C.1 - Introduction

Governments increasingly need to collaborate with each other in order to solve global problems [1]. Information gathering, sharing, processing and retrieval are inherent to such collaborative efforts. Information systems that support international collaborations among governments face challenges that require research on how to

- Protect, control access, filter, summarize, correlate and share information across agencies and organizations without compromising the security, laws, autonomy and culture of the participating countries;
- Interoperate transparently across countries whose heterogeneous information networks differ in design, reliability, performance and technological age; and
- Gather, represent, share, translate and retrieve multilingual information.

Problems that require global solutions include global warming, illicit drugs, biodiversity, pollution control, disaster preparation and mitigation, epidemic diseases, trade disputes and the international digital divide. A common first step towards the solution of any of these problems is the voluntary participation of countries in activities that characterize and mitigate the problem. The extent and success of these activities are monitored by international organizations through information collection and summarization in terms of mutually agreed-upon indicators.

The results of the proposed work will have general applicability to transnational processes of collaborative information collection and monitoring. Currently, significant expense and extensive overhead are incurred by agencies of different countries in the collection of data (about agreed-upon standard indicators) that vary in accuracy, completeness, promptness and compliance with international agreements. Inconsistent or unreliable indicators can invalidate information about regional issues, thus becoming a problem of concern to all involved countries. Information technology can play a major role in solving this problem by facilitating the infrastructure, tools, protocols and communications needed to seamlessly integrate transnational activities into national government procedures. The long-term vision is one where governments and public-service agencies can access and use each other's information infrastructure as if it were a part of a single information grid where information and services are securely deployed, shared in a controlled manner, and available with the necessary quality of service (QoS). The Organization of American States (OAS) refers to this futuristic hemispheric infrastructure as AMIGA, the Americas Multilingual Information Grid Alliance. The proposed project, while working towards this goal, has the following more specific focus:

- **Problem domain:** We propose research in the context of a process of transnational cooperation among all Western Hemisphere governments in dealing with the negative impacts on society of illicit drug production, traffic and consumption. The process is coordinated by CICAD-- the Inter-American Drug Abuse Control Commission, which is a technical body of the member nations of the OAS. The process, called the Multilateral Evaluation Mechanism (MEM), is defined as a "singular and objective process of multilateral governmental evaluation, in dealing with the diverse manifestations of the drug problem." It requires that countries collect, share and analyze extensive amounts of information in accordance with agreed-upon standard indicators presented in the form of a questionnaire. A complete response to this questionnaire requires data obtainable and exchangeable in transnational survey instruments to facilitate the conduct of national epidemiological surveys. These instruments assist countries in determining drug prevalence, the social and economic costs of drugs to the country's economy, and informing policy for drug prevention and treatment. One of the objectives of this project is to provide innovative information technology approaches to the deployment and use of these and other instruments for data collection and analysis throughout the participating countries.

- **Transnational partnership between academia and government agencies:** The proposed work will be conducted by a unique team of researchers from five universities (Carnegie Mellon U., U. of Belize, U. of Colorado, U. of Florida, U. of Massachusetts and Ponticia Universidad Católica Madre y Maestra) and experts from agencies in three different countries (US, Belize and the Dominican Republic). Under the umbrella of the OAS, several ministries and agencies in the three countries will be involved. These include two OAS departments in Washington, D.C. (the Information Systems division and CICAD's Inter-American Observatory on Drugs); the National Drug Abuse Control Council of Belize's Ministry of Health; and the National Drug

Council of the Dominican Republic. The university researchers include experts on speech-based interfaces, machine translation, databases, information retrieval, Internet computing and networking.

- **Research foci:** Information systems that support international collaborations among governments face many problems whose solutions require fundamental research. We propose research on technology and tools for the collection, processing, exchange and integration of information needed by transnational government. Research challenges fall in the following areas (detailed in Section C.5):

    - Spoken dialogue systems for data collection, training and learning;
    - Active data management and security techniques for rule-based secure data sharing and filtering;
    - Information retrieval and machine translation technology for sharing documents and searching information across different languages and countries;
    - Middleware for transnational (heterogeneous) information grids that enable private, secure and dependable automation of collaboration processes and policies, and the delivery of computing services through Internet portals; and
    - Network behavior modeling and optimization for delivery of acceptable quality of service.

This project presents a unique and timely opportunity to address these challenges. The project builds on several planning workshops on international collaboration and digital government sponsored by the NSF and OAS. Activities at these workshops led to the formation of an international team of computer scientists, engineers and personnel in government and non-government agencies who bring complementary expertise and passionate commitment to the project. Perhaps most important, the project benefits from a focused multinational digital government effort, the Multilateral Evaluation Mechanism (MEM). This task, which is managed by OAS CICAD and includes participation by all Western Hemisphere nations, has clearly defined objectives, and provides a remarkable test bed for both tractable and extensible research in multinational digital government processes. The proposed work addresses research challenges in the context of this Multilateral Evaluation Mechanism.

C.2 - The MEM: an example of a transnational digital government process

The MEM involves the collection and analysis of data by, and from, several government agencies and non-government organizations within each country. The data in question corresponds to, or allows the generation of, a standard set of indicators negotiated among OAS member states to stimulate measurement of national anti-drug efforts and inter-state cooperation for their improved execution. For the year-2000 version of the MEM, the indicators measure progress towards sixty-one objectives grouped under five different goals. These goals correspond to five broad categories of data related to sixty-one indicators of how the objectives are being met. The goals are:

1. Prevention of drug use and the treatment of drug abusers;
2. Reduction of drug production;
3. Improvement of drug control and related measures;
4. Optimization of national strategy; and
5. Estimation of the cost of the drug problem for the affected countries.

