Appendix A *Readings*

The following is a paper by Ian K. Crain which appeared in the Auto-Carto Proceedings for 1989. It does an excellent job of explaining the development of Geographic Information Systems in Canada in the past two decades. Canada has been a pioneer of GIS in North America and offers much in the way of experience to anyone interested in developing a GIS.

FROM LAND INVENTORY TO LAND MANAGEMENT
THE EVOLUTION OF AN OPERATIONAL GIS

I.K. Crain
and
C.L. MacDonald
Canada Land Data Systems
Environment Canada, Ottawa, Ontario K1A 0E7

ABSTRACT

Virtually all long-lived information systems, whether spatial or not, proceed (by intent or by unconscious incremental change) through a characteristic evolution from a simple data bank - an inventory tool, to incorporate complex retrieval and statistical processing - an analysis tool, and finally to a system involving modelling and complex decision support capabilities - a management tool. This evolution is evident and still going on in the world's oldest operational geographic information system - the Canada Land Data System (CLDS). The system began with the specific intent of serving the needs of the Canada Land Inventory - to store very extensive and fairly detailed data on the land use potential of Canada. This initial goal having been achieved, increasing demands were placed on the system to explore data and spatial relationships within this huge data bank, and to provide insight for land resource scientists into more localized and specialized concerns - in short, an analysis capability. The CLDS is now entering the third phase of evolution to meet the inevitable demands as a land planning and management tool. The implications of this are mainly: increased user access and interactivity with existing data, forecasting, planning and modelling capabilities, enhanced variety of display outputs, ability to service non-expert users locally and interactively for small-area data sets in a land use planning setting.

INTRODUCTION

Information systems, regardless of their field of application, have certain elements in common - a large central bank of data, facilities for manipulating, retrieving, updating, and reporting of data. The common principles and capabilities have been extensively studied and reported in the literature and will not be re-iterated here. Another commonality amongst information systems, again independent of application - whether it is a personnel system, financial system, management system, is a tendency to progress through a characteristic evolution which can be conveniently divided into three stages. These stages are typical of and perhaps necessary to a successful information system. Systems die, or are murdered, mainly for their inability to evolve in this pattern. In this generalized sense, geographic information systems are no different, and to succeed must be capable of following the same evolutionary pattern.

TABLE 1

PHASE	SYSTEM CAPABILITIES	PRIMARY ACTIVITY	USER/SUPPLIER RELATIONSHIP
INVENTORY	-data input -editing -simple retrieval -routine reporting	- data input	-clear division between client and supplier -little interaction
ANALYSIS	all above plus: -complex retrievals -ad hoc queries -statistical processing -derived reporting (eg. graphics) -derived data sets	-data retrieval and manipulation	-supplier involved in determining output needs -interactive retrievals and direct data access by user
MANAGE	all above plus: -modelling, simulation -decision support tools	-data retrieval and manipulation	-user and supplier indistinguishable -fully interactive

This paper describes the principal stages of the evolutionary process and elaborates them in terms of a particular system (or group of systems) known as the Canada Land Data System (CLDS), which contains the Canada Geographic Information System (CGIS). This system has GIS and has in that period undergone considerable evolution and adaptation, so that it continues to be a highly successful broad-scope GIS.

CHARACTERISTIC STAGES OF INFORMATION SYSTEM EVOLUTION

The stages as outline are essentially arbitrary and dividing lines between them are necessarily fuzzy. It is recognized as well that successful systems probably exist which have not followed this pattern for one reason or another. Such counter-examples should not detract from the general principle described. The three characteristic stages are named (again arbitrarily) by their principal functional purpose as Inventory, Analysis and Management. The main characteristics of the phases are outline in (Table 1).

As the system functionally evolves, the organizational relationship between the user and supplier of system services must change as well. The changes in this environment parallel to some extent the general evolutionary pattern of all EDP organizations as defined by (Nolan 1979).

Phase I: Inventory

In this initial phase of an information system, the reason for its existence or development is to assemble, organize and determine the extent of existence of data on a particular subject. In a personnel system for example, the initial thrust is normally to develop a file of basic common information about each employee. At this phase the system serves to answer data queries of how much?, how many?, where?, with an emphasis on reports which derive directly from the data in the field based on summations, counts and other minor manipulations of the raw data. In short, an inventory function.

The principal activities are data collection, input, and editing - in general those activities which ensure the development of an accurate, high integrity base of primary data.

For a GIS this represents a system whose primary task is the development of a broad integrated collection of mapped or geocoded data which may encompass a broad area, but of very well defined scope, e.g. all the forestry maps for a province, with capabilities to answer direct queries on area by selected criteria, or to redisplay the data in various ways.

During this phase, the relationship between the user and the supplier of system services tends to be simple and consist of a well defined separation of function - a classic customer/client relationship. System functions are well known and relatively simple, the user “menu” is well defined. The supplier’s role is one of ensuring the integrity of the data, ensuring on-going operational continuity of software and hardware and providing standard (albeit complex) reports and outputs routinely or on demand.

