Automating Human Service *Practice* Expertise
ASAP
The Automated Screening and Assessment Package

Susan Millea, Ph.D.
Hogg Foundation for Mental Health
WCH 300
University of Texas at Austin
Austin, TX 78713
Voice (512) 471-5041
Fax (512) 471-9608

Mary Anne Mendall, M.S.S.W.
Mendall Associates
815 Lewiston Drive
San Jose, CA 95136
Voice (408) 978-1386
Fax (408 448-7412

Abstract:

Though the technology has been available for some time, there are few applications of knowledge-based systems (KBS) technology being used in human services. Those systems which have been developed have focused on the representation of complex and sometimes conflicting federal, state and local policies and procedures for service delivery. While the Automated Screening and Assessment Package (ASAP) contains this type of written knowledge, it goes beyond this in the use of AI knowledge engineering methods to capture and represent the more elusive, unformalized and mostly unwritten "practice wisdom" of expert practitioners. The knowledge domain is extremely complex. It involves assessing basic human needs. Knowledge engineering required the participation of multiple experts from varied disciplines and input from potential users and service consumers. Experts had to reach consensus about what knowledge the system should contain to help non-expert practitioners perform holistic assessments of the non-clinical needs of persons with severe mental impairment living in the community. ASAP is being successfully used in Texas. Innovative techniques were also used in the evaluation of ASAP in practice. The research design yielded measurable results demonstrating that use of the tool helped non-experts to approach expert level performance.

Introduction

This paper will present a review of the development, implementation and evaluation of the Automated Screening and Assessment Package (ASAP), a successful human service expert system. Subsequent to that review will be a discussion of what lessons were learned that might be applicable to the development of other innovative systems. ASAP is particularly innovative because:

1) ASAP deals with a complex problem domain involving the assessment of basic human needs such as basic living skills, social supports, housing, income, education/employment, and legal needs, and behavioral issues.
2) The knowledge engineering process required capturing and representing the expertise of multiple experts in multiple disciplines who had to reach a consensus about what the system should contain. Also,
3) An innovative research approach which combined AI research techniques with social work research methodology yielded measurable results demonstrating that use of the tool did indeed improve practitioner performance.

ASAP was developed by the Texas Mental Health and Mental Retardation Department (TXMHMR) in Austin, Texas. TXMHMR is the largest state agency in Texas. It provides services to 35,545 persons with mental retardation and 144,032 persons with mental illness. TXMHMR engaged the services of Mendall Associates to do the knowledge engineering for ASAP. Mendall Associates is a woman-owned and managed consulting firm in San Jose, California. The Center for Social Work Research at the University of Texas at Austin supported the evaluation research of ASAP.

There are few applications of knowledge-based systems technology being used in the human services. Though the technology has been available for some time and has been perceived as potentially useful in human services at least as far back as the 1980's (Henderson 1986; Geiss and Viswanathan 1986), examples of successfully deployed expert systems are limited (Kidd 1992; Simmons 1989). Among the reasons for the lag in the transfer of this technology are a poorly formalized body of practice knowledge (Nurius 1990; Mullen & Schuerman 1988), the expense of the technology, lack of KBS expertise among in-house agency staff, lack of an understanding on the part of AI developers of the complexity of the social service delivery environment (Hasenfeld, 1983, 1992), and disincentives to innovate in public agencies (Ibid.).

The few expert systems which have been developed in human services have focused on the representation of complex and sometimes conflicting federal, state and local regulatory policy; and agency procedures, such as those dictating eligibility for entitlement programs. The *MAGIC program (Kidd & Carlson 1993) and the CLEAN* system (Simmons 1989) are examples of two successfully deployed applications of KBS technology to determine a person's eligibility for public aid. The CAPS program is

an expert system created by the Texas Health and Human Services Commission which also does eligibility screening for 13 state/federally funded programs.

Like the above applications, ASAP assists with the determination of eligibility. It assists practitioners in determining eligibility for TXMHMR services and for case management. In order to determine eligibility for case management, the practitioner is required to identify the individual's unmet needs and assess their impact on the person. Therefore, ASAP, as part of its eligibility determination process, further assists the practitioner in performing a global assessment of the needs of the individual in the areas of housing, income, behavior, basic living skills, work, school, socialization, legal issues and family issues.

Unlike the above applications, the knowledge required to perform an assessment cannot be found in policies and procedures alone. That knowledge comes from the wisdom acquired through practice and through the application of social work knowledge and common sense. What makes ASAP stand out from other human service applications is that AI technology was used to capture and represent the more elusive, unwritten "practice expertise" and common sense of multiple experts with specialized knowledge in different areas of assessment. Demonstrations of ASAP at state and national human service conferences have garnered it praise and recognition for doing what was not thought possible, encoding "practice wisdom." In 1991, ASAP was selected as one of 75 semi-finalists among a field of 1,875 applicants for an award as an "Innovation in State and Local Government", sponsored by the Ford Foundation and Harvard University. It also gained praise at the 1990 International Conference on Human Services Information Technology Applications.

Problem Description and Goals

The Texas Mental Health and Mental Retardation Department is the largest state agency in Texas. It provides services to more than 179,000 individuals through its 27 facilities and 35 community MHMR centers. Services are targeted to those with severely disabling conditions, particularly those whose conditions are complicated by severe health needs; behavior problems; or vocational and training needs to develop skills for independent living.

In 1985 the TXMHMR department moved to a case management paradigm of service as a way of improving service to its consumers and reducing their placements in state facilities. The essence of case management is that a single accountable individual, the Case Manager, performs activities in the service of the client, insuring to the maximum extent possible that the person has access to, and receives, all resources and services needed to achieve and maintain the ability to function in the community. The program is designed for persons with severe, long-term mental illness and for those with mental retardation, regardless of age, who require long-term community services.

By 1987, after extensive training efforts in the methods of case management, quality assurance personnel still found that case managers were not applying service eligibility policies equitably across the state. Also, the quality of needs assessments, which is fundamental to the development of individualized treatment programs for consumers, varied considerably among practitioners. One reason for the variance was that case management staff came from backgrounds ranging from high school graduate to masters level social worker. Another problem was a high staff turnover rate due in large measure to low pay and limited opportunities for advancement. A study done in 1986 showed a 39% turnover rate among case managers within the previous year. This rate was considered by case management supervisors to be typical. Significant, ongoing training was clearly required to maintain quality of care. However, Texas is an exceptionally large state and geographic distances increased the costs of statewide training due to the travel expenses. It was for these reasons that two key administrators, Sally Anderson, who was Director of TXMHMR Information Services and Janet Collins Wright, Coordinator of the TXMHMR Case Management Program turned to AI technology as a way of capturing and delivering scarce expertise to the novice service provider at the point of service delivery when that knowledge was most valuable. These two administrators were the primary "champions" of the technology throughout the development of ASAP.

The early articulated goals of the ASAP project were to:

- provide a more equitable approach to determining those most in need of services;
- develop more comprehensive, better quality assessments of individual consumer need;
- create an opportunity for staff training at the time that training is most valuable.

Once administrators and potential users could react to an early system prototype, they came to appreciate that additional problems could also be addressed if ASAP were connected to a database. While TXMHMR has a statewide client tracking system called CARE, the type of information in CARE is mostly demographic. CARE cannot provide information about the specific needs of persons served. Therefore, the needs of the target population must be estimated. A database of actual consumer needs would be invaluable in making decisions about the allocation of funds and the direction of advocacy efforts. Also, a key concern of potential system users was that they were "drowning in paperwork." In ASAP they saw an alleviation of the paperwork nightmare that now encompasses all of social services. Once information is entered into ASAP, that information can be easily modified and an updated report can be generated. Based on these responses to the early prototype the following

additional goals were added to ASAP:

- to provide more accurate identification of actual needs of persons served;
- to reduce paperwork, particularly for direct service staff.