Within a country, the MEM process requires the participation of several government agencies. Agencies are expected to gather, process and report data relevant to their missions. They may also need to collaborate and share data with each other in order to carry out their mission or to provide indicators required by the MEM process. For example, towards drug use prevention, health agencies may be in charge of data on drug treatment and rehabilitation programs, education agencies may deal with data relative to school programs in preventing drug usage and justice agencies will typically deal with data relative to drug law violations and arrests. To assess the impact of prevention programs on youth in need of drug treatment and rehabilitation, and on drug-related crime, these agencies must share data. Similarly, to optimize drug policy and estimate the drug problem cost, organizations that need to conduct research and evaluations of drug abuse prevention programs will need access to data from several agencies. The MEM recognizes the importance of these collaborative efforts by explicitly specifying two indicators of the existence of mechanisms to effect collaborations. Concretely, Indicator 39 refers to the “existence of a mechanism for effecting coordination, cooperation and timely exchanges of information among national authorities” and Indicator 31 refers to “existence of a mechanism for effecting operational information exchange and collaboration among national authorities with responsibilities for pharmaceuticals and controlled chemical substances.”

Reflecting the transnational nature of the drug problem, the MEM process also recognizes the importance of national agencies being able to access and provide information to agencies from other countries. Concretely, Indicator 32 refers to "existence of a mechanism for effecting operational information exchange and collaboration between countries, in relation to pharmaceuticals and controlled chemical substances" and Indicator 40 refers to the "Existence of a centralized agency for effecting coordination, cooperation and timely exchanges of information between countries in accordance with international agreements." The ability to share information across borders is key to the understanding of drug traffic patterns, the impact of drugs on regional society, security and economy, and the implementation of regional actions to prevent drug use.

The evaluation provided by each country on its programs and efforts in dealing with the drug problem and its manifestations is reviewed by an Inter-Governmental Expert Group and developed into a report. The originating country is then consulted on the contents of this report. The CICAD Executive Secretariat coordinates the process and CICAD itself approves the reports on each government generated by the expert group. The MEM process is based on cooperation and does not impose sanctions. It recognizes that all countries share responsibility for the drug problem and that, by sharing information and strategies, they can develop a clearer picture of needs and weaknesses and identify areas that may warrant closer cooperation, better legislation, more research or additional resources.

To date there has been one MEM evaluation and a second one is under way. Existing problems range from inefficient data collection to inconsistent questionnaires, non-electronic responses, incompatible formats, and manual retrieval of information for analysis purposes. While individuals in most of the government agencies and non-government organizations share ideas and information (e.g., about resources and programs) with colleagues in other countries, this information is shared mainly through interpersonal interaction - conversations at meetings, phone calls, or email. There are few mechanisms in place to enable effective access and sharing of information.

C.3 - Towards automation of the MEM

CICAD sees Internet-based information technology as a primary tool for creating integrated national systems for the compilation and maintenance of statistics and documentation in conformity with CICAD's Multilateral Evaluation Mechanism. The national systems will link into a regional information network grid under development by the Inter-American Observatory on Drugs. This grid will connect governmental and non-governmental institutions that directly or indirectly collect data and documentation important for understanding the current situation in each nation, identifying trends and measuring the impact of specific policies and activities. The MEM is a first step in the implementation of a long-term vision of digital government in Latin America—the so-called America's Multilingual Information Grid Alliance (AMIGA)-- which would consist of the integration of the information grids of OAS member countries for all aspects of scientific, societal and government collaboration.

Currently, the implementation of the MEM process faces a variety of problems that have a similar nature, in spite of differences among infrastructures of distinct countries. Within individual countries, the MEM process requires the creation of dependable frameworks for gathering, organizing, sharing and processing information. The OAS/CICAD must contend with the existence of disjoint information systems of member nations and serve as a collector and conveyor of MEM data from and to those countries. Being unrealistic to force member nations to implement identical digital government infrastructures and procedures, CICAD and other agencies must be able to interface and interoperate with heterogeneous systems. An expected contribution of the proposed work will be a framework for the automation of the MEM process, and solutions of research problems inhibiting the successful deployment of the framework across the participating countries. The identified solutions are expected to scale and generalize so to be applicable to the countries in the Caribbean and the Western Hemisphere.

C.4 - Preliminary work (and Results from Prior NSF Support)

The participants of the proposed project have already done some work towards the identification of research problems in, and establishment of a test bed for, effectively automating the MEM process. The work was funded by an NSF grant for exploratory research. Brief descriptions of other NSF-funded projects on topics relevant to the proposed work are also provided in this section.

NSF Award: EIA-0107686; **Period:** 3/15/01–2/14/02; **Amount:** $72,017; **PI –** J. Fortes**; Title:** Exploratory research on transnational digital government.

The work supported by this award included several interactions, discussions and studies of the nature of, and issues faced by, transnational digital government in the context of the MEM process. It also included the

development of a partnership between agencies of three countries and researchers from several universities. These activities provided the basis and rationale for the work proposed in this proposal. A workshop on transnational digital government was organized from May 9th to May 11th in Belize City, Belize. The participants included representatives of over 20 government departments and non-government organizations across Belize, researchers at the University of Belize, state and government representatives from Mexico and the Dominican Republican, representatives from the OAS, and the principal investigators of this proposal.

The participants, including the PIs of this proposal, discussed ways in which information technology can be used to help government agencies within Belize share information with each other, and in some cases with other countries. The first day introduced participants to CICAD's transnational process instruments for studying, reporting on and exchanging data on illicit drug abuse and control efforts, and presented technology options for collecting and sharing this data at both the national and Hemispheric levels. The instruments, respectively known as the Inter-American Uniform Drug Use Data System and the Uniform Statistical System on Control of the Supply Area, are transnational instruments suitable for exchanging and comparing information among countries.

The second day of the workshop outlined initial content and electronic architecture (hardware, software and networking components) for a national information grid alliance to launch and support transnational digital government in Belize. Following presentations by the individual agencies of the templates highlighting their information assets and needs, participants formed four working groups to discuss public sector operations where collection, flow and automation of information are crucial to accurate national and transnational illicit-drug-related policy and program decisions. The working group topics related to: "Overcoming Language and Cultural Barriers," "Transnational Processes," "Integration and Retrieval of Information," and "Integration of Software and Hardware."