Phase II - Analysis

The second phase of the evolution is spawned by the user’s desire to make more use of the data than direct tabulations and summaries - the desire to explore data relationships in order to shed light on subject matter problems, to help confirm hypotheses or to provide data for research and modelling. In the personnel system example, the demand might be to examine relationships between occupational level and place of residence, or to search for evidence of relationships between salary and ethnic origin, etc.

The emphasis in this phase is much more on complex retrievals, and queries which may generate additional queries in an unexpected or unstructured way, hence the implications of a need for extensive user interaction. Advanced statistical analysis tools are required and the queries tend to relate much more indirectly to the raw data.

The queries, in fact, become more like those typical of a “scientific” system (Crain 1978) in that they do not relate directly to the record (“tell me all about employee number 1234”), but rather span across records seeking relationships (“tell me the age distribution and correlation with salary of female employees living in the city center”).

The principal activity moves from data gathering to data retrieval and manipulation. New data inputs are less likely to be broad and comprehensive, but will be more specialized to focus on particular subject matter concerns.

In the case of a GIS, the primary requirement is for user interaction, usually graphically, with the data - the ability to sort, select, derive new data from old, extract and re-display data on the basis of complex geographic, topological and statistical criteria. New data collections may still represent very high data volumes, but cover smaller areas in more detail.

Ap[art from the obvious need for the system supplier to provide interactive services, the supplier of an analysis system must evolve organizationally to become more involved in subject matter and data context issues. In order to develop the necessary system tools to meet user requirements, considerable emphasis must be placed on determining and anticipating user information needs in the short and long term, and an understanding of the existing and evolving data analysis techniques of subject areas. While most data will still be held centrally, there is a growing decentralization of user access and associated hardware.

Phase III - Management

A true management information system must provide the tools to assist directly in the decision making process - not merely an inventory of what exists, not only analyses of data relationships which might hint at problems or solution. The essential new ingredient for a system to evolve to this phase is the addition of forecasting and planning facilities - the ability for the user to ask "What if....". This implies the need for more advanced mathematical forecasting capabilities. To extend the personnel system example, a personnel management system might for instance, be able to model the process of advancement of minority groups, allowing the manager to forecast potential impact of various strategies for improving their advancement opportunity. A land management GIS might provide answers to "what if?" questions related to alternated land use policies or various strategies for optimization of land use under conflicting demand. It is with these modelling and planning tools, superimposed on a broad base of hard raw data that a GIS reaches maturity as a resource management system.

When a system reaches this point in evolution, the distinction between user and supplier virtually disappears. With the exception of some purely technical EDP staff (such as machine operators and software people) all system "users" must be integrated into an overall system management process Decentralization of data holdings as well as data access will occur and necessary integrated controls must be developed. Knowledge and skill requirements of those involved cannot be divided on the basis of "computer" or "subject matter" and must be divided instead on functional and managerial lines. This change in relationship is by far the most difficult step in the evolution and the greatest barrier to integrating an MIS or GIS truly into the decision process.

EVOLUTION OF THE CANADA LAND DATA SYSTEM

The assembly of systems now known as the Canada Land Data System has origins that go back more than twenty year to the "ARDA project" of 1962. During that long period, the system has evolved in the characteristic pattern, previously described and I would like to use the CLDS to exemplify the process.

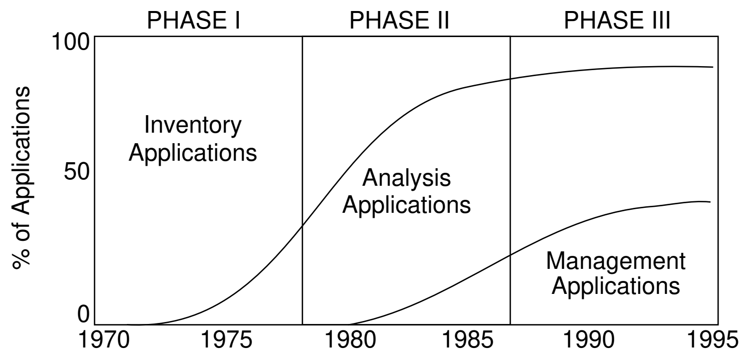
The original purpose for the system (which has had several names, the most well known of which is the Canada Geographic Information System - CGIS) was to store the Canada Land Inventory (CLI). The actual system was delivered in 1968 (now celebrated as the birthdate for CLDS) and full production operation began in 1971. As implied, it commenced life as an inventory system. The initial data collection was of five coverages of land capability for all the arable land of Canada - an enormous area of some 2.7 million Km2.

This required the input of approximately 1500 detailed polygon maps, a process which was essentially complete by 1979. These "early days" were concerned by data entry and routine tabular reporting. Raw data tabulations and summary reports were produced on a national and regional basis in standard formats. While considerable topological processing was built into the system - e.g. ability to do multicoverage overlays, reporting and derived graphics facilities were quite limited.

Somewhere in the interval between the commencement of production operations in 1971 and the completion of the Canada Land Inventory in 1979 the system ceased to be solely an inventory system and evolved into an analysis facility. one clear symptom of this was the introduction in 1974 of an interactive graphics subsystem. Users demanded increased ability to relate variables, to select and study particular areas, and to extract derived data for planning, prioritizing and problem analysis.

The subject matter focus changed from a national resource inventory perspective to more specific issues such as arctic land use, coastal zones, national parks, etc. By 1978 more maps of this regional nature were being input than CLI maps. (Figure 1) schematically outlines this evolution. The changes required to bring the system to Phase II include advance statistical program packages, interactive graphics enquiry facilities through a network of terminals across Canada to provide user access, and improved hard-copy graphics output capability.

Figure 1



Today the system stands fully in Phase II even though data holdings (inventories) still grow at a rate of 1000 maps per year, and is on threshold of Phase III.

Over 350 interactive data bases of complexly overlayed map data are available for user analysis using various systems and subsystems which collectively make up CLDS.

In spite of this, strong demand exists for CLDS to become an integrated land management system and the first steps in this evolutionary phase have begun. Several small applications have used the existing facilities for land planning and management purposes - for camp sit location in National Parks and transportation corridor planning the Beaufort Sea area. Within five years, with suitable system changes, the CLDS will be fully into Phase III.

To accomplish this many new system functions are needed and organizational relationships must change. Land planning models and accompanying facilities to define model parameters from the data bases, and forecast quantitative and qualitative impact must be developed.

At this time, testing is underway of simple linear models which allow for the subjective weighing of land use suitability in order to evaluate the impact of variable land policy options. Data input facilities are being upgraded to reduce the time required to make data available to planners - a necessary step if rapidly changing conditions are to be monitored by resource planners.

The demand is growing for increased local processing capability for local land use management, with continued access to the central regional and national data banks. Preliminary trials are underway to equip mini and micro computers for this purpose.

CHALLENGES TO ACHIEVING PHASE III

The transition to Phase III is not as easy as the movement from inventory to analysis. Many geographic information systems are now facing this transition and face similar challenges. While placed in the specific context of CLDS, these barriers are of general applicability.

1) Resource constraints: The resources (both dollars and human) required to enhance and distribute the processing power as needed in Phase III are substantial. In today's cautious economic climate, especially in government, major one time injections of resources for system enhancement are unlikely. Thus the change will have to be made gradually as funds become available. Software development is always very expensive and time consuming, and user demands will always exceed the capacity of software development resources.

2) Ensuring continual quality control: Quality control of outputs and maintenance of the integrity of the data bases are increasingly difficult to achieve as the system becomes decentralized. Where the distinction between users and supplier becomes blurred, as it will in Phase III, responsibility for ensuring the correctness of data input, its consistency with global quality and subject matter standards, and its protection from corruption become an administrative issue rather than a technical issue. The degree of collective rather than individual responsibility for this incredibly valuable asset increases greatly and is difficult to achieve.

3) Choosing appropriate new technology: This is the constant technical challenge of any evolving system - how to take advantage of the latest technological advance without extreme risk or without jeopardizing future alternatives. It requires constant research and a balance between conservatism - selecting only the proven, and opportunism - selecting the correct new (unproven) machine or software where the risks seem worth it.

4) Ensuring continuity of production operation during change: This is both a technical and administrative challenge which is diffi-

cult to achieve and highly important to on-going success and credibility. It is essential to recognize that a Phase III system in not only a management tool, but will have continued application as an inventory and analysis system as well. The “inventory” of data collected continues to have vital importance and the production processes to maintain it must continue.

SOLUTIONS

The first steps are now underway at CLDS to face the challenges that the transition to Phase III requires. Technical solutions to the need for distributed processing and local data bases are being found through the use of mini computers. The exact configuration is undecided at this time, but the eventual distributed land resource analysis system will likely involve a network of mini computers (or powerful micros) across Canada, linked to the central facilities. As well, it is anticipated that small portable microcomputers will be used in remote locations to analyze small area data sets extracted from the main files. At least one land planning software package is being tested for mini/micro implementation and existing graphics and statistical processing capabilities being upgraded.

The most significant barrier is the organizational/educational one. The disappearance of the distinction between user and supplier can only come through a long process of education and gradual change. The current “user” will need to be greatly more aware of system operations and EDP principles and, moreover, be willing to assume much more responsibility for data control, standards, coordination, quality assurance, etc., than now is the case. The current “suppliers” (EDP professionals) must learn considerably more about the techniques and practices of land use planning, modelling and subject matter data interpretation. It is implied that, in general, knowledge and skill levels of all concerned must be much greater and such multi-qualified people are in extremely short supply. It is expected that this problem will slow the development of a generalized modelling/planning system, although more application specific land management capabilities can be implemented where skill levels are high and appropriate management practices are in place.

As well, the techniques for forecasting land change or quantifying policy impact analysis are not well advanced. Mathematical models which are in use are simple, and clearly preliminary. Much research is required in this field before significant systems advances can be made.

REFERENCES

Crain, I.K. 1978, The Use of Data Base Management Systems in a Scientific Environment: Proceedings Canadian Computer Conference, Canadian Information Processing Society, Montreal, pp. 36-42.

Nolan, Richard L. 1979, Managing the Crises in Data Processing: Harvard Business Review, March-April 1979, pp. 115-126.