All of the above goals were in line with the mission of TXMHMR which is to provide quality care for service consumers.

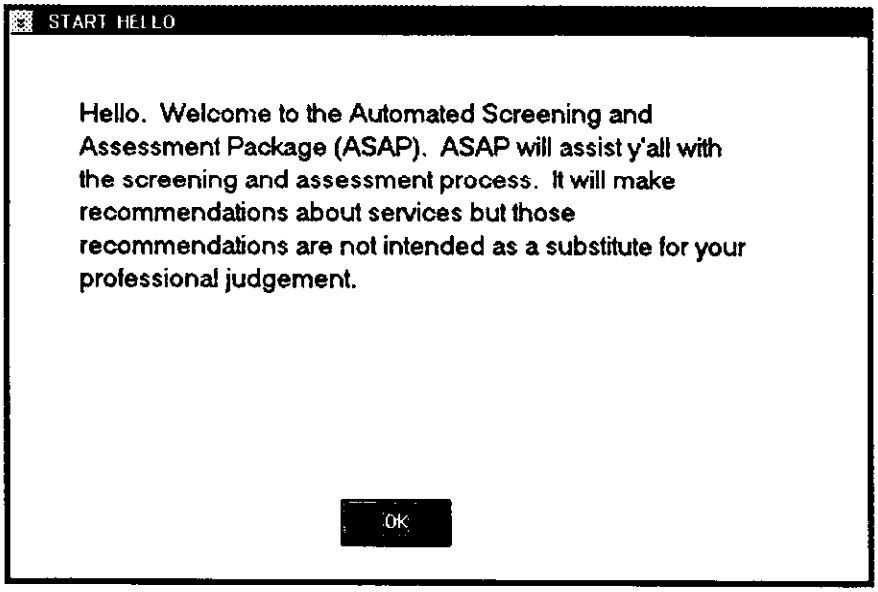

Figure 1. ASAP's Welcome Screen

A Description of the ASAP Application

ASAP is an interactive system which guides the user in performing a global assessment of basic human need. This then helps to determine an applicant's eligibility for TXMHMR case management. Though the policy-based eligibility screening portion of ASAP is useful, it is the ability to guide the assessment process that is emerging as the truly beneficial aspect of ASAP because it brings the expertise of multiple experts to the fingertips of novice practitioners.

ASAP relies heavily on an easy-to-use graphical user interface which delights most users. Social work practitioners as a whole are not computer literate and many are fearful and mistrustful of the computer. Yet, ASAP users quickly overcome their fears after minimal interaction with this friendly interface. You can see the programmer's Southern pride demonstrated in the choice of language on the welcome screen.

A value concern among practitioners in human services is the locus of control when information technology is applied to social work practice. A similar concern has been documented in the use of medical expert systems. As this screen indicates, though the system makes recommendations based on its rule structure, the practitioner may disagree and pursue an alternative course of action, after providing a rationale for this decision. This ability not only keeps the practitioner in ultimate control and accountable, it provides "passive" capture of practice expertise to further assist in knowledge development, and for purposes of professional supervision (Pruger, 1986).

The ASAP assessment process is more than a series of checklists. It is an interactive process which acts on the information entered by the user. Answers to questions determine the flow of subsequent questions. Early in ASAP's development, a paper version of the system was created to allow for preliminary testing of the knowledge base. The paper version required an interviewer to function in the role of the computer. The interviewer had to be fully briefed on all of the rules of the system and on the flow of the questions. Even then, the interview was extremely awkward because it required numerous page turning, frequent stops to calculate the latest list of needs specific to the individual being assessed and occasional consultations with a thick packet of documentation containing all of the systems rules. It is a gross understatement to say that the paper version was not

practical. The reason it was not practical is that the PROCESS through which the system guides the user cannot be represented in a document of manageable size nor can a paper document make the calculations which help the user to manage the assessment process.

ASAP guides the user through a series of questions about each need area in order to identify the full range of the individual's needs. The user is able to use a summary screening format or a detailed assessment format to perform this task. The user can toggle between the two formats at any time.

ASAP creates a list of the identified needs and asks the user to rate the current impact of each of those needs on the individual. It then helps the user to identify the individual's support system and to assess the effectiveness and stability of that support system relevant to each of the significant needs. If the system is unstable, the user is asked additional questions about the likely impact of lost supports. Finally, ASAP calculates a list of UNMET needs containing only those needs with a significant impact which the current support system is NOT effectively addressing. These needs are prioritized by severity.

ASAP further calculates a set of recommendations concerning eligibility for TXMHMR services, for case management and also for a handful of other referrals (e.g. to a substance abuse counselor and to vocational services). It also gives the user a list of "unknown" need areas, which have not been evaluated but probably need to be, for further exploration by the practitioner.

When the user has finished interacting with the system, the data is saved and a report is generated for insertion in the person's chart. It is the successful combination of an expert system, a graphical user interface, a database and a report generating mechanism that makes this system useful in practice.

The Knowledge Engineering Bottleneck

The biggest challenge in developing ASAP was the knowledge engineering required for this complicated domain. ASAP's domain is very complex: it involves understanding the needs of human beings. What are the basic needs of human beings? How does one recognize when basic needs are not met? There are theories of knowledge about this domain, but no definitive,

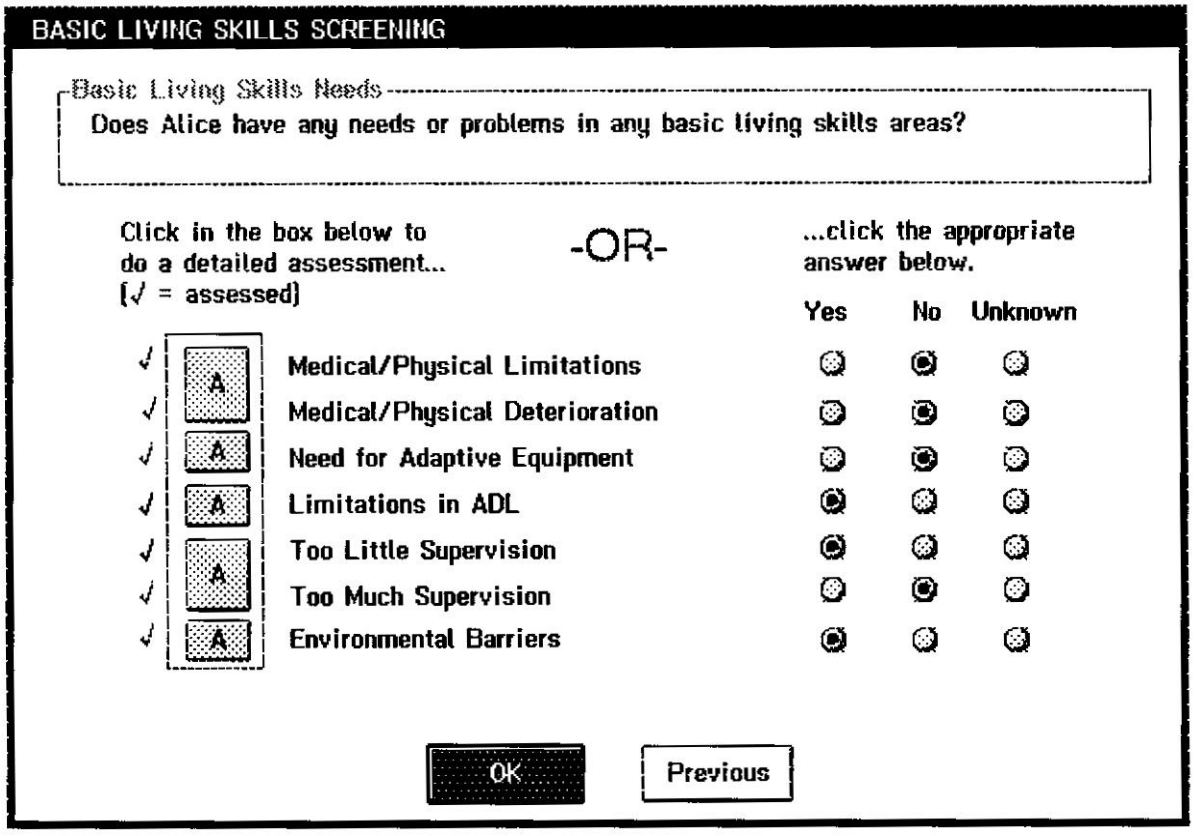

Figure 2. Basic Living Skills Screening, showing sub-components of assessment (ADL means activities of daily living)

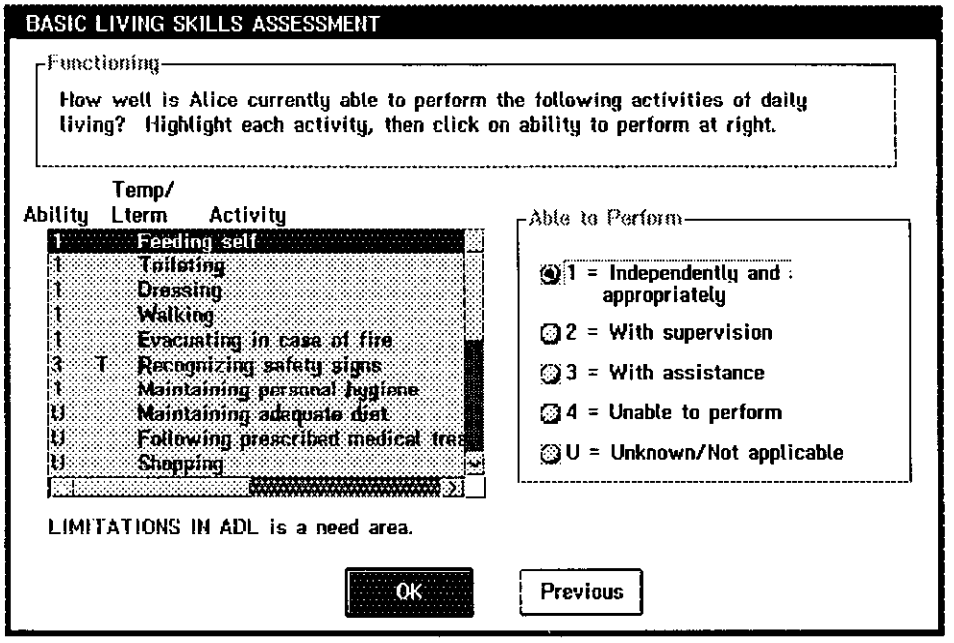

Figure 3. Activities of Daily Living Assessment

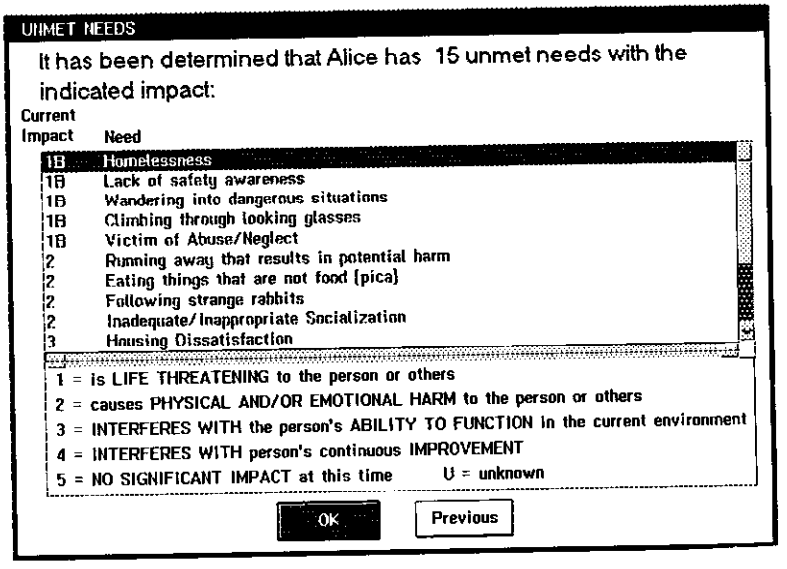

Figure 4. Rating the Impact of Identified Unmet Human Needs

written approach to identifying and assessing the severity of human problems. The challenge was that a knowledge base had to be developed.

In 1988, TXMHMR hired a knowledge engineer to develop the ASAP knowledge base. The knowledge engineer was a social worker with expertise in the assessment domain. It soon became apparent that no one expert or even a handful of experts had all of the knowledge required to understand this complex domain. Substance abuse experts were not necessarily experts in housing. Over the course of the two years when knowledge engineering was the primary focus, TXMHMR provided 26 experts in domains like mental health, mental retardation, case management, vocational services, housing, employment, substance abuse and education.

Experts were also needed who understood the rationale

for policy development and experts who worked with consumers on a day-to-day basis. The team of experts assembled included policy developers, statisticians, case managers, case management supervisors, and information services personnel. Each of the experts understood his or her own individual areas of expertise very well. The danger was that a knowledge base developed by multiple experts could appear fragmented; the result would be an assessment that was superb in assessing specific pockets of need, but did not capture the *integrated* needs of the whole person.

Blending the expertise of these disparate and often opinionated disciplines required iterative interactions among the experts over the course of 2 years. The knowledge engineer met with identified experts singly and in small groups to capture their expertise. Case examples were often used to elicit nuances of the knowledge base. Tape recordings were made of all the interviews and those recordings were studied carefully and used to develop a written representation of the knowledge. The knowledge representation was then reviewed by the expert(s) who supplied the information and revised iteratively until both knowledge engineer and expert were reasonably comfortable with the results. Then the knowledge representation was reviewed and discussed by the group of experts. This group met every other month. Typically, about 10 experts would be present at a time. They would debate about the impact of the individual area of assessment on the assessment of the person as a whole and would challenge each other to think beyond their own domains. The process resulted in a synergy. The integrated knowledge base of human needs assessment was more than the sum of its parts because it provided a holistic assessment of the individual.

The written knowledge base (KB) which the experts reviewed contained the appropriate assessment questions in the specific order in which they should be considered plus a set of IF...THEN statements which contained relevant federal, state and agency policies, assessment expertise, common sense heuristics and instructions to the computer to determine the flow of the interaction with the user. For example,

> **IF** a person is 18 years of age or older
> **AND** the person has a primary diagnosis of schizophrenia, bipolar disorder, or major depression
> **THEN** the person is eligible for TXMHMR state-funded mental health services.

The policy rules were the easiest to develop because most of them were written down. However, even they were not without difficulty because what was written was often vague and not quantifiable. Therefore, much of the knowledge engineering process involved working with policy makers to refine and quantify their policies. For example, one mental health (MH) policy required that persons who do not have a major diagnoses must have a "functional impairment" to be a member of the priority population. Working with the knowledge engineer, the policy makers specified functional impairment as an assessment of 50 or less on the widely used Global Assessment of Functioning Scale. This not only made it easier to code the policy, it also clarified the policy for users. Even now as policies change, ASAP is challenging policy makers by bringing to their attention policies which are vague or inconsistent with other policies.

More complicated to define were the rules containing assessment expertise or "practice wisdom." Experts frequently are not conscious of what they know and need help in specifying the exact process they use to make decisions. Case examples were often used to help experts describe the specific steps they take in assessing particular problems. Interviews were followed by group meetings in which the experts challenged each other and helped each other to move beyond their exclusive domains. Case manager participants were particularly helpful in reminding other experts of the effect a specific approach or certain language might have on an actual consumer. They often brought to the table the most complicated element of all: the element of common sense.

Like common sense, there is nothing simple about *practice* sense, the unarticulated knowledge of expert practitioners. This is the type of knowledge that experts are least aware they are using and find most difficult to discuss. The assessment of social needs was one of the most excruciating to define because it is imbued with this type of knowledge. How does one know when a person's social needs are unmet?

During the knowledge acquisition phase, individual experts presented approaches which asked about the number and quality of social contacts in a given week. Arguments ensued about whether telephone contact should be counted equally with physical contact. What about the fact that different people have different levels of need for social contact? The definition of "normal" eluded everyone. Finally, common sense came to the rescue when one expert introduced the concept of the need for friendship and interpersonal support. We all need at least one friend, one human being with whom we share a trusting relationship. This is perhaps a universal human need. But more than common sense was involved in this process. *Practice* sense was exhibited as the rules were specified. For instance, experts questioned the applicability of the rule

> **IF** a person has NO friends (someone with whom the person shares a trusting relationship)
> **THEN** lack of friends is a need area.

to certain consumer populations, particularly paranoid schizophrenic individuals and possibly some persons with profound mental retardation. These persons might not have the capacity to form truly trusting relationships. This does not preclude them from having supportive social relationships. The rules were re-worked as follows:

IF a person has no friends with whom he/she shares a trusting relationship or a relationship in which both parties enjoy spending time with one another
THEN lack of friends is a need area.

IF a person has no close family members with whom he/she shares a trusting relationship or a relationship in which both parties enjoy spending time with one another
THEN lack of close family members is a need area.

IF lack of friends is a need area
AND lack of close family members is a need area
THEN insufficient interpersonal support is a need area.

IF the person has friends or close family members with whom he/she shares a trusting relationship or both parties enjoy spending time together
AND the person seldom or never shares pleasant times with friends or close family members
THEN insufficient interpersonal support is a need area.

IF the person has friends or close family members with whom he/she shares a trusting relationship or both parties enjoy spending time together
AND the person can count on friends or close family members to provide effective and appropriate assistance in times of need only sometimes, seldom, or never
THEN insufficient interpersonal support is a need area.

After numerous interviews with multiple experts and groups of experts, the core knowledge was ready for the real test: the test of whether it could actually be useful to real practitioners in assessing the needs of real consumers. A paper version of the knowledge base and an early system prototype were used in interviews with practitioners, consumers and family members at 8 different sites across the state to test how well use of the knowledge base assessed the needs of actual consumers. More than 80 practitioners, consumers and family members from four state facilities and four community centers critiqued the knowledge base, modified the questions, adjusted the language and brought their own expertise to bear on the knowledge base. The users at times had expertise which our original group of experts lacked (e.g. in the areas of legal issues, health care and assessment of children).

Consumers and family members identified two areas of need not addressed in the early knowledge base: legal issues and client rights. Furthermore, they made the language more sensitive and insisted on the addition of clarifying questions which guarded against stigmatizing labels. For example, ASAP asks if the person exhibits certain behaviors. Among them is "inappropriate sexual behavior." Consumers insisted that if that behavior was identified, the user should be asked to name the specific offending behavior. Was it rape or inappropriate hugging and kissing? The user should also specify when the most recent incident occurred and whether it was likely to be repeated. An inappropriate kiss delivered 20 years ago and not likely to be repeated is considerably different in meaning from a rape committed last month and likely to be repeated.

The involvement of practitioners, consumers and family members added usefulness and sensitivity to ASAP. They gave ASAP the values of the social work profession and a sensitivity to the human condition.

The knowledge engineering for ASAP is an ongoing process. Recommendations are actively solicited from sites which are using the system in their day to day operation. Those recommendations are carefully tracked by the knowledge engineer and prioritized by all of the users. High priority recommendations are then followed up by consultations with experts and adjustments to the knowledge base. The knowledge engineer is currently working with health experts to develop a separate and more elaborate health assessment for inclusion in the third release of the system.

ASAP is an expert system. Like a human expert, it must continue to evolve and adjust its knowledge base in response to changes in policy, improved practice wisdom and common sense. Its long term success as an expert will depend on its ability to grow in its knowledge .

Development of the ASAP Computer Program

Timelines

1988 - 1990	• Development of the Core Knowledge Base
1988 - 1989	• Development of the first prototype using an expert system shell
1989	• Field Testing of the Core Knowledge Base • Evaluation of the first prototype and expert system shell
1990 - 1992	• Reprogramming of the system in an OS/2 environment using Application Manager and the SQL Server database • Iterative testing and debugging
1991	• Establishment of the ASAP Steering Committee appointed to oversee the implementation and ongoing refinement

	of ASAP
1992 - 1993	• Deployment of ASAP: Version 1 at MH and MR field sites • Evaluation of the effect of ASAP use on the quality of assessments done at field sites
1994	• Deployment of Version 2 which contains increased functionality and policy changes • Deployment at additional field site • On-going knowledge engineering, testing, and maintenance

Rationale for Development Approach

Concurrent with the development of the core ASAP KB, the first ASAP prototype was developed using an expert system shell from Texas Instruments called Personal Consultant. Enough knowledge was encoded in the shell to demonstrate the capability of expert system technology and to assess the ability of this software to represent the knowledge. This was in 1989 when expert system shells did not have quick and seamless links to databases and graphical user interfaces (GUI's). Database calls were time consuming and the database screens differed markedly from screens created within the shell.

The complexity of the knowledge base made it difficult to represent adequately in the selected expert system shell without customizing the Lisp code behind the expert shell. Limited understanding of expert systems technology existed in the agency, and no programmers knew Lisp. Limited funds made it difficult to purchase and test multiple expert system shells, or to hire an expert systems programmer. This type of technology was unproven in human services. Developing the needed skill to adapt a shell would have required considerable staff time for an unknown payoff, which was anticipated to benefit only this one information system. Meanwhile the agency had purchased a graphical user interface development tool (Application Manager) which was expected to be of broader utility to the organization. Programmers were already expected to learn this program. The efficient organizational decision appeared to be to proceed with this latter development tool, have programmers learn to use it, and using the available expertise in database systems technology, create a system which emulated the functionality of an expert system.

System Code and Functionality

The result of this organizational context was that programmers were challenged to develop an expert system without the advantage of the tools that come with an expert system shell. They have largely succeeded. ASAP returns the rule-like explanation for a recommendation made. It has a non-deterministic control structure, and it measurably improves user performance.

By 1990, TXMHMR Information Services had selected a system architecture in which to develop all of its field applications. That architecture included a LAN using a client server architecture. Applications would be designed to run under the OS/2 operating system using the Sybase SQL Server database. The database would connect to the CARE database (the client tracking system) so that users could download information from CARE into local applications.

Since an easy-to-use interface was critical for staff with little computer experience, a graphical interface development tool called Application Manager (AM) was selected to insure the easy development of a friendly user interface. While Application Manager was not an expert system shell, developers believed it had the capacity to allow programmers to build a program that could function as an expert system.

The decision was made to re-program ASAP in this environment.

> **IF** a person is 18 years old or older
> AND has a diagnosis of schizophrenia, bipolar disorder, or major depression
> **THEN** the person is eligible for TXMHMR state-funded mental health services

appears, in part, in AM as

Test: INFO{0}.MH_ASAP_Schiz=1
and
Assign: INFO{0}.Elig_MH_St_Funds$:="y".

Test: INFO{0}.MH_ASAP_Bipolar=1
and
Assign: INFO{0}.Elig_MH_St_Funds$:="y".

Test: INFO{0}.MH_ASAP_Majdep=1
and
Assign: INFO{0}.Elig_MH_St_Funds$:="y".

An attempt was made to collect a representation of the rules, policy rules, summary statements and recommendations in a separate area of code for programmer use. When a policy needs to be changed, the programmer can do so by changing the code in this area. The result is that the change is then automatically made to all of the other areas of code affected by the change. This allows the programmer to avoid having to search the code for the location of the exact procedure to be changed. Although policy changes are much easier using this process, the code must still be exhaustively tested to make sure no unintended results occur, such as changes in screen flow or unexpected affects on other rules. They must also test to insure that no message statements to the user are out of synch with the new rules.

Use of an effective expert system shell would have allowed a clearer separation of the rules from the code. A shell would also have automatically generated user explanations (message statements) to the user directly

from the rules. A shell was not used because 1) the shell selected was not adequate for the task, 2) the cost of high-end shells for experimentation was considered prohibitive both in terms of their purchase price and in the staff time needed to learn a new program, and 3) those costs interfered with organizational expectations that programmers develop skill in Application Manager. This agency went with what it could afford and with what it knew how to do best.

The selection of AM as the development tool had advantages. It saved the time and expense required to purchase and test expert system shells, most of which were still evolving. It cut the cost of the end product for the user by cutting out the cost of runtime versions of an expert system shell. It was relatively easy to quickly develop colorful and artistic screens that invite even the most timid user to interact with the computer. It was efficient in that it allowed programmers to learn a development program in which their expertise was expected to be developed. It took advantage of what the agency did best which was conventional programming. Finally, it resulted in the successful deployment of a system which shares scarce expertise with end users.

The disadvantages were that it took exceptionally clever programming to get this tool to look and act like an expert system. It does indeed function as an expert system. It returns a rule-like explanation when a recommendation is made, it has a non-deterministic control structure and it measurably improves performance of its target users. This is a tribute to the skill of the two programmers who coded ASAP, Barbara Holub and Dan Kern.

However, much of the system's intelligence remains imbedded in the program code. This makes modification a cumbersome task requiring considerable testing and debugging. Maintenance of ASAP will be more time consuming and difficult than it would have been if an expert system shell had been used. ASAP may eventually have to migrate to an expert system shell.

The key point is that without this approach which used available resources, ASAP might not exist. The human service literature on expert systems integration suggests that the focus of work needs to be on development of the knowledge models to be encoded, rather than the technical development of the programs, since that expertise exists in other fields and can be adapted (Nurius 1990). The approach to development chosen by the agency allowed the project to proceed, to focus on the knowledge acquisition, and to develop a system which could help garner additional support both inside and outside the agency to secure additional resources for further development.

Deployment

ASAP was first installed in the Community Services division of a State School for persons with mental retardation in a multi-racial, rural Texas community. Five Case Managers working with community based clients in about a fifty mile radius from the state school were the first to use ASAP. At the end of nine months, practitioners were able to determine whether they wished to continue using ASAP. Not only did they elect to do so, they have become advocates for the tool. The first site has now been using ASAP for nearly 2 years. Developers have installed ASAP in two additional sites in Texas. The latter of these two installations is testing ASAP's applicability to assist multi-agency collaboration in serving the homeless mentally ill. In addition to these installations in Texas, ASAP is being used in a related research project at 2 sites in California[1]. This research is expected to lead to the development of a broadly useful product for human service needs assessment.

Evaluation and Use

The dominant feature of a KBS that makes it hard to validate is that it is both a piece of software and a model of human knowledge and reasoning in a particular domain (O'Keefe & O'Leary 1991). Even if one can verify that the system works correctly (Gasching et al. 1983), this does not mean that the right knowledge was encoded (O'Keefe & O'Leary 1989). Knowledge-based systems are evaluated essentially to determine 1) whether a particular system works right, and 2) whether it is the right system. *Verification* tends to relate to whether a system works right. *Validation* addresses the question of whether this is the right system (in McGraw & Harbison-Briggs 1989). Establishing that a given program is the "right system" is a less straight forward and frequently more difficult task. The evaluation of ASAP included both verification and validation.

The discussion which follows focuses on the question of validation because this is another area where we believe ASAP was innovative and can contribute to the field of AI. The limitations of the technical product were recognized, but using the resources available within the agency allowed the ability to test whether a KBS-encoded model of practice knowledge could assist this area of human service delivery. The feeling of designers was that the real strength of ASAP was in the knowledge. If this could be demonstrated, even on a small scale, resources could be leveraged to further develop ASAP technically. This was strategic thinking.

Though other methods for validating KBS systems exist, the primary methods currently used in practice are the prototyping of systems, letting users react to partially developed systems and obtaining their feedback; and the application of Turing tests in which reports generated by the computer and those of human experts are evaluated by experts in a blind study to see if the computer can match expert performance (Buchanan & Shortliffe 1983). Another measure of effectiveness is the quantification of payoffs generated by KBS use. This helps establish the

[1]This research is being funded by NIMH.

utility of a system.

Like KBS validation, the validation of human service innovations also involves testing in practice. The purpose is to determine not only how well a new intervention works, but under what conditions and for whom (Thomas 1985; Rothman 1980). For instance, verbal psychotherapy may be an effective and cost efficient treatment for persons with some diagnoses such as mild reactive depression, but may be ineffective or require adjunct medical intervention for persons with other diagnoses, such as manic-depressive illness or schizophrenia. Establishing validity in this field is typically a lengthy iterative process which *must* occur in practice because of the complexities of the raw materials of human service practice: They deal with human social systems (Hasenfeld 1983).

Both the product and context are essential in establishing the *validity* of human service interventions (Rothman 1980). The importance of context is recognized elsewhere, but it is typically considered important for the successful *deployment* of proven innovations, including information systems (Kling 1993; Gibson & Smilor; Tornatzky & Fleischer 1990). In social work, *context contributes directly to the actual validation of a new technology* (Rothman 1980, Thomas 1985). The KBS literature, which is calling for an openness to new methods for validation (Ayel & Laurent 1991, O'Leary 1991), seems to recognize the importance of context to validation (Partridge 1986, in McGraw and Harbison-Briggs 1989). Drawing from a rich inter-disciplinary literature (Millea 1994), the evaluation of ASAP addressed the following questions.

- Did ASAP meet user needs?
- Was ASAP effective as a social service intervention?
- Did the ASAP technology function as intended?
- Did ASAP meet organizational needs?
- Was the context appropriate for the innovation?

Each of these will be addressed below.

Did ASAP meet user needs?

ASAP met user needs. Through semi-structured interviews and questionnaires it became clear that ASAP was perceived to meet practitioner needs in nearly all areas identified prior to implementation. The anticipated benefits of using ASAP were all perceived as having met or exceeded expectations. These included reducing redundant tasks, saving time, improving accuracy in client screening, and obtaining help in generating new ideas to assist clients. For difficult cases all of the assessment domains in ASAP were rated as "very useful". Further, of all problems /concerns raised by Case Managers prior to implementation (such as losing data, forgetting to save work, wasting time, getting lost in the program), most did not occur at all and only two were rated as "a minor problem".

Among the perceived impacts on performance, Case Managers found themselves "definitely and consistently" going into more depth in considering client need areas, and "frequently" identifying more client supports, more barriers to service, more needs, and documenting these regardless of the resources available to meet the needs. Clearly, more and better quality information was being obtained to describe the actual needs of the service consumers. These were the hoped for benefits of experts involved in system development.

ASAP has also been found in three successive implementations to be useful in helping practitioners determine when a client no longer needs Case Management services. This latter finding is important because it supports an argument that ASAP assists the development of practitioner judgment, and does not simply lead to an over-identification of consumer needs (and possible overuse of the service system).

Was ASAP effective as a social service intervention?

The Case Managers all clearly perceived that ASAP was improving the quality of their assessments. This "exceeded expectations" according to questionnaire data. In addition to measuring these users' perceptions, the actual change in their behavior when using ASAP was also measured. This involved the use of single subject methodology, an appropriate quantitative methodology for small scale studies (Bloom & Fischer 1982; Hersen & Barlow 1976; Jayaratne & Levy 1979). This approach provided a key outcome measure of ASAP's effectiveness in practice, and has been replicated.

The performance of non-expert practitioners was compared to that of experts over time and across the same actual client cases to determine ASAP's impact in guiding their behavior. All Case Managers had been trained with the TXMHMR case management curriculum prior to the implementation of ASAP, though their levels of experience varied. Two behaviors were evaluated: The number of needs identified and the specification of those needs. Case Managers and Experts completed assessments using the same written case materials. They did not discuss cases with one another.

When using ASAP, Case Managers approached the experts in the number of needs they identified. This improvement diminished when ASAP was removed from their use, indicating it was ASAP and not some other factor influencing the change in behavior. More importantly, Case Managers improved in the *actual specification* of consumer need, from identifying roughly 30% of the same needs experts had identified for the same cases, to identifying about 70% of the same expert-identified needs. This 70% agreement with experts actually exceeds the level of agreement among experts considered acceptable in studies of mental health assessment. It is approximately equal to the level of

inter-rater agreement achieved by the experts in this study. The non-experts were performing at approximately expert level with ASAP. These findings were replicated.

Figure 5 exhibits the performance of one of the Case Managers over time. All five of the Case Managers improved in their agreement with the experts about what the consumers' needs were when using ASAP. For 4 of the 5 Case Managers the level of this improvement was significant ($p \leq .05$). Case Managers have voluntarily continued to use ASAP in their work for nearly two years.

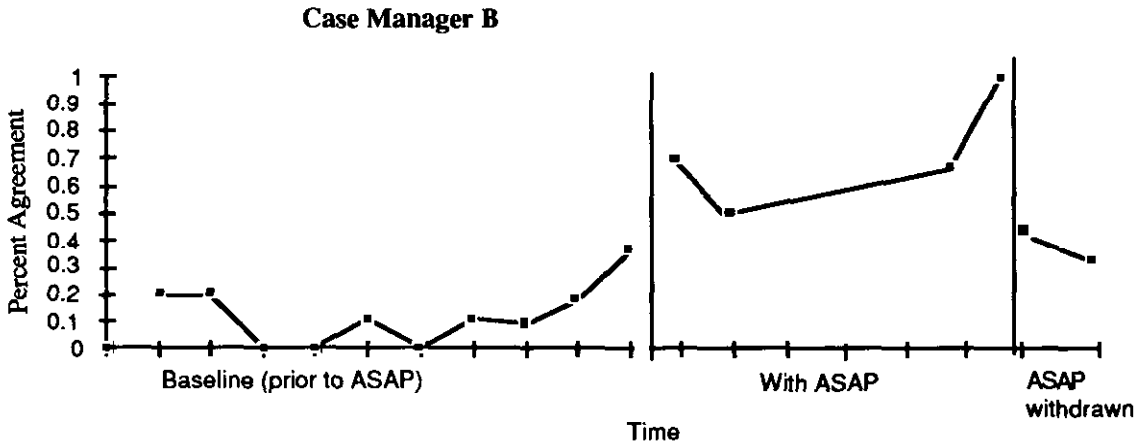

Figure 5. Case Manager B Percent Agreement with Experts In Single Subject Study Using Withdrawal of ASAP Treatment

Did the ASAP technology function as intended?

Clearly, ASAP was a success with its intended users. Case Managers liked the tool. Outcome measures of their performance revealed its effectiveness as a service intervention. It was achieving desired results. But these findings did not address other areas of the technology, such as the functioning of the computer program. For this, a log of error reports and user recommendations was maintained and analyzed.

The Information Services Department evaluated these reports as either errors or enhancement requests. An alternative analysis was able to distinguish comments about knowledge content from those regarding the computer representation of that content, and issues of system functionality from those of the KB. The result was a quantification of results across these categories which highlighted both the resiliency of the knowledge model and the difficulties of the program code. This in turn guided an organizational decision to support re-coding of the program. Though developers encountered a number of technical problems after initial deployment, and re-coding has been required, the innovation overall was very positively evaluated.

The opportunity to re-code also reflects the learning of the programmers, who were neither trained nor experienced in expert systems technology, but were adept at database development. They have adopted aspects of AI development techniques in their work such as attempting to separate the system rules from the operational code. They have been able to successfully modify system rules to maintain the system.

In addition to the log of errors and recommendations, a Turing test was conducted in which experts were presented with case materials. For each case, experts were to select the better of two coded needs assessment reports, one of which had been produced by an expert using ASAP and the other by a human expert. The process of conducting the Turing test generated an important discussion among experts about the trade offs in report content in general. Experts prefer greater detail about the particular individual because it helps them serve the consumer better. However, recording sensitive information about persons with mental impairment is a value-laden and politically sensitive process. It has the potential to negatively portray vulnerable individuals. Without knowing the source of the reports, the experts selected the ASAP-produced reports as the better quality needs assessment report, because of the tempered language. The *preferred* reports, however, were those which were more descriptive of the consumer. These reports had been produced by human experts. The tempered language, while more appropriate for a client record also sacrificed descriptive richness. This in turn made serving consumers more difficult.

The relationship between the KB and report output became more apparent in the second installation when quality assurance (QA) personnel were not familiar with the ASAP program, as they had been in the first site. ASAP was initially found to be inadequate by these reviewers, based on ASAP-generated reports in the client record. When QAs had access to the ASAP program, they

determined that the program actually functioned quite well for its intended purpose. The KB was sound. The reports are adequate for persons familiar with the ASAP program. The difficulty was that for persons unfamiliar with the program, the reports did not reflect the process by which needs were identified. This difficulty in the reports could be rectified, it was determined, by having reports pick up the rule statements which reflected the practice logic. Accomplishing this is testing the limits of the development tool functionality. The report writer for AM is not very flexible for this type of program.

For the *non*-expert practitioner, ASAP provides a very good tool. It captures a comprehensive description of the individual consumer while avoiding negative connotations of the individual in its reports. Demonstrating both the abilities of the knowledge model and the limitations of the development tool are helping to build organizational support to secure additional resources which could be used to re-program ASAP. This is a strategic approach to innovation development. In the meantime, ASAP remains a useful, functional and maintainable system.

Did ASAP meet organizational needs?

ASAP functions as an expert system. It shares previously scarce expertise in remote areas of the state. It provides a non-deterministic control path, returns the rule-like explanation invoked when a recommendation is made, and has measurably improved the performance of users in the direction of the expert. Had evaluation focused entirely or primarily on the technical aspects of ASAP's functioning, the project likely would have failed.

Instead, information about the technical functioning of the tool, along with its impact in practice was provided to organizational decision makers, who were then observed by the external evaluator. Their decision to support further development or stop the project was considered an indicator of ASAP's validity. The strength of the outcomes, e.g. the payoffs, had to be sufficient to garner support from key organizational decision makers.

ASAP succeeded at the organizational level. It received continued support for development in the face of legislative funding cutbacks in excess of $1 million. Organizational decision makers perceived technical functioning to be secondary to the importance of the knowledge model in the first implementation. ASAP appears to have succeeded organizationally because it was consistent with the organizational mission and because it was considered a prudent use of organizational resources. It also had the support of two organizational champions who perceived the utility of this tool for quality improvement in social service delivery under a total quality management (TQM) paradigm.

Was the context appropriate for the innovation?

One reason why innovations which appear to function well fail when introduced into practice has to do with the context in which they are introduced (Gibson & Smilor 1991; Fleischer & Roitman 1990; Rothman). That context can perhaps best be described in terms of the *communication* context. It includes things such as the level of administrative and technical support for getting an innovation in place, and supporting the users while they are learning to use the new system; electronic communication factors, professional values of practitioners, cultural factors and geographic factors. In introducing ASAP into practice, the context was observed in order to understand whether, in the event of failure, inadequacies of the communication context rather than shortcomings of the ASAP program were at play. All of these factors noted were observed to influence the outcome of ASAP's effectiveness determination.

Communicating about the technologies involved in ASAP literally required programmers and practitioners to learn one anothers' language. The difficulty of the communication process was exemplified by the group's decision to bring in a group facilitator. The facilitator guided the process which fostered cross-disciplinary learning. Had the group not addressed the communication issues it is possible that the project might not have survived, despite findings that ASAP was effective in practice. Tracking these processes helped to emphasize the importance of organizational communication in the development and implementation of expert systems which cut across disciplinary functions.

Use and evaluation are linked. Current research is focused on the use of ASAP in other service settings, each of which has slight variations from previous implementations. Evaluation occurs at the individual practitioner level and at the level of the service organization. We are collecting a 'critical mass' of client cases in the database. As that database develops, we will be able to test it to assess the usefulness of ASAP for guiding state planning. Eventually, it is believed that evaluation can be based on client outcome measures which this tool will help to identify.

Payoffs

ASAP's earliest payoff was that the development of the knowledge base encouraged policy makers to rethink and refine their policies. One of the policy makers involved in ASAP's development, an expert practitioner, had such an aversion to computers that she walked out on a presentation encouraging the use of computers in social services. Initially skeptical about the ASAP project, she became a strong advocate because of the effect that KB development had on the policies of the agency. That process required a clear statement of policy rules so that they could be encoded. In her words, had the project gotten no further than this, it would have been an accomplishment. The human service literature supports this stance on the potential payoffs of expert systems in practice (Mullen & Schuerman 1989; Nurius 1990).

Another payoff was that a Needs Assessment

Knowledge Base was developed and is continuing to evolve. The expertise of the original team of 26 experts has been enhanced by expertise from other experts and from well over 100 field staff, consumers and family members. Having captured that expertise, practitioners have something tangible to react to which fosters critical thinking about the processes involved in assessment. This has observably heightened awareness of practitioners who use the system.

ASAP has improved the quality of the needs assessments of non-expert practitioners in the actual practice environment, when compared to identified service experts. Improvement was demonstrated both in the total number of needs identified and in the specification of those needs.

ASAP has also increased the efficiency of direct service staff. Assessments, which used to take 1 1/2-2 hours, were reduced by half to about 45 minutes by the end of the study. This has been replicated in other settings. The production of a completed client assessment report is now nearly instantaneous, reducing dependence on clerical assistance and a delay of up to 10 working days to receive a prepared report from a word processing pool. Quality assurance personnel have approved the reports generated. Also, because information is retained in a database, updating reports takes only a few minutes. Previously, annual reviews of client progress required up to two hours because the entire report had to be replicated.

For the first time, a database of the actual needs of service consumers, assessed in a consistent manner, exists for this agency. As that database is developed it can provide the information needed to assess the effectiveness of services based on consumer outcomes. It can also provide information which was previously inaccessible about the service-seeking patterns of consumers, and can help direct the effective allocation of resources.

Additional payoffs include the development of processes for the acquisition of human service practice knowledge, and of methodologies for evaluating these innovations in human service practice. Both have been used successfully in additional work toward the integration of KBS technology in human service practice.

The model of human needs assessment appears to have utility in a broader human service setting, even though that part of the knowledge base containing local agency policies (and related state and federal regulations) may not be useful in other human service organizations. If a method of assessing need and screening for service eligibility can be developed which is useful in multiple agency settings it can help agencies collaborate to serve persons with multiple and complex needs such as those typically served by the Texas Mental Health and Mental Retardation Department. This can help bring about a "one stop shopping" concept for service consumers, with the potential for both improved quality of services and greater efficiencies in their delivery. Though still potentials, these are powerful payoffs.

Maintenance

Human service providers and administrators have become increasingly aware of ASAP's potential to assist service provision. Because ASAP is seen as a successful service intervention, it has been reshaped throughout its development to work in increasingly broad environments. The purpose of the tool has changed over time. Once seen as a training tool for case managers in the process of eligibility determination, it is now perceived as an interactive process tool for intake by all direct service staff; as a tool for completing comprehensive assessments; and for efficiently producing related reports. ASAP has not yet attained a purely "maintenance" mode. Some understanding of maintenance requirements is emerging, however.

ASAP will undoubtedly require the part time attention of the two programmers who developed the program, and the programmer who has developed the reports, Bill Hargrove. Agency policies change annually and those changes will need to be encoded. The practice knowledge for assessment is also changing. Like any expert, ASAP will need to keep up with the latest practice knowledge as it emerges or it will quickly become obsolete. As the ASAP database grows, there will undoubtedly be requests for both routine and ad hoc reports to help inform policy decisions. It has yet to be determined whether these reports will be developed by the ASAP development team or in-house by technical staff at the user sites.

ASAP will also continue to need the attention of the ASAP Steering Committee. This group consists of some original members of the development team, service program staff and administrators, information services personnel, several Assistant Deputy Commissioners for mental health and mental retardation services, the Knowledge Engineer and external Evaluator. This committee is responsible for implementing the system statewide. The Steering Committee has formed several task groups which are taking responsibility for things like developing standardized testing and implementation procedures, and developing a users' group.

An informal users' group is beginning to form and there are plans to support this structurally within the agency. A mechanism has been developed for users to request modifications to ASAP. The users' group will need representatives from the different sites where ASAP is used. Their role will be to help evaluate and prioritize the changes being requested, and to decide which changes to submit to programming staff; and which changes to refer to an expert for elaboration. Some of these users might also be engaged in testing new versions of the software before release to the general user population.

A part time knowledge engineer will be needed to help users to evaluate the changes they are requesting in light of the effect each change has on the overall knowledge base. Since end users are frequently not experts, the knowledge engineer will need to work with identified experts to enhance the knowledge base as needed. The

knowledge engineer will also need to re-test the system to determine if the knowledge is accurately represented.

Repeated evaluation is also critical to maintaining the integrity of the knowledge base. Periodic evaluations should be conducted to insure that changes to the knowledge base have not compromised the ability to achieve the desired results. Evaluation is also a necessary component for introducing ASAP into new environments.

Finally, as its been since the inception of ASAP, project champions, will always be needed. Hopefully, the Steering Committee will be the formal embodiment of the champions. Informal champions would also be helpful. A project as expensive and elaborate as ASAP exists only because there were individuals in positions of authority who had a vision and a willingness to put their authority and their funds behind that vision. That kind of support will continue to be a necessity .

Summary and Lessons Learned

It has been six years since the first expert was interviewed by the ASAP knowledge engineer. During that time, the following lessons were learned. Hopefully these lessons will help guide and encourage the development of other innovative computer systems, particularly in the human services.

- It is possible to capture and encode social work practice expertise and common sense heuristics.

- Access to automated expertise can help non-expert human service practitioners approach expert level performance.

- The use of multiple experts is not only possible but necessary when dealing with complex practice expertise that cuts across disciplines. That use can result in a synergistic effect.

- The use of multiple experts is also plausible in the evaluation of expert systems.

- Integration of multiple technologies may be required for innovations to be useful to end users.

- Early and on-going involvement of end users and service consumers is invaluable to the development process.

- Evaluation-in-practice goes beyond 'prototyping'. It can include quantitative outcome measures, even on a small scale. These can be aggregated through sequential installations.

- Evaluation of these innovations includes the technology (both computer and knowledge domain), the users, the organizational conditions, and the quality of communication. These factors don't just affect successful deployment, they affect the validity of the tool.

- The success of innovative projects requires the support of visionaries and project champions with organizational clout.

Acknowledgements

As indicated, ASAP is the result of the committed efforts of many people. In addition to those noted in the text we wish to acknowledge the following for their contribution: Melissa Mattes, Olivia Flournoy, Tari Nixon, and Dr. Vijay Ganju from TXMHMR, and Albert P. Mendall from Mendall Associates. We also wish to thank the experts, field staff, service consumers and their families who helped develop and test this product. Mary Anne Mendall was the Knowledge Engineer for ASAP. Susan Millea was the external Evaluator.

References

Ayel, M. & Laurent, J. P. eds. 1991. *Validation, Verification and Test of Knowledge-Based Systems.* Chichester, England: John Wiley & Sons.

Bloom, M. & Fischer, J. 1982. *Evaluating Practice: Guidelines for the Accountable Professional.* Englewood Cliffs, NJ: Prentice-Hall.

Buchanan, B.G., & Shortliffe, E.H. eds. 1983. *Rule-Based Expert Systems: The MYCIN Experiments of the Heuristic Programming Project.* Reading, MA: Addison-Wesley.

Eveland, J. D., and Tornatzky, L. G. 1990. The Deployment of Technology. In L. Tornatzky & M. Fleischer (Eds.), *The Processes of Technological Innovation* (pp. 117-147). Lexington, MA: Lexington Books.

Fleischer, M and Roitman, D. 1990. Implementation. In L. Tornatzky & M. Fleischer (Eds.), *The Processes of Technological Innovation* (pp. 197-232). Lexington, MA: Lexington Books.

Gasching, J, Klahr, P., Pople, H., Shortliffe, E., & Terry, A. 1983. Evaluation of Expert Systems: Issues and Case Studies. In Frederick Hayes-Roth, Donald Waterman, and Douglas Lenat, Eds. *Building Expert Systems.* Reading, MA: Addison-Wesley Publishing.

Geiss, R.G., & Viswanathan, N. 1986. Information Technology and Helping People. In R.G. Geiss and N. Viswanathan (Eds.), *The Human Edge: Information Technology and Helping People.* New York: Haworth Press.

Gibson, D., Smilor, R. 1991. Key Variables in Technology Transfer: A Field-Study Based Empirical Analysis. *Journal of Engineering and Technology Management* (8): 287-312.

Hasenfeld, Y. 1983. Human Service Organizations. Englewood Cliffs, NJ: Prentice-Hall.

Hasenfeld, Y. 1992. The nature of human service organizations. In Yeheskel Hasenfeld, Ed., *Human Services as Complex Organizations*. Newbury Park, CA: Sage Publications. 3-23.

Henderson, J. C. 1986. Emerging Trends and Issues in Decision Support Systems and Related Technologies: Implications for Organizations. In R.G. Geiss and N. Viswanathan (Eds.), *The Human Edge: Information Technology and Helping People*. New York: Haworth Press.

Hersen, M. & Barlow, D. 1976. Single-Case Experimental Designs: Strategies for Studying Behavior Change. New York: Pergamon Press.

Jayaratne , S. & Levy, R. 1979. Empirical Clinical Practice. New York: Columbia University Press.

Kidd, R. & Carlson, R. 1992. A Truly MAGIC Solution. In Proceedings of the Fourth Annual Innovative Applications in Artificial Intelligence Conference, 237-247. Menlo Park, CA: AAAI Press.

Kling, R. 1993. Organizational Analysis in Computer Science. *The Information Society*. (9): 71-87.

McGraw, K. & Harbison-Briggs, K. 1989. *Knowledge acquisition principles and guidelines*. Englewood Cliffs, NJ: Prentice Hall.

Millea, S. 1994. Automating Human Service Expertise: The Validation of Knowledge-Based Expert Systems in Social Work Practice. Ph.D. diss., School of Social Work, University of Texas at Austin.

Mullen, E. & Schuerman, J. 1988. Expert Systems and the Development of Knowledge in Social Welfare. Unpublished paper prepared for the Conference on Empiricism in Clinical Practice, August, 1988, Great Barrington, MA.

Nurius, P. 1990. A Review of Automated Assessment. In R Reinoehl & T Hanna (Eds.) *Computer Literacy in Human Services*. (Special Edition) *Computers in Human Services* (6): 265-282.

O'Keefe, R. & O'Leary, D. 1989. The Verification and Validation of Expert Systems. Paper presented at the Eleventh International Joint Conference in Artificial Intelligence. San Jose, CA

O'Leary, D. E. 1991. Design, Development and Validation of Expert Systems: A Survey of Developers. In Marc Ayel and Jean-Pierre Laurent, Eds. *Validation, Verification and Test of Knowledge-based Systems*. Chichester: John Wiley and Sons. 3-20.

Pruger, R. 1986. Information Technology in Support of Service-Delivery Decisions. In Geiss, R.G., & Viswanathan (Eds.), The Human Edge: Information Technology and Helping People. New York: Haworth Press. 212-234.

Rothman, J. 1980. Social R and D: Research and Development in the Human Services. Englewood Cliffs, NJ: Prentice Hall.

Simmons, M. 1989. Automated Welfare Eligibility. *Government Technology*. August.

Thomas, E. 1985. The Validity of Design and Development and Related Concepts in Developmental Research. *Social Work Research and Abstracts* (20): 50-55.

Tornatzky, L. , Eveland, J. D., & Fleischer, M. 1990. Technological Innovation: Definitions and Perspectives. In L. Tornatzky & M. Fleischer (Eds.), *The Processes of Technological Innovation* Lexington, MA: Lexington Books. 9-25.