced digital information system. Report No 205, NASA, Space Technology Laboratories, Earth Resources Laboratory, NSTL Station.
- Tom, C. and L. Miller, 1974. A review of computer-based resource information systems. Land use planning information report 2, part 2, Colorado State University, Fort Collins, Colorado.
- Tomlinson, R.F., H.W. Calkins and D.F. Marble, 1976. Computer handling of geographical data. The Unesco press.
- USDA, 1983. National soils handbook. USDA, Soil Conservation Service. U.S. Government Printing Office, Washington DC.
- Valenzuela C.R., 1985. The element soil as a basic component in a digital geographic information system. Ph.D. Thesis, Agronomy Department, Purdue University, West Lafayette, Indiana.
- Yahner, J.E., 1979. Use of soil maps in Indiana's farmland reassessment. Agronomy guide AY-216, Cooperative Extension Service, Purdue University, West Lafayette, Indiana.

Table 5.9.1 Area estimates and percentage of occurence of soil associations in Indiana.

CODE SYMBOL	SOIL ASSOCIATIONS	AREA IN HA	PERCENT
123456789012312312345126H123451231234567121231234P123123123456789012345000000000000000000000000000000000000	GENESSE-EEL-SHOALS FDX-GENESSE-EEL SLOAN-ROSS-VINCENNES-ZIPP STENDAL-HAYMOND-WAKELAND-NOLIN WHEDLING-HUNTINGTON-LINDSIDE HOUGHTON-ADRIAN MAUMEE-GILFORD-SEBEWA RENSSELAER-DARROCH-WHITAKER SEBEWA-GILFORD-HOMER LYLES-AYRSHIRE-PRINCETON MILFORD-BONO-RENSSELAER PATTON-LYLES-HENSHAW ZIPP-MARKLAND-MCGARY TRACY-DOOR-LYDICK ELSTON-SHIPSHE-WARSAW OSHTEMO-FOX FOX-OCKLEY-WESTLAND PARKE-NEGLEY OAKVILLE-ADRIAN PLAINFIELD-MAUMEE-OSHTEMO PRINCETON-BLOOMFIELD-AYRSHIRE ALFORD RAGSDALE-RAUB SABLE-IPAVA FINCASTLE-RAGSDALE REESVILLE-RAGSDALE RESVILLE-RAGSDALE RESVILLE-NAPPANEE PARR-BROOKSTON BLOUNT-PEWAMO HOYTVILLE-RAGSDALE RIMMI-CROSIER-BROOKSTON MIAMI-CROSIER-BROOKSTON MIAMI-CROSIER-BROOKSTON MIAMI-RUSSEL-FINCASTLE-RAGSDALE RUSSEL-HENNEPIN-FINCASTLE RARKHAM-ELLIOT-PEWAMO BARTLE-PEDGA-DUBOIS WEINBACH-WHEELING AVONBURG-CLERMONT HOSMER LE-PEDGA-DUBOIS WEINBACH-WHEELSTON-TILSIT CINCINNATI-VIGO-AVA CINCINNATI-VIGO-AVA CINCINNATI-ROSSMOYNE WELLSTON-ZANESVILLE-BERKS CRIDER-BASTER-CORYDON BERKS-GILPIN-WEIKERT CORYDON-WEIKERT-BERKS EDEN-SWITZERLAND LAKE MICHIGAN	202, 050 187, 225 406, 000 65, 800 70, 450 333, 350 131, 150 233, 930 331, 150 231, 930 331, 150 231, 930 331, 930 331, 930 331, 930 331, 930 331, 930 331, 930 331, 930 231, 930 231, 930 231, 930 231, 930 231, 930 231, 930 273, 700 173, 700 173, 700 173, 970 173,	387394192654062103193388884643160849067151663120577076066 2104003100000000000110001100117501045050060011124403010

Table 5.9.2 Weighted average corn yield estimates for soil associations in Indiana.