The third day of the workshop opened with the presentation of a successful digital government initiative by a representative of the Mexican State of Colima, and continued with a second round of working groups which collapsed those of the previous day and dealt with Belize's "National Information Grid Alliance" and "Transnational Digital Government." The Transnational Digital Government Working Group suggested immediate action to constitute a Web portal with the tools, formats, methodology useful in grid construction and an index of national agency data. The Web portal would serve as a central point of information and contact during the construction of the grid. The constitution of an interest group was also suggested. The Web portal site and its metadata would be managed by the CICAD Executive Secretariat with assistance from all participants.

NSF Award: EIA - 9975275; **Period:** 9/1/1999 – 8/31/2003; **Amount:** $1,200,000; **PI-**J. Fortes**; Co-PIs:** R. Eigenmann, M. Livny, V. Taylor; **Title:** Network-computer for Computer Architecture Research and Education (NETCARE); **Publications**: [2, 3, 4]

We have conceived, developed and deployed the first distributed computing infrastructure that allows researchers, educators and students throughout the world to access and run tools for design, simulation and programming of computer architectures. NETCARE allows such tools to be shared, tested, evaluated and used via conventional web browsers, in addition to providing tool information, codes and examples.

NSF Award: EIA-0075284; **Period**: 9/1/2000-8/31/2001; **Amount**: $100,000; **PI**- S. Y. W. Su, **Co-PIs**: H. Lam and S. Bai; **Title**: Research on Advanced Technologies to Support Internet-based Scalable E-business Enterprises (ISEE); **Publications**: [5, 6].

We are developing an information infrastructure for supporting Internet-based scalable e-business enterprises. It consists of a network of ISEE hubs, each of which has a number of replicable e-business servers providing various e-services to individuals and businesses. These servers manage business events, rules and processes, and perform constraint satisfaction processing, cost-benefit analysis and selection, brokering, supplier selection, and other functions needed in a collaborative e-business environment.

NSF Award: NCR-9508274; **Period:** 9/1/1995 – 8/31/1999; **Amount:** $460,000; **PI:** D. Towsley; **Co-PI:** J. Kurose; **Title:** Multimedia communications in wide-area networks; **Publications**: [7, 8, 9]

This project focused on multicast communications, quality-of-service, video streaming, and performance modeling. Noteworthy is the work coupling forward error correction with reliable multicast [9], which won the 1998 IEEE William Bennett Award and the work modeling TCP behavior [8], which was runner up for the 2000 IEEE William Bennett Award.

NSF Award: EIA-9996246; **Period**: 09/15/97 - 08/31/00; **Amount**: $1,800,000; **PI**: R. Cole**; Co-PIs:** D. Massaro, M. Macon; **Title**: Challenges in CISE: Creating Conversational Agents for Language Training: Technologies for the Next Generation of Interactive Systems; **Publications**: [10, 11].

We have developed interactive learning tools now being used in classrooms in several schools to help children with disabilities learn classroom subjects and speech and language skills. Research led to improved speech

recognition, text-to-speech synthesis and facial animation technologies that were integrated into an animated agent that was incorporated into a wide variety of learning applications. These applications were shown to improve vocabulary acquisition and use, and speech recognition and production skills of profoundly deaf children. This work was recently featured on ABC's Prime Time Thursday and the NSF home page.

C.5 - Proposed work

All MEM-related processes require that information be collected, processed and distributed. They can be classified in three different categories with different characteristics, as summarized in Table 1. Figure 1 outlines a generic process of information collection and use that broadly applies all of these processes. The transitions between the process stages shown in Figure 1 could potentially involve transnational activities. For example, it may be necessary to filter and reformat data collected in one country before agencies from other countries can use that data to generate statistics.

Table 1 - High-level categorization of types of MEM-related processes for information collection, processing and usage.

	Survey-based	Record-based	Report-based
Examples	School and household surveys of drug use	Statistics related to hospital admissions, police arrests and conviction records	OAS reviews of confidential self-evaluations done by each country on the basis of MEM indicators
Data/information collection	Individual anonymous surveys	Public and confidential records generated by government employees	Reports and data generated by national agencies and organizations
Information processing	Statistical tools used to generate shareable summaries and consolidated information	Software developed by OAS for online and offline consolidation of information	Information consolidation done manually or with customized software. Self-evaluations reviewed and rated by OAS.
Periodicity	Every 2 years	Continuous	Continuously done by OAS, every other year for individual countries
Type of results	Statistics	Statistics	Reports and documents
Variations	Identical surveys for each person	Predefined forms used	Country-specific; evolving processes

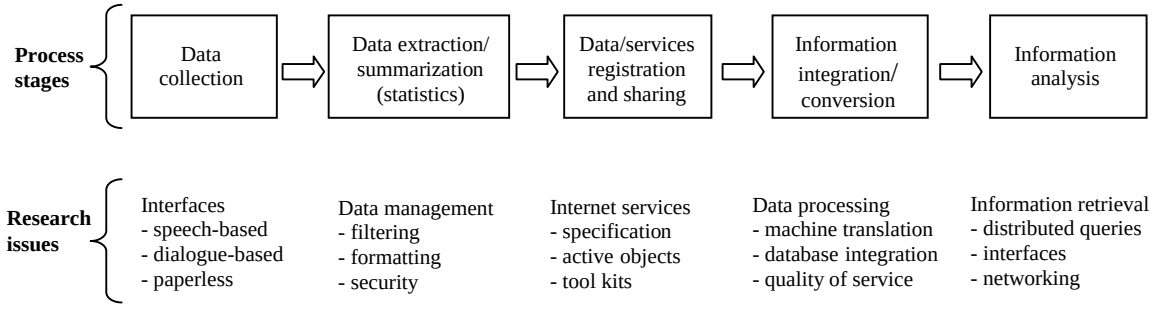

Figure 1 - Opportunities for advanced information technology contributions to the effective implementation of the different stages of the processes needed to generate and use MEM indicators

Proposed Research on Interfaces: Spoken Dialogue Systems for Data Collection, Training and Learning

Data collection from diverse populations such as students, farmers, police and hospital workers, and households poses challenges with respect to rates of response, consistency, anonymity, accuracy and cost-effectiveness. Cultural and educational factors limit the extent to which providers of data can interpret and respond to conventional paper-based surveys, which could contain more than fifty questions. These issues can be effectively addressed through speech-based dialogue-oriented surveys that can leverage a range of information and communication technologies. Such technologies could be as conventional as ubiquitous phone services, and as sophisticated as the increasingly inexpensive palmtop electronic assistants. We propose to address the research issues that arise in the process of deploying several proven dialogue-oriented interfaces in countries like Belize and the Dominican Republic.

During the planning workshop in Belize, we showed a video of profoundly deaf children interacting with an animated conversational agent to learn language skills. The video (produced by ABC TV and shown on Prime Time Thursday) shows Baldi, an animated agent that interacts with children through speech to teach them to identify and name objects in pictures. Baldi produces excellent visible and auditory speech, and provides both spoken and visual feedback to the students based on their responses to Baldi's questions. After the video presentation, we demonstrated (in real time) how authoring tools in the CSLU Toolkit could be used to develop new multimedia applications incorporating Baldi. The workshop participants were unanimous in their opinion that similar applications, customized for their needs, would provide invaluable tools both for collecting essential MEM data and educating their constituents.

The goal of our work is to develop English and Spanish dialogue systems for collecting MEM data, and to use these systems as test beds for collaborative research. In the first year of the project, we will work with colleagues in Belize and the Dominican Republic to design and test a set of English and Spanish spoken dialogue systems for data collection. This process should proceed smoothly because core technologies, including English and Spanish speech recognition, speech synthesis and animation systems are already in place; because authoring tools for data collection (and learning) applications have been developed and tested, and we have substantial experience (and success) developing similar data collection tools with colleagues in Latin American countries (Mexico, Brazil, Chile).

The proposed work will be conducted using the CSLU Toolkit [12,13,14], a comprehensive set of tools and technologies enabling research and development of spoken dialogue systems. We will work with interested personnel in government and non-government agencies to design spoken dialogue systems using the CSLU Toolkit's authoring tools both for collecting MEM data and for designing applications to educate individuals about illicit drugs. Data collection systems will be designed to provide an engaging experience with animated agents that ask questions in English or Spanish.

The resulting systems will record and save all spoken responses and other behaviors (e.g., mouse clicks on objects in pictures). Eventually, spoken responses will be recognized automatically using English and Spanish recognition systems. However, experience shows that to achieve acceptable recognition performance in a new task domain, the system's recognition responses must be verified manually. The corrected data will then be used to conduct research to improve recognition technologies. Thus, spoken dialogue data collection serves dual roles: the systems serve as instruments for collecting MEM data automatically, and as test beds for research on English and Spanish recognition technologies. The systems will serve as test beds for research on acoustic modeling (collecting and modeling acoustic phonetic data from native speakers), pronunciation modeling (modeling dialects and accents as phonological processes within words), and language modeling (using statistics on co-occurrence of words to improve recognition.

Professor Cole and his colleagues have substantial experience developing and deploying conversational interfaces for data collection, information access and learning. In 1995, Cole and his team at the OGI developed a spoken dialogue system for taking the U.S. Census over the telephone. The system was successfully tested in the 1995 Census trial [15]. The CSLU Toolkit, developed by Cole and his colleagues at OGI, has been used to collect speech data in 22 languages, resulting in over 20 different speech corpora that are freely distributed worldwide and used widely for research. Under an NSF Challenge grant, new authoring tools were developed incorporating animated characters, transforming the toolkit into a more powerful tool for data collection and learning. Baldi has been used to collect speech data from over 2000 U.S. and Mexican children. In the past five years, the CSLU

Toolkit has been installed in over 6000 sites worldwide, and over 200 teachers, parents and students have learned to develop applications with the toolkit's authoring tools. Through grants from NSF/CONACyT and NSF/CNPq programs, Spanish and Portuguese recognition and synthesis technologies were developed, enabling multilingual data collection and learning applications in these languages. Under an NSF ITR grant, new data collection and learning tools are now under development in both the U.S. and Mexico. These data collection and learning tools use English and Spanish animated agents to help children with reading problems learn to read.

To succeed, the proposed work on spoken dialogue systems requires international collaboration. Conversational systems must be designed, deployed and evaluated "in-country." This is because statistical techniques used to model linguistic behaviors, communication channels and acoustic environments require speech data from native speakers over local networks. Moreover, development of spoken dialogue systems requires insights about both linguistic behaviors and cultural attitudes that could impact system use. These linguistic and cultural issues are often subtle, but can have disastrous consequences if they are ignored; e.g., resulting in rude or incomprehensible interfaces. Systems must be designed to suit the needs of the agencies, their personnel and their clients. Developing good systems requires the active participation and insights of these users at all stages of the design process.

Proposed Research on Data Management: Event-Trigger-Rule (ETR) Management and Security Technology

Two major goals of agencies in charge of data collection within and across countries are the protection of privacy and national autonomy, and the optimization of resources by using single survey instruments and data collection efforts for multiple purposes. These seemingly conflicting goals can be reconciled by using information technology to automatically filter data and route/share/update information as needed and allowed by rules of the involved agencies. In the envisioned environment of networked agencies, many events that happen at one agency can be of interest and importance to other agencies/users and their application systems. Examples of events include the publication or modification of a document, the posting of a Web page, the actions taken by users through a browser, signals from external devices, before (or after) the invocation of an existing application, before (or after) the enactment of a work process, exceptional conditions, etc. Users and/or application systems can register for some events, specify some data conditions to form event filters, and receive event notifications (by email, application program activation, paging, FAX, or other means) of when these events occur and the specified data conditions are satisfied. Events and event notifications are powerful means to allow distributed objects, users and organizations to interact/inter-operate in a loosely coupled computing environment like the one envisioned in this project.

Interactions/interoperations among agencies are often subject to security, privacy and data integrity constraints, and government policies, strategies, and regulations. These constraints can best be specified as rules using a high-level rule specification language. They can be enforced by a rule processing system instead of being hard-coded in application programs. This is because the former approach is much easier for people to understand, use and modify. There are two general types of rules: constraints and action-oriented rules. The former type is used to specify data constraints in forms of attribute constraints (e.g., information about the nationality of a hospitalized drug abuser can only be national, international and undocumented) and inter-attributed constraints (e.g., if the patient is international, then an agency in his/her country of origin should be contacted without releasing the name of the patient). The latter type of rules consists of a condition specification and an action specification. If the condition part is evaluated to True, then the operations specified in the action part are performed. If we tie events with rules, then the occurrence or posting of an event can automatically trigger the processing of rules.

Event-Trigger-Rule (ETR) management: Event-Condition-Action (ECA) rule systems [16] have been used in active database management systems [16, 17, 18, 19], commercial systems for business applications, business rule repositories and rule mining [20, 21, 22, 23, 24]. We propose to generalize the ECA paradigm into an Event-Trigger-Rule (ETR) paradigm [25, 26, 27], which supports flexible interoperation and collaboration among loosely coupled agencies of different countries over the Internet. Unlike the ECA paradigm, we separate events and rule specifications so that they can be defined by different organizations. We then use trigger specifications to tie events with structured rules. A rule that represents a small granule of logic and control can participate in a number of rule structures to form larger granules of logic and control; thus it becomes reusable. We also distinguish "triggering events" from events that participate in a composite event expression (or event history) in our trigger specification. An example of a trigger specification is "When E1 or E2 occurs, verify if E3 and E4 have also occurred within a specified time window. If so, activate a structure of rules." Events allow loosely coupled systems from different organizations to inter-operate and collaborate with minimal dependency (via event publish-subscribe actions).

Agencies can control 1) who can access their data using role-based access control specified by rules, 2) when to release and deliver their information to users/organizations by controlling local events and rules triggered by a timer, and 3) when to allow external information and events to affect local data and/or to activate local applications by controlling local triggers. The above listed capabilities will allow 1) the enforcement of government and inter-government constraints, policies and regulations, 2) the automatic dissemination of data to relevant users and organizations in a timely manner, 3) the triggering of rules to validate the received data, translate them into the language known to the receiving site and convert the data into useful format, and 4) the implementation of dynamic interagency workflows by incorporating events, triggers and rules in workflow models in contrast with the static or conditioned process sequencing allowed by existing workflow management systems [28]. We propose to investigate the above four uses of events, triggers and rules. The proposed work will made use of a prototype ETR server reported in [27, 29, 30] and extend it to meet our needs in this research effort.

Security Technology: One of the challenges in Internet-based agency collaboration is security control for accessing distributed resources. In a single organization, it is easy to control the access to resources by using identity-based authentication and role-based access control. In the transnational network environment which involves a number of collaborative agencies, trust agreement needs to be established among these agencies to give them the right to grant the right of accessing each other's resources to their own personnel. Also, data transmitted through the Internet can be intercepted and/or modified unless encryption and decryption techniques are applied. In our investigation, we shall focus on the integration of a number of known security control technologies such as PKI [31, 32, 33], role-based access control [34, 35, 36] and distributed trust management [37, 38] to develop a new integrated security model and implement a security server so that participating organizations will have absolute control over what, when and how their resources will be accessed by others. The rule specification language of the ETR server and its corresponding GUIs will be used to define security policies and access control rules.

Proposed Research on Internet Services: Portals, Virtual Private Grids, Distributed Active Objects, E-services

Transnational efforts to characterize and monitor how each country addresses all manifestations of the illicit drug problem must rely on consistent use by all nations of like instruments to collect, assess, present and analyze data. This is a challenge, given the large number of OAS countries and the different capabilities of their information systems, government organizations, workforce and basic infrastructures. In some countries, the agencies and organizations do neither have the hardware/software resources nor the expertise to use proper data collection and consolidation procedures and tools for statistical analysis. For example, Table 1 shows that processing of MEM-related information may require the use of commercial statistical packages (e.g. SPSS), OAS-developed legacy codes or customized software developed elsewhere. Local realities also affect the ability of individual countries to apply standard approaches - workable solutions are often learned through trial and error instead of reuse of working solutions. For example, the OAS has generated survey instruments (questionnaires) that countries can use to collect statistical data on illicit drug usage in schools. However, the challenge remains on how to have all countries administer these surveys so that they have statistical significance in a transnational sense (e.g. with respect to size, age and make-up of the surveyed populations). Currently, significant effort and expense is required from the OAS and its member nations in education workshops, communications, frequent travel to attend meetings, duplicated efforts in deploying infrastructure, and manual integration of information from different agencies.

In the long term, one of the implications (and benefits) of the implementation of MEM-like processes will be the creation of the necessary infrastructures and information technology professionals. However, this will happen on different clocks for different countries. An alternative solution that can have immediate effects is to use the Internet to allow countries to access, learn, obtain and use the necessary tools, resources and knowledge in a single web site. This approach would have an immediate enabling impact on the implementation of surveys, statistical studies and other information processing tasks. The site would act as a portal to tools, instruments, documents, reports, discussion groups, educational materials, and many other resources that can be easily located and transferred to national infrastructures under development. In effect, it would provide agencies with a tool kit, associated instructions and examples of how to deploy MEM processes. Consistency, shared experiences, collaboration, resource leveraging and the opportunity to reuse existing software to develop new tools are natural side effects of a portal to MEM resources. Such a portal (which can be used for either centralized or distributed resources and information) is also desirable in the foreseeable future when individual countries own and share resources. It would thus be a first step towards the long-term vision of AMIGA, a distributed infrastructure that appears as a single information grid to all its users.

Our experience in research and development of web-based computing portals [39,40,41] and active distributed object technology can be readily applied to this problem.

Portals and Virtual Private Information Grids: We have deployed several live, production-quality systems that have thousands of users distributed around the world. These systems include the following:

- NETCARE: a portal for computer architecture and parallel programming tools (http://www.ecn.purdue.edu/netcare)
- NANOHUB: a portal for nanotechnology device simulation and design tools (http://www.ecn.purdue.edu/nanohub)
- SIGMICRO Architecture Center: a portal to tools, articles, experts, forums, industry and test machines for the computer architecture research and education community (http://maul.ecn.purdue.edu:8080/)
- Portals for VLSI design tools and computational electronics (http://www.ecn.purdue.edu/punch)

The above-mentioned portals are powered by PUNCH [42], a distributed computing infrastructure that appears to its users as a portal to upwards of fifty computing tools that can transparently run on geographically distributed machines. PUNCH has more than one thousand users, most of which are located in twelve countries in the Americas, Europe and Asia. This year alone, PUNCH has logged more than one million hits through the above-mentioned. These hits take place as users run the tools via conventional browsers and not simply as they access documents. In this sense, PUNCH provides users with an Internet-computing operating system that allows them to use the Internet as their virtual computer anywhere, anytime. We believe that technology and concepts similar to those used in PUNCH can be applied to the construction and efficient operation of a portal for MEM-related activities. The immediate construction of such a portal was one of the strongest recommendations of the Transnational Digital Government workshop that took place in Belize (as discussed in Section C.4). It would also complement the planned development of a portal to illicit drug information by the OAS.

Several research issues arise in the context of an implementation of the proposed portal. Users must be able to control the privacy of the data, tools and information that they wish to generate, share or provide. This implies the need for secure communication (see previous discussion under *Security Technology*) and personal accounts with access control. We plan to expand shadow account mechanisms [43] to transparently multiplex large numbers of users onto a finite number of physical accounts on machines that provide the portal resources. In addition to the need for communication to be secure, it should be transparent to users, i.e. users should be able to use data and tools in the portal as if they were in their own computer. This can be done by using a global virtual file system such as the one proposed for PUNCH [44]. However, issues of performance and interoperability need to be better understood in the context of the proposed work and different approaches may have to be considered. The long-term goal is a system that provides each agency or each user with its virtual private information grid which, in turn, is transparently mapped into the physical information grid where resources of all kinds remain under control of their owners.

***Active distributed object technology*:** The processes of construction and maintenance of a portal that aggregates information and services from different countries are also a challenge. Ideally, these processes should allow for decentralized contributions of all types of objects (i.e. data, tools, documents, messages, information, etc). We need to provide a powerful data model for transnational agencies to specify various types of data (reports, surveys, statistic summaries, etc.) and software systems that they are willing to make accessible to other agencies. These data and systems can be uniformly modeled as distributed objects using the existing distributed object technologies exemplified by CORBA [45], DCOM [46], and Enterprise Java Beans [47]. However, distributed objects are "passive" in the sense that some other data objects have to initiate the access of data objects and system objects can only respond to method calls from other data objects. They cannot "actively" enact processes or perform operations to enforce government and inter-government constraints, policies, and regulations in response to some events. They do not provide a way to add "knowledge" to the existing resources, which are needed for the integration, management and manipulation of distributed resources.

In this project, we shall develop an active distributed object technology, which incorporates events, rules, and processes in the modeling of various types of inter-government resources and develop GUI tools for capturing the active properties of active distributed objects. To give distributed objects their active properties, we shall extend the traditional object model into an active object model. In this model, objects of a class are modeled in terms of attributes and methods just like the traditional object model [48, 49, 50, 51, 52, 53]. Additionally, events, triggers, constraints, and action-oriented rules, which can be used to specify many kinds of knowledge at a high-level of

abstraction (instead of program code), can be optionally included in object class specifications. The extended specifications of object classes will be used to "wrap" existing data and software/application systems as active distributed objects and to generate code. Thus, interoperations among distributed objects at run-time may post events and trigger the processing of rules by using the services of an existing Event-Trigger-Rule server. This object modeling and processing work will make use of the experience we have in building object-oriented knowledge base systems [54, 55, 56].

We shall develop graphic user interface tools based on the active object model to allow collaborating agencies to specify data that they want to make available to other agencies. There are two ways that they can publicize the specifications. One is to register with a broker server, which centrally maintains the meta information about all the sharable data in the network and provides browsing and querying facilities to different agencies/users in order to "discover" which agencies have what data. The other way is to publish their specifications as web pages on their own web sites. A web crawler can be developed to automatically process these web pages and gather the meta-information into the broker server. Both approaches can be used to publicize their sharable data. Our proposed effort is to develop the broker server, the web crawler and graphic user interface tools for registration and data resource discovery purposes. The development of the broker server will be based on the emerging technology: the Universal Description, Discovery and Integration [57] with an extension to allow constraints to be specified on the registered data and information access requests. Thus, the broker server can use these constraints to do constraint satisfaction processing to identify the proper sites that contain the right data to satisfy the information access requests. Thus, the requests can then be directed to these selected sites. The proposed broker is therefore a constraint-based broker, which is different from many existing brokers developed by the agent community.

***E-service Specification, Discovery and Access Technology*:** We need tools and mechanisms for collaborative agencies to publicize not just data but also those software/application systems' functionalities that the agencies are able and agree to make available for use by other agencies. The emerging technology being developed by the information technology community is to treat the sharable systems functions or operations as e-services. A language such as the Web Service Definition Language[58] can be used to define the e-services (automated as well as manual services) and the information needed for software systems to access these e-services. Similar to data specifications, the e-service specifications can be either registered with the broker server or published as web pages to be crawled by the web crawler on behalf of the proposed constraint-based broker. By browsing and querying the broker, users can find the available e-services in the network. Software systems (e.g., agents) that query the broker can also obtain the URLs that identify the proper e-service sites. We shall adopt the Simple Object Access Protocol (SOAP)[59] for activating the operations associated with the e-services at these sites. SOAP messages in the XML format can be transmitted between web sites using HTTP. We shall extend WSDL to include the specification of constraints so that constraints of e-service requests can be matched against the constraints associated with e-service specifications by the constraint-based broker to identify the suitable e-service provider(s). The developed technology based on the emerging industrial standards such as WSDL and SOAP will enable the sharing of software/application systems to process data and also simplify the development of new application systems by using the functions provided by the existing systems.

Proposed Research on Information Retrieval: Distributed Queries and Machine Translation

Distributed query processing technology: The envisioned system should provide a querying facility to allow users to issue queries and receive relevant data without having to know: 1) the locations or sites where the retrieved data come from (location transparency); 2) whether multiple copies of the retrieved data exist (replication transparency); 3) what information/database systems are managing the retrieved data, what operating systems they run on, and what hardware systems are used (system transparency); and 4) what communication network or information infrastructure is used to transmit data, queries, messages, etc. (communication transparency). Such a facility will provide more precision in information access than keyword-based search engines. Querying distributed and heterogeneous databases is a well-established technology in the database management field. When and wherever appropriate, we shall apply the existing techniques for distributed query processing and optimization, query decomposition, and data integration for meeting the transparency requirements. However, the envisioned environment presents several additional challenges to be addressed by the proposed research as discussed in the following three paragraphs.

The traditional way to integrate heterogeneous databases is to use a common data model (e.g., the relational model or an object-oriented model) to model the data resources available at different sites to produce a number of

"local conceptual schemas" and integrate these local schemas to produce the "global conceptual schema". Users' queries are then processed against the global conceptual schema or views defined over it. This scheme assumes that these schemas are rather stable (i.e., not subject to changes very often) and accessing distributed data is the only objective. In the envisioned environment, these assumptions may not be correct for two reasons: 1) participating organizations and their data and application systems resources may be added or removed frequently; 2) the sharing of software/application systems is also involved. We shall develop new distributed query processing and optimization strategies to deal with the dynamic nature of the proposed information grid. Information needed by a user or agency may have to be derived by using the e-services of a number of application systems (e.g., statistic data processing systems) of different countries. For example, by giving some input data D1 to system S1, it can produce data D2, which can be fed to system S2 to produce data D3. This process may continue through several systems until the final data needed by the user/agency can be derived. We propose to study techniques for data derivation that involve the use of e-services of existing systems. We shall use our experiences in query processing and optimization and data mediation and validation [60, 61, 62, 63] to carry out the proposed tasks described above.

The citizens of the countries involved in this project speak two different national languages (English and Spanish) and numerous indigenous languages. These languages will be the natural and most efficient vehicles for data collection and archival in the countries where they are spoken but will lead to repositories of information that are multilingual. User interfaces, forms and interfaces for portals in different languages need to be developed to make it easier for international users. Machine translation will greatly facilitate either the access to such information by monolingual speakers or the conversion of multilingual documents into a given language. Machine translation techniques have already shown the potential to become practical tools for multilingual information retrieval and integration, particularly when the application domain is well defined. We propose to investigate fluent multilingual interaction in the context of the MEM. Communication among governments, citizens, non-government organizations (NGOs) and foreign information providers on the Internet requires bridging the linguistic gap in several ways:

- Machine Translation (MT), readily customizable to the application domain (e.g. MEM indicators, governmental procedures, etc.): we will enhance corpus-based MT methods developed at CMU -- example-based MT and statistical-MT -- to be automatically domain adaptable to provide customizable translation [64];
- Translingual retrieval, beyond translating isolated query words. Out-of-context translations of individual query terms to search engines for retrieving documents in different languages results in a 30-70% degradation of precision and recall. Instead, at CMU we have been investigating methods based on generalized vector spaces, latent semantic indexing, and example-based thesauri to provide contextualized query equivalents with very little cross-language degradation [65,66]; and
- Summarization of document clusters (e.g. retrieval cohorts) in any language in order to provide more efficient search and localization of information in the context of AMIGA.

The above technologies require major scientific and technological extensions beyond the current base systems at CMU. For example, domain-adaptable MT requires either rapid domain knowledge acquisition via coding of semantic relations [67] or automated retraining [68] using standard language models or smoothed exponential models for domain lexicon, phraseology and any syntactic or structural properties to render appropriate translations. Exponential models enable smooth combination of evidence for alternative translations, including the base-level (prior) model interpolated with domain-oriented constructions based on all available parallel training text from the domain. Long-range constructions in MT will be handled by smoothed generalized-example-based MT [64]. When substantial examples of similar text have been previously translated these serve as priors for the statistical translator shifting the probability mass to the known translation sequences and finding the maximal-likelihood transitions and fillers among the translated segments. The key research question is rapid semi- or fully-automated adaptation of the MT system to the necessary domains to produce translations that minimize semantic errors. With respect to summarization, we propose to investigate genre-oriented summarization, where news articles, government reports or medical texts (for instance) are summarized differently according the properties of the texts and the expected use of the summaries. Summarization eases the translation burden when real-time information access is desired -- so that translated summaries may be quickly scanned, drilling down to full translation of documents only if or when pertinent to the user's current task. The summarization work will build upon CMU's high-performance Maximal-Marginal-Relevance technology [69,70]. In essence, MMR seeks maximal relevance and maximal diversity of information included in summaries as means to achieve informativeness while maintaining brevity.

Research on Networking and Quality-of-Service (QoS): Network Behavior and Proxy-based Services

The services provided by the network supporting MEM processes are quite variable - we have observed loss rates between the US and other hemispheric countries that fluctuate between 1-3% and 15-20%. Furthermore, tariff structures vary significantly from country to country and can pose significant constraints on communications. For example, within Belize, Internet services are provided by a monopoly. Thus bandwidth is scarce and expensive. Moreover, connect time is expensive. On the other hand, although bandwidth remains scarce within the Dominican Republic, connect time is cheap. In both countries, however, computing power in the form of PCs remains affordable. This variability in services provided by the network and the availability of networking resources poses significant challenges regarding how to provide quality of service to the end user; this is the focus of the following networking research issues to be pursued in this project (discussed below in turn):

- Characterizing and modeling end-to-end loss and delay behavior, and end-user behavior,
- Developing proxy-based network services to compensate for variable network quality and tariff structures.

***End-to-end loss and delay behavior in wide-area networks*:** Proper development of proxy-based mechanisms to enhance network services requires a solid understanding of end-to-end network behavior. Such an understanding can be used to produce parsimonious loss/delay models that can drive the development of advanced services and the development of efficient performance evaluation tools. We outline research to be done on this topic.

We are currently developing discrete time hidden Markov models to characterize the combined end-to-end loss and delay behavior. This effort builds on our previous work in this area [71,72,73] as well as that of [74]. Our preliminary results are promising and we propose to extend this effort in several directions.

Currently, we model end-to-end delay through a histogram that depends on the state of the hidden Markov model. We propose to consider a more parsimonious representation where we replace the histogram with a class of distributions (e.g., family of gamma distributions) described by one or two parameters. The research task is to identify the appropriate class of distributions and then to develop estimators for the loss rates and parameters of the delay distributions. Second, we will consider a representation of end-to-end delay in terms of jitter coupled with the delay of the first packet. Our preliminary results suggest that this may provide estimators of end-to-end delay with better statistical properties. Third, we will develop *continuous time* hidden Markov models of end-to-end loss and delay behavior. This will be better suited for evaluating the performance of an application running on top of TCP. Last, we will apply hidden Markov models to model end-user behavior.

The models that result from the above efforts will be used to drive network simulations of applications developed within this project and to provide on-line predictions of network and user behavior for use by proxies in tailoring services to the end user as will be described next.

Proxy-based services: There has been considerable work during the preceding ten years on providing QoS at the network level within the Internet. For a variety of reasons, technical, economic, and political, little of this work has been deployed within the Internet. One reason is that large backbone networks provide good service through over-provisioning. However, bottlenecks occur at peering and access points because of economic and technological reasons. Furthermore, as stated earlier, networking resources are scarce for a significant subset of end-users.

We propose research on the use of proxies for enhancing QoS perceived by applications and for reducing the costs to end-users. We envision two types of proxies, those residing on dedicated equipment located near network peering points and access points, and those running as processes on end-user equipment. These proxies will reduce user connect times and latencies by locating data and computing resources close to the users. In order to combat high connect time costs, they will also batch data, such as email, which have relaxed latency requirements. Once a batch has been prepared, the proxy can open a network connection and input it into the network. Last, they can also mask network impairments from the user/application by transparently adapting to network conditions as determined using the inference techniques described earlier

Several fundamental problems arise in the context of these services. First, we need to instrument the proxies to continuously monitor paths between proxies and between proxies and clients. This instrumentation should focus on loss, bandwidth, and delay behavior. The results of this monitoring can be used as inputs to: (i) proxy selection by a client in order to obtain needed services and (ii) QoS adaptation between proxies and clients. The performance results will also be made available to applications running on top of our infrastructure. The proxies also need to be instrumented to monitor user behavior in order to determine appropriate batch sizes and the time between opening a connection in order to transfer a batch.

The second problem that we will address is of how to use these observations to provide the sets of services described earlier. This work will again rely on being able to quickly infer network behavior in an on-line manner

and use the inferences to optimize performance. For example, one approach to latency reduction will be to prefetch the most likely accessed resources based on a probabilistic model of user behavior, [75]. Whether to prefetch or not will be based on inferred network performance as well as the likelihood that the resource will be used.

C.6 - Participants, Project Management and Plan of Work

A unique team of investigators and transnational government experts from three different countries will carry out the proposed work in the context of a partnership involving

- Researchers from universities in the US, Belize and Dominican Republic
- OAS experts located in the US, Belize and Dominican Republic
- Members of the staff of agencies in Belize and the Dominican Republic who work in MEM-related processes

Figure 2 is a diagram of the interactions among the participants. The roles of the project partners are as follows:

- *Research on information technology for transnational digital government*: Researchers from the universities of the three countries will conduct the research described in Section C.5. They will interact closely among themselves and with members of the agencies in each country.
- *Test beds for, and deployment of, research results*: Experts from the OAS and the government agencies of Belize and the Dominican Republic will test, integrate into operational practices, and provide feedback on the outcomes of the research; they will also help assess and characterize the problems faced by MEM processes and the efficacy of proposed solutions; the agencies of Belize and the Dominican Republic will interact through the OAS in the deployment of processes that require transnational activities involving the two countries.
- *Transfer of research results to transnational government processes*: Researchers of the Universities of Belize and the Dominican Republic will assist the participating agencies in developing the information infrastructures and training the staff needed for the testing and deployment of research techniques into government processes.

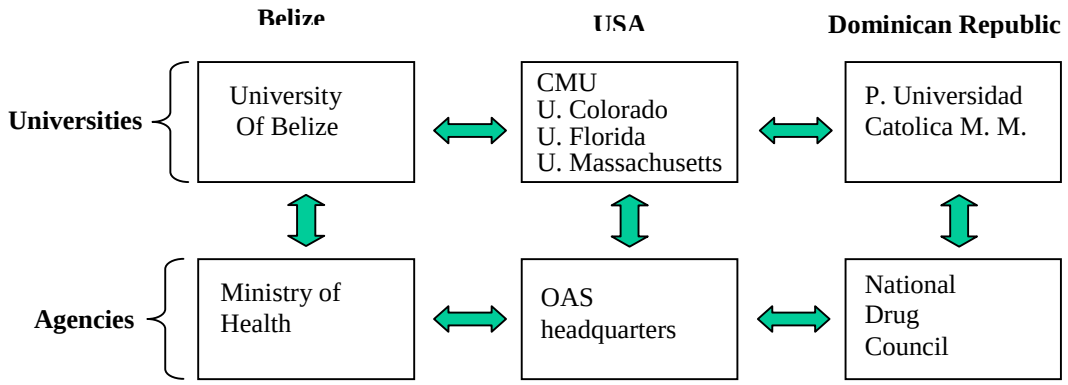

Figure 2 - Project participant interactions: universities interact closely with each other and with agencies in their own countries; local universities and the OAS play a key role in the deployment of research results into transnational processes.

The PI will be responsible for ensuring that the activities across the different participating organizations are coordinated and done in a timely fashion. In their areas of expertise, the PI and Co-PIs will be responsible for coordinating research and development of solutions and working prototypes for problems faced by transnational government. They are also responsible for collaborating with agencies in integrating their systems and technologies in the MEM implementation mechanisms. In Belize, collaborations will be coordinated by Ornel Brooks of the Belize National Drug Abuse Control Council and Charles McSweeney of the University of Belize. In the US, collaborations will be overseen by Ruth Connolly of the OAS and PI José Fortes. In the Dominican Republic, