Data Modeling as Applied to Surveying and Mapping Data

by

C.A. Chamberlain and K. Lochhead
Geodetic Survey Division
Canada Centre for Surveying
Energy, Mines and Resources Canada
Ottawa , Ontario , Canada
K1A 0E9

ABSTRACT

The Geodetic Survey Division of the Canada Centre for Surveying is replacing the National Geodetic Data Base (NGDB) with the National Geodetic Information System (NGIS). For the NGIS to be successful, it was recognized that a sound, well engineered data model was essential. The methodology chosen to design the data model was Nijssen's Information Analysis Methodology (NIAM), a binary modeling technique that is supported by a Computer Aided Software Engineering (CASE) tool, PC-IAST. An NGIS prototype has also been developed using Digital Equipment of Canada's Relational Database (Rdb) management system and COGNOS Corporations POWERHOUSE 4th generation language.

This paper addresses the need for, and the advantages of using a strong engineering approach to data modeling and describes the use of the NIAM methodology in NGIS development. The paper identifies the relationship between the data model, data structures, the design and development of a database and the use of automated tools for systems development. In conclusion, critical success factors for the continuation of the N.G.I.S. developments are identified and the benefits that will accrue are enumerated.

THE NEED FOR DATA MODELING

There are many advantages to creating a data model before an information system is designed. In fact data modeling should be the initial step in the overall planning exercise. The reasons are very logical and provide the foundation from which to build an information system.

The first important point to recognize is that in any organization is that data has intrinsic value. It reflects the organization and its mandate. Data possesses far more permanence and stability than any other component in the organization. This approach is referred to as a "data-driven" approach and can be a valuable tool in designing systems that are more responsive to organizational needs. It has been proven that systems designed by a data driven approach have a better long term return on investment than information systems designed using traditional approaches. It seems logical, then, to begin designing an information system from this point of view.

Experience at Geodetic Survey illustrates this point well. In the past, many different geodetic type databases were developed as needs were identified. Consequently, interaction between these databases or any type of data sharing is non-existent. Examples of these "stand-alone" databases are the NGDB, an astronomic database, a levelling database, and a Doppler observation database. This approach is an "application" oriented development environment and is an example of how many organizations handled the database question for years. Using this approach, data bases simply become expensive file systems.

A new approach must be found to efficiently design information systems to manage the massive volumes of data and provide the means of extracting information from this data. One such new approach is referred to as the "subject database" approach. An example is extracting station vertical data and relating it to GPS determined station heights or Doppler determined station heights. To accomplish this task, one would have to access 3 or more different databases and somehow link the data in an "off-line" method or a difficult, inconvenient "on-line" method. This is a typical "application oriented" database design. It is desirable to have this data modeled once in the system and a unified connection to all other information created. This can be accomplished if the information system is designed using a structured engineering approach. Each piece of data or "entity" has certain properties or "attributes". These entities also possess certain common relationships that join them to other entity occurrences, either of the same or different entity type. In the above example the entity type "station" has various attributes such as all the various heights associated with it. This identification of the entities, their attributes and relationships is the heart of a data modeling approach to information systems design.

The data modeling exercise is completely independent of any data base management system (DBMS) and in most cases is process or application independent as well.

The benefits of a data driven or subject database approach over the traditional approach are:

- reducing data redundancy
- reduce the effort for updating these information systems
- reduce effort in program modification brought about by changes in the data definition
- increase data accessibility of data by providing multiple access paths to this data
- encourage the use of standardized data-naming conventions
- provide data security
- provide recoverability
- reduce operator intervention and associated production errors
- promote data communications services

These benefits are strong motivating factors for any organization to use data modeling.

DATA MODELING APPROACH

Bringing about a successful integration of information depends largely on the functions, the size, and complexity of an organization. To integrate information requires rigorous, analytical techniques, backed by powerful tools and a methodology that oversees the total life-cycle of an entire development project.

The methodology chosen for this project was the Nijssen Information Analysis Methodology (NIAM) (Nijssen 1979; Verheijen 1982; Isloor 1987). The six phased approach of the NIAM Information Analysis Methodology results in a comprehensive, easy to understand, rigorous, stable and verifiable requirements specification. User participation, project planning, data modeling and prototype construction are critical to this approach. The following discussion on the data modeling and application summarizes the details found in references (Isloor & Chamberlain ,1988; Systemhouse,1988).

The NIAM Information Analysis Methodology is divided into six phases as follows:

- Developing the Project Model
- Developing the Functional Model
- Developing the Information Flow Model
- Developing the Semantic Information Model
- Generating the Data Model and a Relational Schema
- Validating the Requirements Model

Figure 1 graphically shows the development stages, NIAM Phases, and associated deliverables.

Developing the Project Model

This phase is designed to facilitate an understanding and commitment. The Project Model created during this analysis will provide a clear definition of the roles, responsibilities, activities, and milestones associated with each stage of the project. The Project Model identifies two crucial aspects: Project Organization and Project Activities.

Vital to the Project Organization phase is the identification of three critical resources from within the project organization: Executive Responsible, Acceptor, and Project Manager. The end result of the analysis of this phase is a formal document called the Project Plan. This plan identifies the problem scope, objectives, and defines the roles, phases and deliverables. In addition, the Project Plan presents the initial schedule for the project.

Developing the Functional Model

The intent of the Functional Model is to gain a clear and stable picture of the hierarchy of basic tasks that will serve as a foundation upon which to build the Information System.

The functional model is a hierarchy of basic tasks in a business or operation within the defined scope of the project. This phase takes an unbiased, objective look at the various tasks. In a series of brainstorming sessions, the project team isolates and analyzes the system functions, breaking them down into subfunctions until a hierarchy evolves that describes the total operation. This method of analysis, termed " top-down decomposition", provides checkpoints for completeness and relevancy at each level.

Although basic functions remain stable, how one handles these functions is subject to change. For instance, an analysis of the current functions in light of today's changing technologies could cause the automation of one or more previously manual functions or could lead to reorganization. For this reason, the functional model looks beyond the current methods, structures, technologies and organization. Its intent is to give a clear and stable picture of the functions that will serve as the foundation for an information system that can adapt to change. Figure 2 shows the functional model formed during this project.

Developing the Information Flow Model

The information flow model represents the logical sequence of the functions by means of the information that must flow between the functions. This model serves as the basis for the derivation of the scope of the information system models. It also depicts the interfaces to environments external to the function in the functional model.

In this phase, the project team develops and establishes relationships between functions identified in the functional model. Like the Functional model, the Information Flow model is not tied to the current organizational structure, methods and processes.

Information flow is divided into two areas: input used by the function; and output produced by the function. Figure 3 is an example of an information flow diagram.

Developing the Semantic Information Model

The Semantic Information Model (Nijssen 1979; Verheijen 1982) addresses what information is used and graphically identifies relationships, behavioral constraints and validation requirements. This model is a rich semantic network that concentrates on the meaning of information rather than its structure. The information flows that were identified in the Information Flow Model are decomposed into objects and their references. These individual objects are then related to one another to show how they work within the function.

Figure 4 shows a part of the Semantic Information Model for a portion of the technical equipment information flow from Figure 3.

Generating The Data Model and Relational Schema

Using rigorous rules, the binary facts and constraints of the Semantic Information Model are grouped to form a Data Model. This model provides an overview of the data structure.

The Data Model is a normalized model generated from the Semantic Information Model. By providing an overview of the entire data structure, the Data Model assists in the selection of a Database Management System (DBMS) that could be used for the prototype. The next step from the Data Model is to use specific target DBMS pipes to generate relational schemas for the chosen target DBMS.

Validating the Requirements Model

The models developed thus far have provided a rigorous and computable set of requirement specifications. To further enhance the value of these models, a validation prototype is created for verification. The validation prototype generates the prototype database, collects, and loads real data (applying the rules determined during the development of the Semantic Information Model), and simulates logical processes.

APPLICATION OF THE METHODOLOGY AT GEODETIC SURVEY

The application of the methodology for the modelling of survey related data in NGIS proceeded as three major phases. They were :

- Project planning
- Data planning
- Data validation prototype

Project Planning Phase

At the beginning of this phase, a project team of twelve personnel (9 Survey Engineers and 3 computer scientists) were trained in the NIAM Methodology. The project team participated in the specific definition of the project and constructed a functional description of NGIS. The team also created a high level information flow model of NGIS functions and their external interfaces. The purpose of this phase was to gain a complete understanding of the project. The elapsed time for this phase was approximately 2 months.

Data Planning Phase

The initiation of this phase started with the review of the results of the project planning phase. IBM's Joint Application Development (JAD) technique with an external consultant acting as a facilitator was used in this phase to develop the data model.The results of the analysis were:

- The **Functional Model** was developed. The highest levels of the model are shown are shown in Figure 2. Each box or function on this figure was broken into sub-functions and defined in detail to gain a clear picture of the basic tasks to be performed.

- The **Information Flow Model** was developed to specify the data flows between the functions developed in the functional model. Refer to Figure 3 for an example.
- The **Semantic Information Model** was developed for each piece of information. The relationships, constraints and validation requirements of the data model were developed and graphically defined. Refer to Figure 4 for an example.

Details of the Semantic Information Model were:

- Number of objects modeled = 200.
- Number of attributes of objects modeled = 500
- Number of subtypes in the model = 40
- Number of constraints modeled = 90

In developing the data model, the entire team worked together to resolve all outstanding questions before a concept was finalized. Often the initial attempt at defining a concept was done by a subset of the team as appropriate to the area under consideration. Members of the team consulted other staff of Geodetic Survey Division on an individual basis, when such discussion was deemed essential to the analysis. A series of integration work sessions were held by the project team in order to ensure that an overlap or duplication would be identified and resolved. The elapsed time for this phase was approximately four months and the effort involved was 3 person years.

Data Validation Prototype Phase

The computer program PC-IAST (Control Data Corporation) was used to capture the Semantic Information Model pictorially. This tool is a product from the emerging field of Computer Aided Software Engineering (CASE). Using rigorous, automated rules, the binary facts and constraints of the Semantic Information Model were grouped by PC-IAST to generate the data model.

In the case of NGIS, this grouping produced approximately 500 records. Analysis of these records showed that based on sound design principles several records needed to be combined into larger records. Suppression of several reference classes and allowing null values in the Semantic Information Model allowed PC-IAST to regenerate a data model resulting in approximately 230 records. PC-IAST's SQL schema generator, in conjunction with some manual adjustments, created an SQL Schema for the NGIS data prototype.

The technical environment chosen to develop the NGIS prototype was the DEC/VAX 8300 using Rdb as the data manager with POWERHOUSE (Cognos Corporation) as a prototype generator. The prototyping environment is shown in Figure 5. The 200 screen NGIS prototype consists of:

- Three subsystems: Geodetic, Publications and Software. The model was divided into three subsystems to test the prototyping environment with two smaller models.
- In each subsystem the first screen established user identity and enforced user authorization. Subsequent menu screens provided access to various logically organized subsystems within the three major subsystems.
- A Prototype evaluation subsystem was developed which was automatically invoked at the end of using every logical subsystem to capture user comments (when they are fresh). This system gathered all comments and suggestions in a file which provided an excellent repository for information when making modifications to the Data Model.

The success of this prototype development is best illustrated by the following points ;

- Expertise in developing of information systems and using automated CASE tools for development, such as POWERHOUSE and PC-IAST, was developed within Geodetic Survey.
- An accurate data model was developed that will be used in the final final implementation of NGIS .
- The prototype helped identify many management issues and technical complexities that must be addressed before a successful implementation of NGIS is possible.

FUTURE PLANS

The roadmap for the development of a full implementation of NGIS is shown in Figure 6. A project must be initiated to perform each of the functions shown in the boxes. The modeling of the processes, as identified in the prototype development, is to be carried out in parallel with the standardization of codes (such as province codes, station numbers etc). Once the codes have been standardized, we can then determine the data conversion requirements to fit the data model. These activities will be labelled as Phase 1.

The next step, Phase 2 will be a feasibility study that will be carried out to develop a cost/benefit analysis, an organizational impact study and to address hardware, software and data communications issues.

Phase 3 is the acquisition of any hardware and software needed for the system. In parallel with the acquisition, work will be performed on the design of NGIS in the chosen hardware/software environment.

Phase 4 will be the development and full implementation of the NGIS.

Present plans call for the system to be fully developed and implemented by March 31, 1991.

CONCLUSION

The NGIS will form a solid foundation for data management at the Geodetic Survey. The following critical success factors will ensure that the project remains on schedule and budget:

- a commitment by management to the project
- commitment of a dedicated, enthusiastic project team and project leader
- the use of a structured data modeling approach
- the use of automated and rigorous CASE tools for all phases of the development life cycle
- the use of the Joint Application Methodology.

When completed, the NGIS will provide:

- better data for users and information for managers
- enhanced planning and control capabilities for the Division
- enhanced technology transfer
- an integrated environment for data processing and management
- a qualitative improvement in the products and services of the Division.

REFERENCES

Isloor,S.S, 1987 *A Logical,Comprehensive,and Automated Methodology for Data-Driven Systems Development Using Semantic Information Modeling* ,Advance Paper, Proceedings of the First International Workshop on CASE, CASE(87),Vol.2,Cambridge,Mass.

Isloor, S.S.& C. Chamberlain, 1988., *Strategic Scientific Data Planning in Geodetic Surveys*

Martin, J. 1982 ,*Strategic Data Planning Methodologies* , Prentice Hall Inc., Englewood Cliffs, N.J., U.S.A.

Nijssen, G., Van Assche, F., & J. Snijders,1979 End User Tools for Information Systems Requirement Definition, In *Formal Models and Practical Tools for Information System Design*, H. Schneider, Ed., North-Holland Publishing Company, Amsterdam, The Netherlands,

SYSTEMHOUSE ,1988 *NGIS Strategic Data Plan - Analysis,Design and Construction of a Validation Prototype* , Geodetic Survey Contract Report ,

Verheijen, G.M.A., & J. Van Beekum , 1982 NIAM: An Information Analysis Method, In *Information Systems Design Methodologies: A Comparative Review*, Proceedings of the IFIP TC8 Working Conference, Noordwijkerhout, The Netherlands, North-Holland Publishing Company

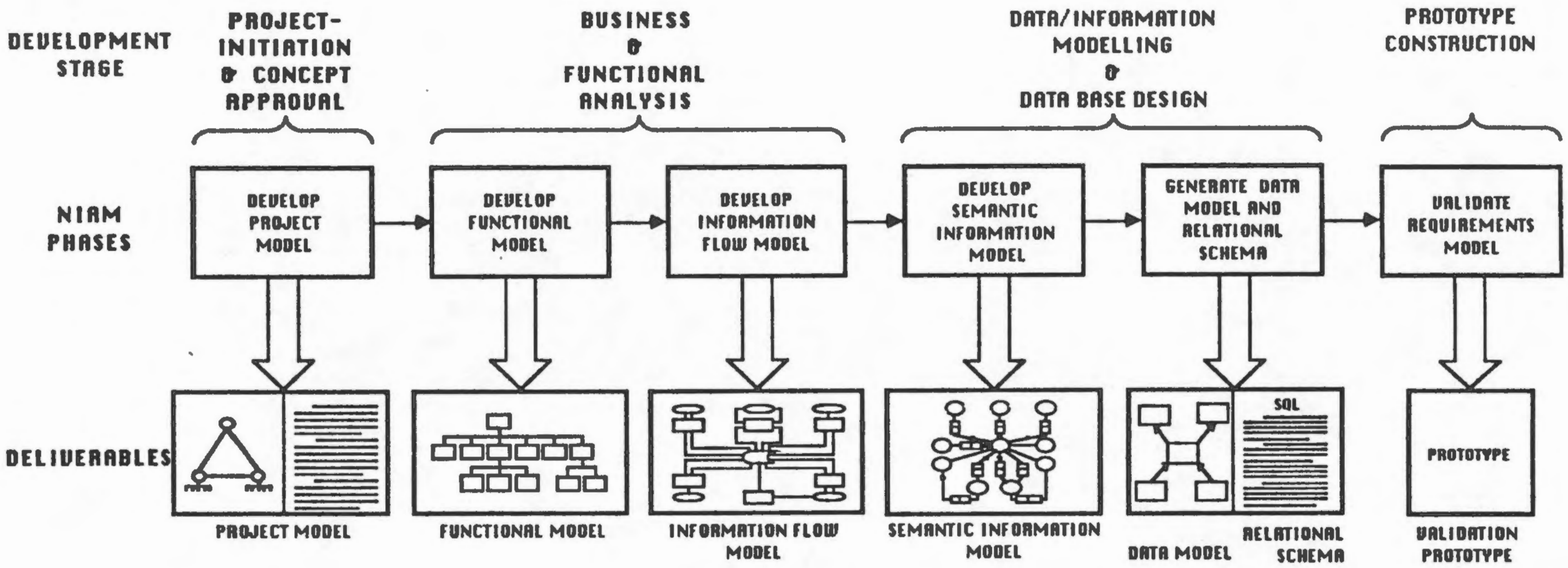

Figure 1: Phases of NIAM Information Analysis Methodology

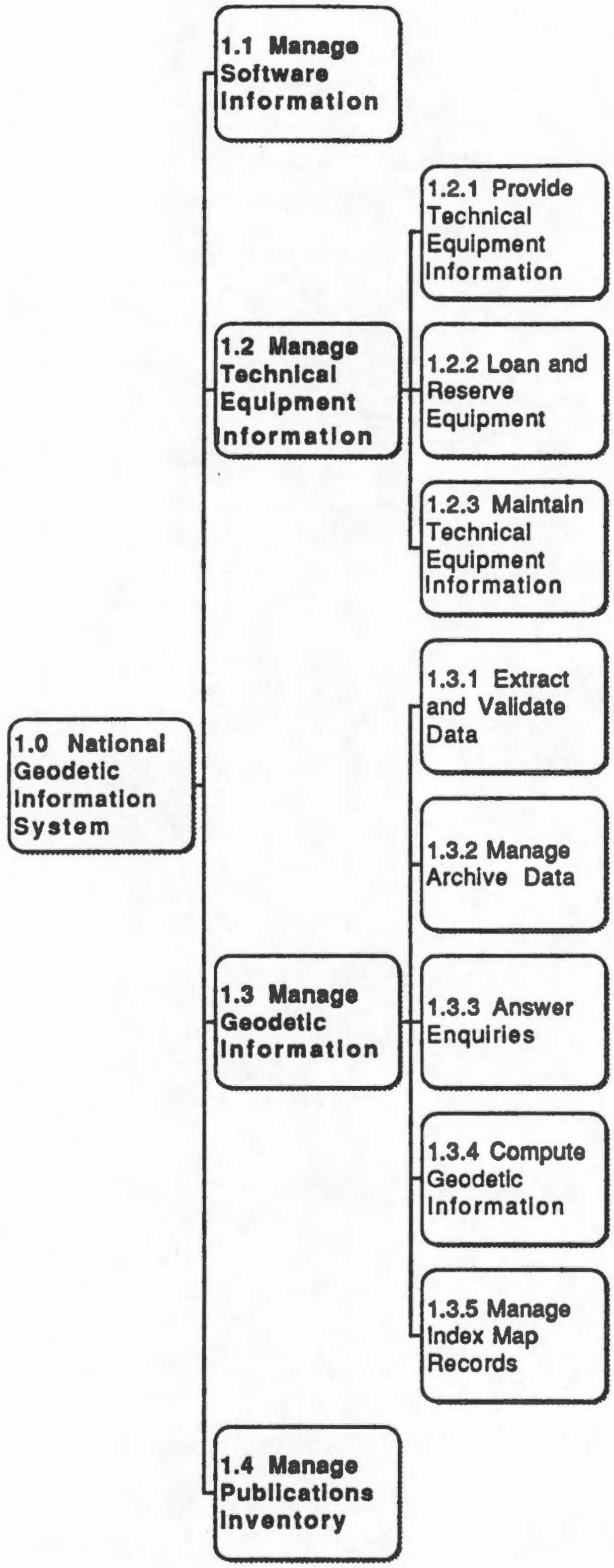

Figure 2 : Functional Model for the National Geodetic Information System

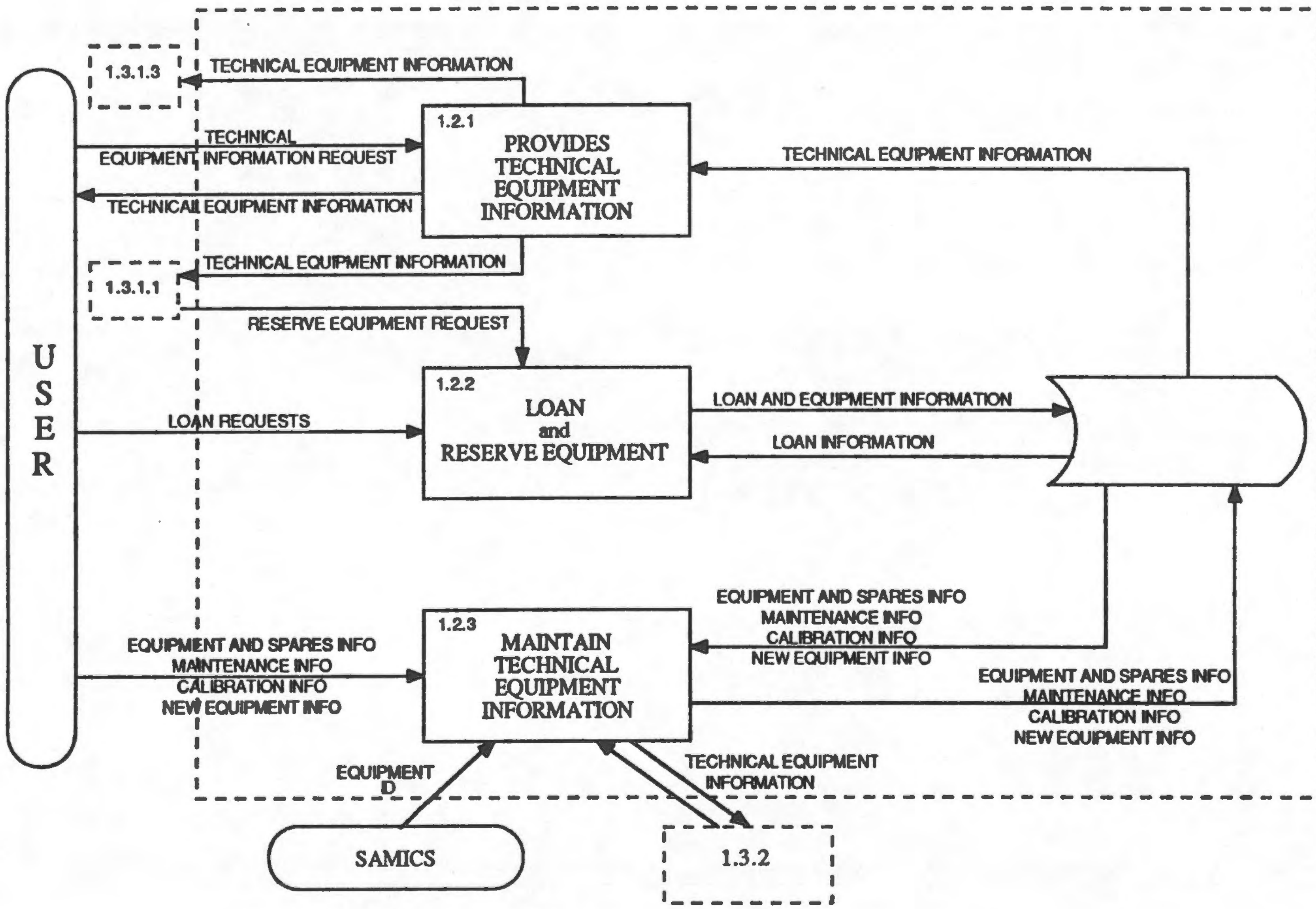

Figure 3: Information Flow Diagram of Function "1.2 Manage Technical Equipment Information" of Figure 2.

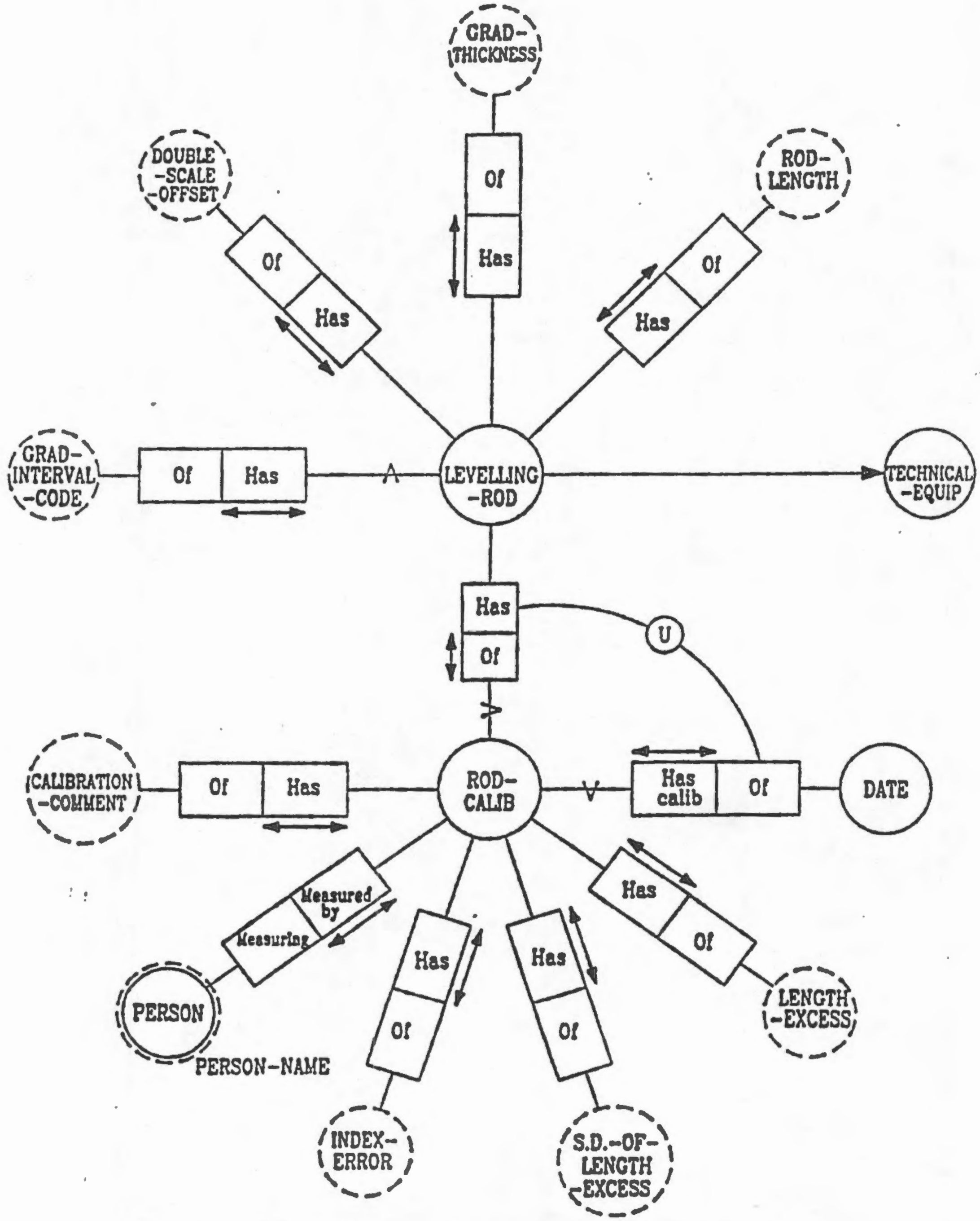

Figure 4: Information Structure Diagram for a portion of Technical Equipment Information for Figure 3.

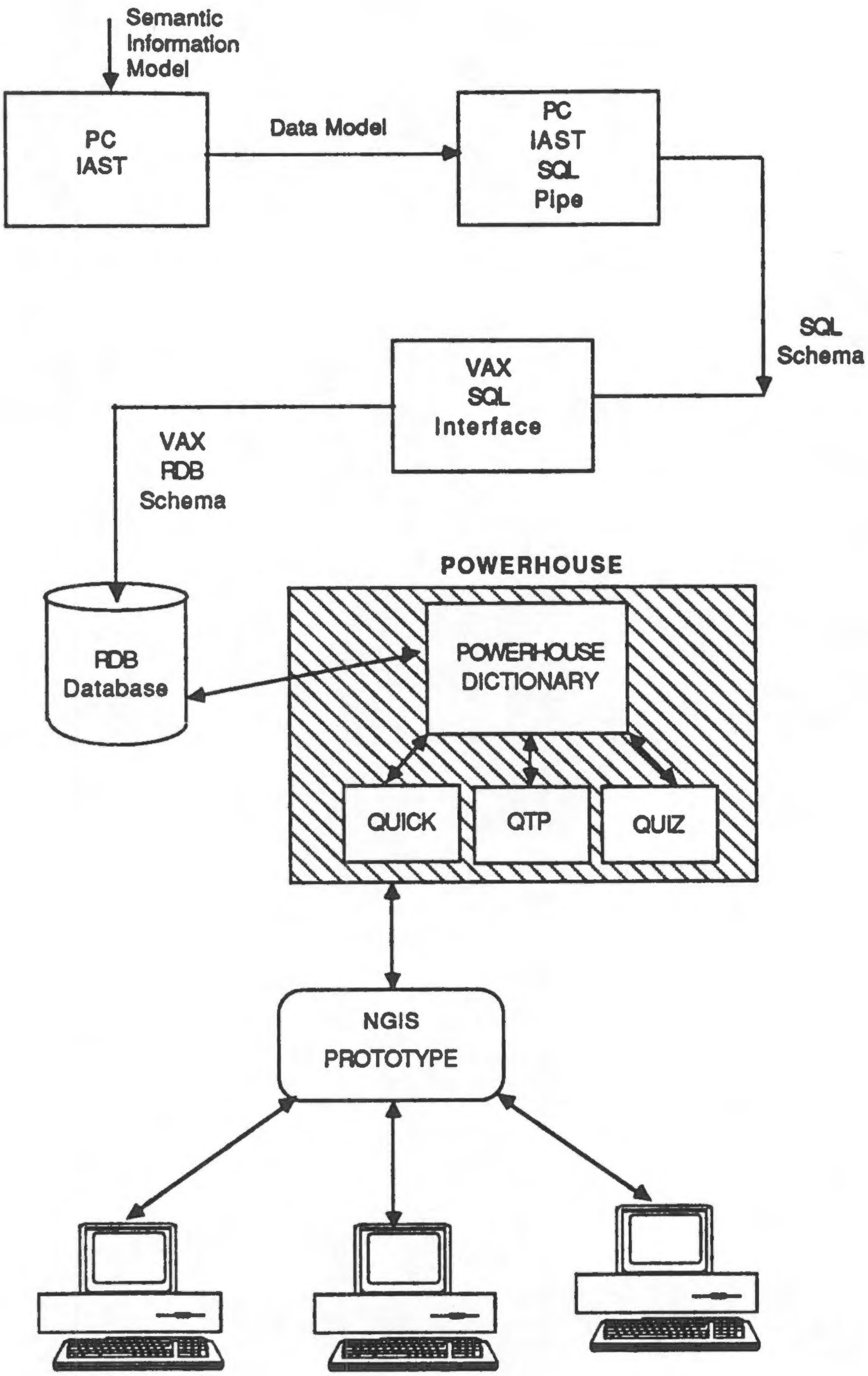

Figure 5: Prototyping Software Environment

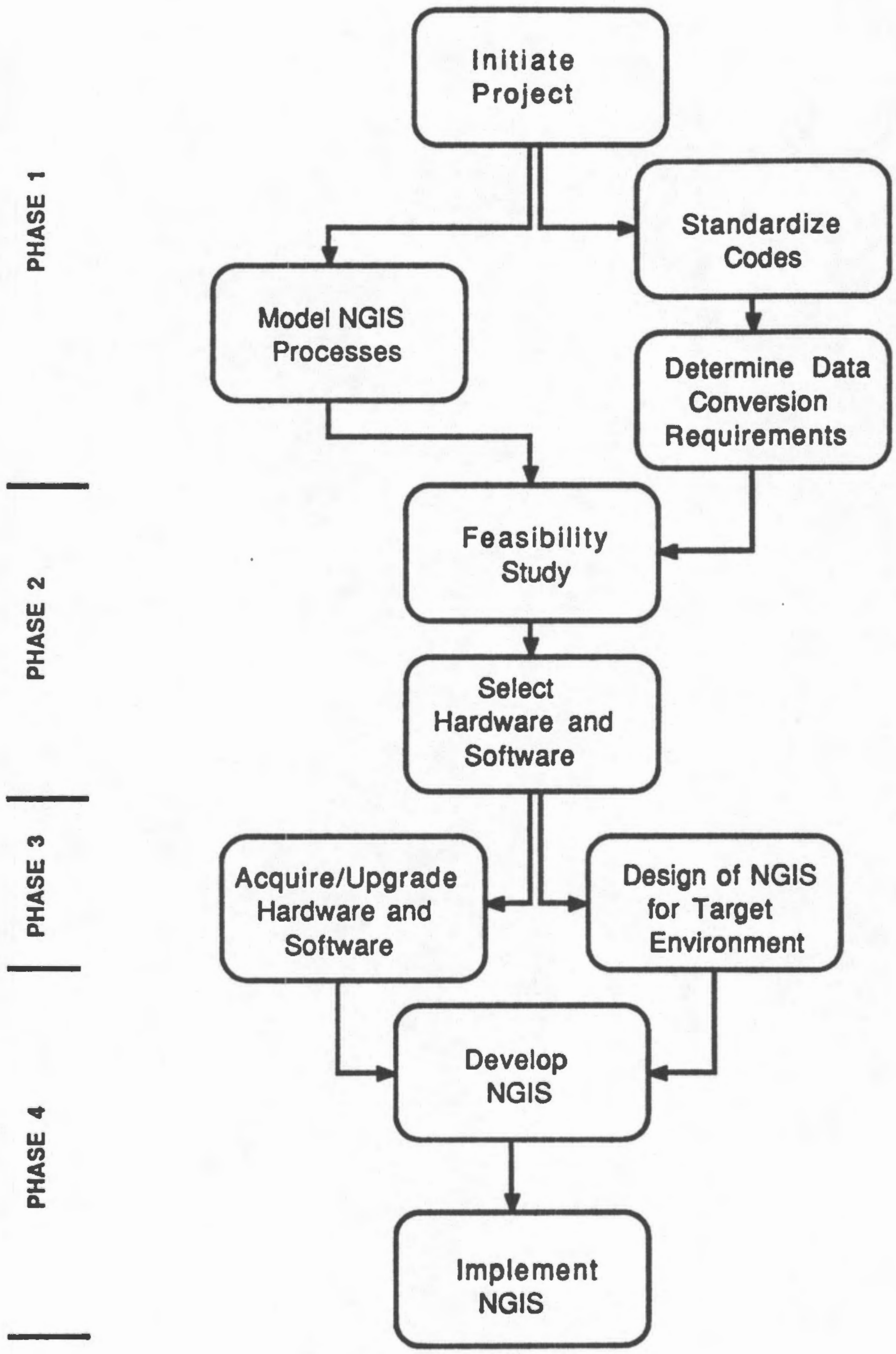

Figure 6 : NGIS : Plans for the Future