SOIL ASSOCIATIONS	MAP SYMBOL	CODE	CORN#
GENESSE-EEL-SHOALS FOX-GENESEE-EEL SLOAN-ROSS-VINCENNES-ZIPP STENDAL-HAYMOND-WAKELAND-NOLIN WHEELING-HUNTINGTON-LINDSIDE HOUGHTON-ADRIAN MAUMEE-GILFORD-SEBEWA RENSSELAER-DARROCH-WHITAKER SEBEWA-GILFORD-HOMER LYLES-AYRSHIRE-PRINCETON MILFORD-BONO-RENSSELAER PATTON-LYLES-HENSHAW ZIPP-MARKLAND-MCGARY TRACY-DOOR-LYDICK ELSTON-SHIPSHE-WARSAW OSHTEMO-FOX FOX-OCKLEY-WESTLAND PARKE-NEGLEY OAKVILLE-ADRIAN PLAINFIELD-MAUMEE-OSHTEMO PRINCETON-BLOOMFIELD-AYRSHIRE ALFORD RAGSDALE-RAUB SABLE-IPAVA FINCASTLE-RAGSDALE REESVILLE-RAGSDALE IVA-VIGO BROOKSTON-ODELL-CORWIN CROSSIER-BROOKSTON BLOUNT-PEWAMO HOYTVILLE-NAPPANEE PARR-BROOKSTON RIDDLES-TRACY-CHELSEA MIAMI-CROSSY-BROOKSTON RIDDLES-TRACY-CHELSEA MIAMI-CROSSY-BROOKSTON RIDDLES-TRACY-CHELSEA MIAMI-CROSSY-BROOKSTON RIDDLES-TRACY-CHELSEA MIAMI-CROSSY-BROOKSTON MIAMI-HENNEPIN-CROSBY MIAMI-CROSSY-BROOKSTON MIAMI-HENNEPIN-CROSBY MIAMI-RUSSELL-FINCASTLE-RAGSDALE RUSSELL-HENNEPIN-FINCASTLE WARKHAM-ELLIOTT-PEWAMO BARTLE-PEOGA-DUBOIS WEINBACH-WHEELING AVONBURG-CLERMONT HOSMER ZANESVILLE-WELLSTON-TILSIT CINCINNATI-VIGO-AVA	123451231231231234512GH12345123121234	12345678901200000000000000000000000000000000000	1154941209114120000000000000000000000000000
EDEN-SWITZERLAND LAKE MICHIGAN	R3	55 60	022

^{* =} Bu/acre

Table 5.9.3 Soil associations and area estimates for some counties in Indiana.

COUNTY	CODE	SYMBOL	SOIL ASSOCIATIONS	AREA IN HA
BENTON	4			
	8 33 36	C1 J1 L1 L4	RENSSELAER-DARROCH-WHITAKER BROOKSTON-ODELL-CORWIN PARR-BROOKSTON MIAMI-CROSBY-BROOKSTON	10,500 68,725 24,225 2,700
CARROL	8			
	1 17 25 30 35 36 37 38	A1 E4 I3 J3 L3 L4 L5 L6	GENESSE-EEL-SHOALS FOX-OCKLEY-WESTLAND FINCASTLE-RAGSDALE CROSBY-BROOKSTON MIAMI-CROSIER-BROOKSTON-RIDDLES MIAMI-CROSBY-BROOKSTON MIAMI-HENNEPIN-CROSBY MIAMI-RUSSELL-FINCASTLE-RAGSDALE	8,650 18,625 33,800 975 175 25 25 34,125
CLINTON	12			
	23 25 30 36 38	I1 I3 J3 L4 L6	RAGSDALE-RAUB FINCASTLE-RAGSDALE CROSBY-BROOKSTON MIAMI-CROSBY-BROOKSTON MIAMI-RUSSELL-FINCASTLE-RAGSDALE	8,100 17,225 28,525 46,600 4,100
POUNTAIN	24			
	1 15 17 23 26 37 39	A1 E2 E4 I1 I4 L7	GENESSE-EEL-SHOALS ELSTON-SHIPSHE-WARSAW FOX-OCKLEY-WESTLAND RAGSDALE-RAUB RESVILLE-RAGSDALE MIAMI-HENNEPIN-CROSBY RUSSELL-HENNEPIN-FINCASTLE	5,650 10,600 23,825 7,850 17,375 25 41,625
JASPER .	37			
	1 6 7 80 28 33	A1 B1 B2 C1 F2 J1 L1	GENESSE-EEL-SHOALS HOUGHTON-ADRIAN MAUMEE-GILFORD-SEBEWA RENSSELAER-DARROCH WHITAKER PLAINFIELD-MAUMES-OSHTEMO BROCKSTON-ODELL-CORWIN PARR-BROOKSTON	200 6, 100 59, 900 35, 125 15, 200 10, 350 19, 000
MONTGOMERY	54			
	1735636789 99003	E4 I1 I3 I4 L1 L4 L5 L5 L7	FOX-OCKLEY-WESTLAND RAGSDALE-RAUB FINCASTLE-RAGSDALE REESVILLE-RAGSDALE PARR-BROOKSTON MIAMI-CROSBY-BROOKSTON MIAMI-HENNEPIN-CROSBY MIAMI-RUSSELL-FINCASTLE-RAGSDALE RUSSELL-HENNEPIN-FINCASTLE	18,750 6,750 19,600 2,950 3,325 5,475 7,825 64,225 1,650

Table 5.9.4 Soil associations and corresponding interpretive information in the State of Indiana.

CODE SYMBOL	AREA IN HA	EROSION	SLOPE	DRAINAGE	FARM	CORN	<u>%О</u> . <u>М</u> .
12345678901231231234512GH12345123123456789012345000000000000000000000000000000000000	202, 050 187, 8225 400, 800 065, 800 065, 800 067, 8150 107, 400 107,	111121111111110111101111001111111111111	11111111111111111111111111111111111111	114114444447111111111446777467779411711717777777777		######################################	NNN-N4NDN-NDNNN-ND4NNDNNDND-NNDN-NN
where Erosion Slope Drainage Prime farm Corn yield Org. Matter	1(nearly 1: 4(hilly) 1(well), 2 4(noorly)	medium), evel), 2((moderate	ely wel	1); 4(vert sting), 3 (1), 3(son (25 %) 1), 4(vert (2, 6-3, 5%)	newhat	; рост	·1y),



Landsat mosaic FIGURE 5.9.1

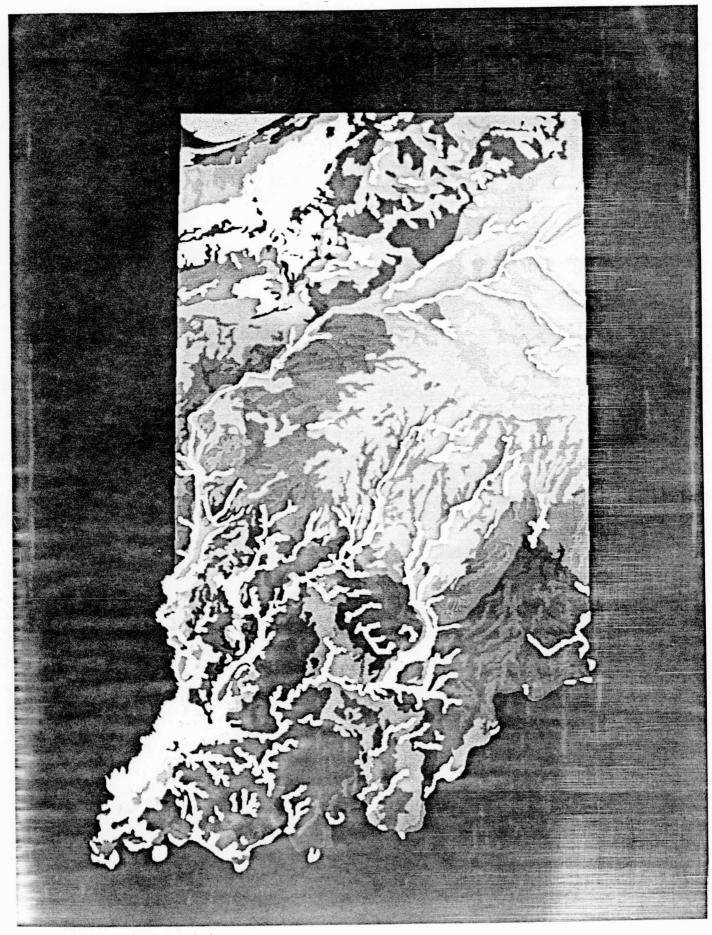


FIGURE 5.9.2 Soil associations



FIGURE 5.9.3 Parent materials



FIGURE 5.9.4 Potential erosion hazard



FIGURE 5.9.5 Dominant drainage

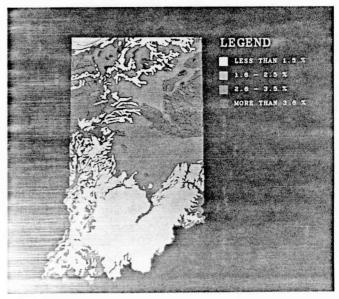


FIGURE 5.9.6 Organic matter content



FIGURE 5.9.7 Estimated maize yields



FIGURE 5.9.8 Administrative (county) units

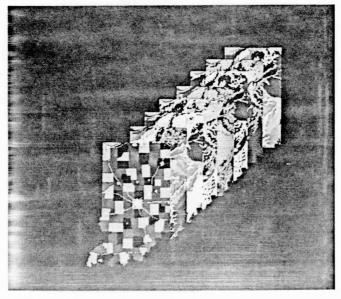


FIGURE 5.9.9 Image database