KNOWLEDGE BASE SYSTEM TECHNOLOGY IN THE DEFENSE MAPPING AGENCY'S DIGITAL PRODUCTION SYSTEM

Luanne I. Mulawa

Defense Mapping Agency
3200 South Second Street
St Louis, Missouri 63118

ABSTRACT

The implementation of Knowledge Base System technology for mapping, charting, and geodetic production applications is emerging out of its infancy. Consequently, literature on Knowledge Base Systems similar in size and complexity to the Defense Mapping Agency (DMA's) Digital Production System's (DPS's) is rare.

The DMA made a commitment to this applied artificial intelligence technology in an effort to secure a production capability with sufficient flexibility to respond to changing requirements and operational experience.

The goal of this paper is to provide an overview of the DPS with an emphasis on the Data Extraction Knowledge Base System. This Knowledge Base System was designed to provide decision support for automated feature extraction.

Knowledge engineers, programmers, and project managers coordinating, directing, or managing a "corporate" KBS development effort should gain valuable insight into the potential benefits associated with developing a large cartographic KBS for production application.

Key Words: artificial intelligence; digital system; knowledge base; feature extraction; mapping

INTRODUCTION

The Defense Mapping Agency (DMA) produces and distributes mapping, charting, and geodetic products; services; and training for the Department of Defense (DoD).

A host of new weapon systems, training systems, and intelligence analysis systems are being developed which will rely upon digital mapping, charting, and geodetic (MC&G) data for operation. The data requirements for these new systems are rapidly overwhelming current production resources.

In 1982, DMA initiated an ambitious modernization program. There were three fundamental driving forces that lead to this modernization program: DMA needed an end-to-end digital production system to effectively satisfy both existing and emerging requirements for digital MC&G data; management needed the capabilities to be more flexible and responsive to new requirements and technology; and DMA needed to increase production throughput and reduce operating costs for MC&G production.

This 10 year development initiative is rapidly approaching operational readiness. By the mid 1990s DMA mapping and charting will be revolutionized with the full deployment of this state-of-the-art cartographic production system referred to as the Digital Production System (DPS).

The DPS will provide DMA with a softcopy production system that will be integrated into DMA's existing hybrid (hardcopy/softcopy) mapping, charting, and geodetic production environment.

DPS's functionality and operations are distributed among seven "segments". The Production Management Segment (PM/S) controls and monitors DMA production. It performs production planning and programming activities; generates production plans, assignments, and schedules; and manages production resources. The Data Services Segment (DS/S) manages system communication services; manages temporary and archived storage services; and manages the DPS MC&G Data Base.

There are five segments that make up the DPS production pipe-line. The first of these segments is the Source Acquisition Segment (SA/S). This segment provides DPS with the capabilities to manage DMA image library holdings and the geodetic control data base. The Source Preparation Segment (SP/S) collects and evaluates all source materials applicable to a specific assignment; conducts point mensuration for geopositioning; and assembles assignment packages required for DE/S and PG/S production purposes. The assignment packages include instructions, specifications, and source material listings required to complete a production assignment. The Hardcopy Exploitation Segment (HE/S) performs triangulation (outputs include rectified and ortho-rectified imagery). The Data Extraction Segment (DE/S) extracts terrain and feature data from softcopy imagery. This data is edited and then stored in the MC&G Data Base to support product finishing. The last segment in the production pipe-line is the Production Generation Segment (PG/S). PG/S transforms, integrates, and generates graphic and digital data for hardcopy and digital product finishing.

Within DPS there are KBSs that provide decision support for planning of production programs; stereo extraction of cartographic features; and data segregation, symbolization, conflict detection and resolution, generalization, and labeling for product finishing. These KBSs will play a central role in DPS performance and flexibility, and in protecting DPS from technological obsolescence. Each KBS was designed differently due to segment functional and performance requirements, how each KBS operates within the segment's overall design, and differing strategies and philosophies of segment developers.

The remainder of this paper will focus on the design, operations, testing, and maintenance of DPS's most mature large scale production KBS, the Data Extraction KBS. This system was developed under contract with General Dynamic's Electronic Division. The DE/S KBS is a decision aid that will revolutionize feature extraction

for DMA. It has been designed to place cartographic knowledge as well as, systematic search and reasoning skills at the finger tips of all DE/S cartographers in the production data extraction environment. With the objectives being improved production quality and throughput.

DATA EXTRACTION SEGMENT

The Data Extraction Segment provides the capabilities in the DPS environment to extract terrain elevation and feature data from digital stereo imagery using softcopy photogrammetric equipment; to edit and merge newly extracted data with existing MC&G data; to perform quality checks to ensure that the extraction process has conformed to production assignment specifications, and to store the resulting digital elevation data and feature data files in the MC&G Data Base for later use in production processes such as product finishing.

In order to achieve DPS production requirements, the DE/S production environment consists of production and support activities. The DE/S production activities are segment control, data extraction, data integration, quality control and quality assurance. The support activities are training and maintenance.

Segment control includes the functions of data access; internal and external communication; resource allocation and utilization; production monitoring; and production control.

Data extraction is performed on the softcopy (digital) Terrain/Feature Extraction (TE/FE) workstation. Data extraction can be thought of as the most significant production activity in the DE/S production work flow.

To support the collection of terrain data, DE/S functionality has Automated Terrain Extraction (ATE) and Interactive Terrain Extraction (ITE) subsystems. ATE provides the capability to automatically collect and output elevation data at fixed and variable grid spacings without the aid of the cartographer. ATE is the first in a sequence of processes performed for data extraction. ITE follows and utilizes the output generated from the ATE process. The ITE provides the capabilities to collect geomorphic feature data and fill adverse areas where the ATE was unable to extract elevation data.

To support the collection of feature data, the DE/S provides the cartographer with an Automated Feature Extraction (AFE) subsystem. This subsystem contains a KBS and computer vision (CV) tools. The KBS supports the AFE process in the identification, attribution, and consistency checking of delineated feature data. The computer vision tools are designed to assist the cartographer in the automatic delineation, metric measurement, smoothing and thinning of feature data. The DE/S provides the cartographer with an option to perform automated delineation using CV tools or interactive delineation using graphic/editing tools. Once a delineation has been completed and accepted by the cartographer, the KBS is automatically invoked for the sequential execution of automated feature identification, attribution, and consistency checking of the delineation. The cartographer must respectively review all candidate identification, attribution, and consistency check queues.

During the review process the cartographer has the opportunity to override or accept a KBS suggestion.

Data integration is performed on a Global Edit workstation. This workstation is a TE/FE workstation configured with additional storage space to accommodate feature and terrain data from multiple assignment areas. Data integration provides for the merging of terrain and feature data between adjacent data sets and for interactive editing of data.

Quality assurance provides for inspection and verification of extracted data before release of the data to the MC&G Data Base.

As for DE/S support activities, training utilizes DE/S subsystems as required. This training includes a Self-Paced Autotutorial System (SPAS) and conventional training courses for DE/S managers, operators, users, and programmers. SPAS training is computer-based instruction that is interactively done on the TE/FE workstation.

DE/S maintenance consists of a collection of functions which aid in the general upkeep of DE/S hardware and software resources. These functions include system startup procedures, fault detection, software backups, corrective maintenance, preventative maintenance, equipment status reports, KBS maintenance, hardware upgrades, and software upgrades.

The maintenance subfunction of primary interest to this paper is KBS maintenance. KBS maintenance provides a means for correcting problems with the KBS, for responding to extraction and product specification changes, and for enhancing the performance of the KBS.

The DE/S KBS maintenance function is supported by the following resources: DE/S cartographers; Summary Reports; a Knowledge Engineer (KE) workstation; and a TE/FE workstation. DE/S cartographers are utilized as "domain experts" for knowledge acquisition activities in support of KBS rule development and performance analysis. The Summary Report is the primary source of information for assessing KBS performance. Essentially, it is an audit trail of KBS performance for an extraction assignment. Information collected includes statistical information about extracted data, cartographer comments, and the number of times KBS suggestions were overridden by the cartographer. The KE workstation is used for Knowledge Base rule editing and verification, data base management, and report management. The TE/FE workstation is used during Production Center scheduled maintenance sessions to test candidate changes and new rules for the KBS.

KNOWLEDGE BASE SYSTEM DESIGN

Much ingenuity and effort has been expended tackling the design and integration of the DE/S KBS. Located in the AFE subsystem, the underlying KBS software architecture is written in Ada and uses VAX relational data base software to store the specification files from which the AFE Knowledge Base is built. The operational concept of the KBS is to support automated feature extraction by serving as a decision aid in the

identification, attribution, and consistency checking of feature data.

Following classic Knowledge Base System design concepts, the DE/S KBS separates the system's declarative knowledge from the code that controls the inference and search procedures that are contained in the system. The declarative knowledge is kept in the AFE Knowledge Base and the control knowledge is kept in the custom designed inference engine. Basically, the inference engine controls the system when it searches the knowledge base.

The DE/S KBS is a "virtual KBS". That is, every DE/S TE/FE workstation is configured with a KBS tailored to the assignment to be worked. Meta-knowledge rules are used to filter out all rules that are not relevant to the assignment's extraction specification requirements and geographic area.

The baseline approach to AFE Knowledge Base population and maintenance is driven by the Consolidated Product Extraction Specification Database (CPESDB). Approximately eighty percent of all knowledge base rules are automatically generated from the CPESDB. In automatically generating these tens of thousands of rules, KBS software architecture includes a rule translator that draws on the extraction specifications, data content specification, glossary, and delineation set portions of the CPESDB. These are the rules that define the features, attributions, and capture conditions for an extraction specification, the allowable values for an feature-attribute pair, and the data base representation for feature delineation. The AFE Knowledge Base rules not automatically generated from the CPESDB are facts and heuristics acquired through knowledge engineering.

The AFE Knowledge Base on the TE/FE workstation is composed of two data bases: the AFE Spec Data Base and the KB Rules Data Base. The AFE Spec Data Base contains all rules that are automatically derived from the CPESDB. The KB Rules Data Base contains all rules that constitute "value added cartographic knowledge" obtained through knowledge engineering. The total number of rules making up a workstation's AFE Knowledge Base varies by extraction assignment. For a typical extraction assignment approximately 10,000 rules are downloaded into workstation memory.

KNOWLEDGE BASE SYSTEM OPERATIONS

PLANNING AND STRATEGY

Prior to the beginning of an extraction assignment the KBS includes rule filtering to minimize the size of the compiled AFE Knowledge Base. Rule filters as specified by KBS operations are addressed:

Extraction Specification Filtering:

The DPS Extraction Specifications are generated in softcopy. These specifications are converted into AFE Spec Data Base Rules. The KBS Rule Compiler filters out all rules with a finding, hypothesis, or fact reference to a feature (cartographic entity such as a road, river, or building) that is not in the list of valid features for the extraction assignment.

Region Filtering:

The KBS Rule Compiler filters out all rules with a finding, hypothesis, or fact reference to a region (geographic area) that is not the region for the extraction assignment. A region is an area of homogeneity based on physiography, hydrography, culture, and vegetation with a discrete boundary defined in terms of latitude and longitude coordinate pairs.

Country Filtering:

The KBS Rule Compiler filters out rules with a finding, hypothesis, or fact reference to a country that is not a valid country for the extraction assignment.

Context Filtering:

The KBS Rule Compiler filters out rules with a finding, hypothesis, or fact reference to a context that is not a valid context. A context area is an area of like purpose or usage (ie. agriculture area, built-up area, or mountainous area) for the extraction assignment.

Prior to delineating any features, the cartographer is asked to roughly delineate areas of similar context on the TE/FE workstation. The KBS assists the cartographer in labeling the contexts.

AUTOMATED IDENTIFICATION ASSISTANCE

Once a delineation is completed and accepted the KBS automated identification assistance process is invoked and the identification rule class is applied. This rule class has two subclasses, a metric partition subclass and a feature subclass. A metric partition is a category (or bin) for feature assignment based on the metric aspects of size, shape, and height. All features are assigned to one or more of the 24 currently defined metric partitions. The KBS basically "sees" the delineation in terms of size, shape, and height metrics. The metric partition rules are used to filter out all but those features having the same metric characteristics as the delineation. The feature subclass is designed to determine the features within the metric partition that are the most probable candidates for identifying the delineation. This is achieved by accumulating evidence based on region, context, country, extraction specification, and spatial association characteristics.

The candidate feature identification list must handle up to eight feature codes, any one of which is selectable as the approved feature. The approved feature is either selected from the candidate list or interactively entered by the cartographer.

One reason for the speed of the identification process is that the KBS is selective about which features it will identify. During the knowledge acquisition phase, DMA "domain experts" helped the contractor identify those features that occur most frequently during feature extraction. These features are commonly referred to as "priority features". These are the only features that are available to the KBS for the automated identification process. This priority feature list is fine-tuned to geographic areas during region knowledge acquisition and rule development studies. Which means some

features not on the list are added, while others on the list are deleted.

AUTOMATED ATTRIBUTION ASSISTANCE

Once the stereo delineation has been identified the automated attribution assistance process is invoked and the attribution rule class is applied. An attribute is an inherent characteristic or descriptor of a feature. An attribute value is a numeric or text value assigned to an attribute. The necessary attributes and the order for attribution is determined from the AFE Spec Data Base. Initially, the KBS will evaluate some attributes using the values that can be retrieved from the metrics of the delineation. If possible, the KBS will use these values to automatically evaluate other attributes through procedural attachment calculations. Next, the KBS will automatically evaluate some attributes using the values that can be retrieved from extraction specification and default knowledge stored in the AFE Knowledge Base. Some attribute values cannot be inferred by the KBS. These values must be entered by the cartographer. However, at each step in the attribution process, the KBS will attempt to infer and automatically evaluate some attributes from other attribute values and user input. The cartographer can accept or override and interactively enter any attribute value.

CONSISTENCY CHECKING

Once automated attribution processing is complete consistency check rules evaluate the consistency between feature attribute values. These rules help enforce data quality.

VALIDATION CHECKING

Validation checks made by the KBS fall into four categories: capture condition checks, delineation type (point, line, area) checks, delineation placement (ground, top of feature) checks, and water extraction checks. These rules also help enforce data quality.

KNOWLEDGE ACQUISITION

The process of acquiring knowledge for the DE/S KBS includes the collection of facts, heuristics, specifications, experience, procedures, and logic to accurately assist in the identification, attribution, and consistency checking of cartographic features. It also includes the knowledge engineering activity to analyze the gathered facts and heuristics and to translate them into the machine readable knowledge representation for the AFE Knowledge Base.

Because it takes experimentation to achieve high levels of performance, the AFE Knowledge Base has evolved gradually. During the early phases of knowledge acquisition, the activities of this task were driven by the requirement to support the system architectural design, software preliminary design specification, and the KBS prototyping tasks. The greatest priority of knowledge acquisition was directed at support to the KBS design so as to insure a design that was adequately defined to accommodate the needs of the future production knowledge base.

To ensure that the KBS was user oriented and met requirements, an incremental, iterative prototyping approach using government cartographers and contractor KE personnel was used.

During this prototyping activity regular interactive terminal sessions were conducted as part of the knowledge engineering process. These interactive terminal sessions provided an avenue for rapid testing of ideas in developing the design and operations concept for the KBS. The designs that didn't work well were quickly discovered and discarded, saving months of work.

Later phases of the knowledge acquisition and engineering tasks were directed at actually constructing a knowledge base suitable for production and on improving overall KBS performance. During this phase in the development life-cycle an emphasis was placed on ramping-up knowledge engineering skills of government KBS personnel.

Knowledge acquisition is an on-going effort. Right now, emphasis is being placed on region knowledge acquisition to support region rule development for geographic areas scheduled for early production. Spatial association rules to enhance verification and validation of extracted data and product specific rules to fine-tune the performance of the KBS for extraction specific assignments are also current subjects of knowledge acquisition/engineering.

KNOWLEDGE REPRESENTATION

RULE CLASS SPECIFICATION

The basic structure of a rule is (<rule_name>...). The <rule_name> has three principle components:
1) <afe_subfunction> (relates to class),
2) <rule_objective> (relates to subclass), and
3) <rule_type>.
The integrated conceptual representation of the rule <rule_name> is <afe_subfunction>-<rule_objective>-<rule_type>.

The AFE state space is defined through the use of the three rule types. Partitioning of the state space is done with the notion of rule classes and subclasses. Classes include planning, identification, attribution, consistency check, and validation. Classes are referenced in the <afe_subfunction>, or first element of the rule name. Each class has a set of subclasses which further prunes the search space. Subclasses are referenced in the <rule_objective>, or second element of the rule name.

RULE TYPE SPECIFICATION

Three rule types are defined to implement the abductive, suspended data acquisition inferencing methodology employed in AFE. Each is described:

Goal Rule Type: Goal rules supply a list of hypotheses in response to an inference query request. The function of this rule type is to provide the list of candidate solutions. The conceptual representation of a goal rule is-

(RULE-NAME QUERY HYPOTHESIS)

RULE-NAME is rule class dependent.

QUERY is the subject of the information request.

HYPOTHESIS is a candidate response selection presented to the cartographer.

Evidence Rule Type: Hypotheses accrue evidence through the inference evaluation of evidence rule facts. The function of this rule type is to support or refute the hypothesis. The conceptual representation of an evidence rule is-

(RULE-NAME HYPOTHESIS FACT WEIGHT)

RULE-NAME is rule class dependent.

HYPOTHESIS specifies which hypothesis in the (GOAL LIST) the evidence rule supports.

FACT is a formula that can be resolved to either true or false by the inference engine.

WEIGHT specifies an integer value between -100 and +100 which may be added to accrued evidence support of hypothesis if the fact is evaluated to be true.

Fact Find Rule Type: Fact find rules are utilized to suspend computationally expensive data acquisition during the inference process until it is absolutely necessary to resolve an evidence rule fact. The function of this rule type is to acquire additional information needed to resolve evidence rule facts. The conceptual representation of a fact find rule is-

((RULE-NAME) (UNRESOLVED-FACT) (RESOLVER-FACT) (PROC-SPEC) COST)

RULE-NAME is rule class dependent.

UNRESOLVED-FACT specifies information required to complete the evidence rule evaluation.

RESOLVER-FACT assists in selecting the correct fact finding rule from a list of several candidate fact-find rules.

PROC-SPEC contains information about the procedure to run and its argument bindings.

COST is positive integer value that indicates how much effort must be expended to invoke the procedure.

Orders the fact find agenda, attempting the highest ratio of the accumulated weight divided by the cost first.

RULE STRUCTURE SPECIFICATION

Two rule data structure specifications exist and are detailed here:

The extracted rule data structures provide an intermediary storage mechanism during rule compilation. The rule data is fetched from the respective rule database into an extracted rule form on its way to the rule compiler. The extracted rule form is also used for editing rules on the KE workstation (to be discussed later) since the extracted rule data structures store the individual components of the rule as a text string. This speeds the display process considerably.

The compiled rule data structures store the rules at the primitive level where all strings have been converted to integer tokens or enumerated types. This representation increases the inference engine throughput by eliminating the string handling required to process symbolically in Ada.

INFERENCE STRATEGY

The true power of the DE/S KBS is in the effective management of the reasoning process by utilizing a customized abductive paradigm.

The inference engine is built around the general task of developing "hypotheses", which account for a set of "findings". Since a finding may only partially support a hypothesis, a strength-of-evidence mechanism provides the ability to determine to what extent a finding is diagnostic of a given hypothesis. In the course of determining the total strength (weight) of evidence in favor of a hypothesis, it may be necessary to obtain additional findings or facts. The inference engine has a mechanism for assessing the cost of obtaining this additional information.

Procedurally, for this abductive paradigm to work, an initial query is presented to the system and a list of possible solutions is obtained. Evidence is accumulated to support or refute the hypotheses. A hypothesis accumulating 100 or more points during evidence accrual is considered to be the "winner" for the given query. A winning hypothesis is recursively submitted to the inference process as a new query. It is in this mechanism that hierarchical partitioning of the problem space is allowed. Therefore, the final solution set for an initial query could span multiple nested queries.

As soon as any single inference "cycle" fails to produce a hypothesis with at least 100 points, or a set of hypotheses does not exist for a nested query, the inference process is complete and the result (the given query when a nested hypothesis set does not exist for the query, or the full list of ordered hypotheses when a "winner" is not obtained) is returned to the invoking process. The algorithms to effect this methodology have been implemented as part of the production KBS code.

INFERENCE INITIALIZATION

The hypothesis list is obtained through the use of goal rules. A query specifying the AFE subfunction, rule objective, and finding is submitted to the inference engine. The AFE subfunction and rule objective are rule partitioning indices used to fetch from a set of relevant rules. This list of rules is then pruned by matching the finding of each rule to the specified finding in the query. The set of remaining rules becomes the goal refinement list where each hypothesis on the list has an associated score.

EVIDENCE EVALUATION PHASE

The ordered list of hypotheses (goal refinement list) is refined by evaluating evidence rules. Following the inference initialization phase, a set of evidence rules is obtained for each hypothesis on the goal refinements list. An evidence rule has a fact (that can be evaluated) and a weight that is to be added to the corresponding hypothesis if the evidence rule evaluated to TRUE.

Facts are defined recursively through the use of the logicals AND, OR, and NOT. The logicals behave intuitively, that is, short-circuit evaluation is performed where possible. For example, an AND logical assumes value FALSE as soon as any single fact component of the AND is FALSE. Thus, fact ordering may increase the efficiency of the inference process by placing the most critical fact components first and the least determinant fact components last.

Fact evaluation proceeds depth first. Nesting (if specified) is attempted before any subsequent facts at the same level are evaluated. Values for facts are typically obtained by fetching data from the blackboard (a repository of status information) and performing the prescribed operation, such as an =, >, <, etc. Values may also be determined by making boolean-like queries to the blackboard. If, during the evaluation process, a value cannot be determined immediately for a specific fact, the fact is placed on the fact find agenda. Fact evaluation continues until the facts for the current evidence rule are exhausted or a boolean value is obtained for the evidence rule through short-circuit evaluation. If the evidence rule assumes a value of TRUE, the related hypothesis's score is updated by adding the evidence rule's weight. It should be noted that evidence weight may be positive or negative.

FACT FIND PHASE

Should necessary information be unavailable during evidence rule fact evaluation, the unresolvable fact is placed on the fact find agenda. Fact find rules are defined for data calculations that are computationally expensive and thus processing is deferred until the data is absolutely required to fully evaluate an evidence rule fact. In some instances, an evidence rule may have its value determined without processing the fact find agenda as could happen due to sort-circuit fact evaluation.

When the evidence evaluation phase is completed without determining a winning hypothesis and unresolvable fact reside on the fact find agenda, the fact find phase is initiated. For each entry on the agenda, a fact find rule is found by fetching a set of candidate fact find rules and resolving the set down to a single rule.

It is entirely possible that there may not be a fact find rule for any one or all of the fact find agenda items. Fact find agenda items without attached fact find rules are so tagged and moved to the bottom of the agenda. The fact find agenda is then popped (in a conceptual sense, the fact find agenda items are never deleted from the agenda, only reordered) and the specified procedure is run. If the called procedure returns a value, each evidence rule linked to the fact find agenda item is again attempted. Only this time the evaluation is attempted in an inside-out path. That is, from the once unresolvable

fact's position in the fact out through its parent logical (AND, OR, NOT). Of course, if the fact was at the top-level, it would merely be evaluated and be assigned as the value of the evidence rule. If any evidence rule is fully evaluated to TRUE or FALSE and unresolved facts from the evidence rule remain on the fact find agenda, references to the evaluated evidence rule are removed from any associated fact find agenda items and the agenda item's weight to cost ratio is updated. This process continues until either the fact find agenda is exhausted or a winning hypothesis is obtained.

KNOWLEDGE BASE SYSTEM TESTING

DEVELOPMENT TESTING

The goal of testing during the KBS development phase was threefold:

1. To provide an opportunity to see how the AFE KBS handles a normal stream of problems and to modify the system when ever it encounters cases that it isn't prepared to handle correctly.

2. To provide an opportunity to see if the AFE KBS performance is acceptable for production.

3. To give government cartographers an early opportunity to learn how to use the AFE KBS to help them perform their work.

An overview of the major activities conducted during the development phase to verify and validate the functional, performance, interface, and design requirements of the KBS are described as follows:

Trial Extractions:

The primary purpose of the Trial Extractions was to evaluate the performance of the KB Rules Data Base. This included acquiring, formalizing, encoding, testing, and improving the knowledge. The principle goals for trial extractions were: 1) Verifying identification and attribute rules: defaults, regional, inferred, product-specific, and metric procedures; 2) Evaluating the content and format requirements of the Summary Report; and 3) Improving the design of the KE workstation.

Knowledge Base Milestone Evaluations:

Verification is a well-defined task that has been used in the development of the KB Milestone Evaluation Reports. The goal of these reports has been to ensure that KB Rules were written without error and that the semantics and syntax for the rule-based knowledge representation was consistent and complete. In addition, these reports have attempted to assess the performance of the AFE KBS during automated identification and the automated attribution assistance processes. The key results obtained from the last milestone evaluation report are as follow:

At this point in the development life-cycle, the AFE KBS automated identification process provides a list of candidate feature identification codes 66 percent of the time for all possible delineations over all DE/S products. Region and product specific rule developments for the

next KB Rules Data Base version should substantially improve the performance of this process for production.

Automated attribution assistance includes the automatic population of attribute values before and during the attribution process. For this milestone, automated attribution assistance was provided for 81 percent of all attributes for all features over all DE/S products.

Acceptable performance for identification and attribution processes for full production is "success" in 80 percent of all cases that an experienced cartographer could solve.

Software Build Tests:

Software Build Test procedures were written by the contractor and witnessed by government KBS personnel to verify and validate KBS and KE workstation functionality.

Factory Exercise and Demonstrations:

The Factory Exercise and Demonstrations (FEDS) activity associated with the KE workstation served a dual purpose. FEDS allowed the KE workstation functionality to be evaluated for operational readiness and it provided government KBS personnel with an opportunity to obtain early hands-on experience using the KE workstation. Side benefits from this activity included the development of a KE workstation training curriculum and benchmark tests for KE workstation functionality.

PRODUCTION TESTING

The operational concept for the testing of candidate AFE Knowledge Base changes in full production is summarized as follows:

Development and initial testing of candidate rule changes is done on the KE workstation; once this phase is completed robustness testing is conducted through an alpha and beta testing cycle on the TE/FE workstations.

Initial testing of candidate AFE Knowledge Base changes is done on the KE workstation through rule edit and verification functionality. The objective is to verify the syntactical correctness of the rules and the correct handling or rules by the inference engine.

The next level of testing is alpha testing which applies the candidate AFE Knowledge Base to a pseudo production assignment on the TE/FE MFWS. This testing is scheduled on a production TE/FE workstation configured for (off-line) maintenance. This allows the KE to conduct experimentation and analysis of the KBS in a unconstrained fashion. Once an acceptable level of stability and performance is achieved, beta testing is scheduled.

Beta testing applies the candidate AFE Knowledge Base to multiple production assignments. The key to the success of beta testing is selectively identifying production assignments that will exercise the maintenance action taken. Analysis is conducted through review and evaluation of the Summary Reports. Once a satisfactory confidence level is achieved during

beta testing the candidate changes are consolidated into a new production AFE Knowledge Base.

The final test is an integration test of the consolidated changes. Successful conclusion of this test results in the installation of the new production AFE Knowledge Base under configuration management and control.

KNOWLEDGE BASE SYSTEM MAINTENANCE

Unlike conventional computer software that are only infrequently updated, the DE/S KBS by the nature of its design will contain knowledge that will be in a constant state of change and evolution. Therefore, the system was designed to facilitate sustained maintenance. This flexibility is necessary not only to fine-tune the quality of DPS production operations, but also to enhance the performance of the DE/S KBS itself. Thus, the system allows experimentation with KBS behavior without contaminating or perturbing the production environment.

The DE/S KBS is a large strategic system that has a full-time government maintenance staff cognizant of the KBS architecture and the data extraction activity. This staff is physically located in the production environment and is responsible for troubleshooting relevant system code as well as for translating programming and cartographic issues into the AFE KBS. They are also readily accessible for operational support.

Since rules are meant to be changed, KBS maintenance procedures include capturing decision statistics to support refinement of the knowledge base employed during feature extraction. It also includes incorporation of changes and enhancements to KBS rules, programs, and data structures. The KBS Administrator manages this process to assure that it is accomplished in an efficient and effective manner within the constraints of configuration management and control.

Knowledge Base maintenance activities include:

1. Monitoring and assessing the performance of the KBS to determine where and when improvements can be made.

2. Investigating user overrides to KBS suggestions to assess trends that may impair production.

3. Generating and verifying candidate rule improvements to the KBS and, as applicable, and modifications to governing extraction specifications using the KE workstation.

4. Designing, scheduling, and implementing experiments to test candidate rule improvements to the KBS.

5. Analyzing test results to determine the value of the proposed changes to the KBS.

6. Documenting and communicating all test results through appropriate configuration management and control channels.

7. Recording and tracking maintenance modifications to the KBS.

8. Managing maintenance of the KE workstation.

9. Maintaining KBS rule documentation.

A variety of events can invoke KBS maintenance. In each case, knowledge engineering effort is required to reconcile or restructure the current state of knowledge:

1. Changes to specifications.
   - New feature or feature components
   - New attributes for existing features or feature components
   - Changes in attribute values
   - Changes in capture conditions
   - Changes in delineation representation

2. Omissions, inconsistencies, or errors identified via users during productions.

3. Performance enhancement opportunities identified via KBS testing, operations, and maintenance activities.

4. Ongoing knowledge acquisition tasks.

5. Changes in DMA production requirements.

KBS maintenance personnel use existing Ada software to correct KBS errors of omission, and to install upgrades/enhancements. The major piece of equipment used for AFE Knowledge Base maintenance and support is the Knowledge Engineer (KE) workstation. The KE workstation's hardware platform is a DEC 3200 VAXstation configured with DECwindows. The KE workstation is a very robust maintenance tool. It is here that KBS personnel have at their disposal the tools necessary for Knowledge Base rule edit/verification management, database management, and reports management.

KBS maintenance functions supported are as follow:

- Review of Summary Reports and cumulative statistics.
- Review of candidate knowledge base rules and tests.
- Review and modification of the AFE Knowledge Base.
- Exercising the inference process.
- Management of files and reports generated during a KE workstation maintenance session.

Basic functionality provided by the KE workstation includes:

- Loading of rules to be used during the KE workstation maintenance session.
- Edit reports which aid in the editing and addition of rules to the baseline rules.
- Hardcopy which provides the ability to print out rules.
- Revert which allows applied Edit Reports to be removed and returns to the baseline rules loaded.
- Search of the rules loaded for specific information.
- Feature/attribute operations which allow the addition or deletion of codes from the rule base.

- Report generation capabilities for performance analysis. (This is accomplished using Ingres Tools.)

CONCLUSION

The DE/S KBS has already demonstrated that it will provide genuine assistance to the DE/S cartographer in the knowledge intensive extraction processes of feature identification, attribution, and consistency checking. The rules, being for the most part tightly derived from the CPESDB, will help enforce a high degree of consistency in feature extraction. Of the KBS processes, automation of attribution has clearly demonstrated that it will have a profound impact on improving feature extraction performance throughput.

The key software elements such as the inference engine and the relational data base software are well proven and the Ada software that controls the extraction operation is a robust design.

The knowledge acquisition activity is proceeding methodically with particular attention being paid to developing region rules to support near term production areas, spatial association rules to enhance consistency checking, and product specific rules to improve overall KBS performance.

Staffing, training, and maintenance policies and procedures have been established and integrated into the DMA production environment.

Performance assessments of the AFE Knowledge Base have been for the most part restricted to verification. Future testing efforts will focus on designing and developing benchmark tests for validation.

ACKNOWLEDGEMENTS

953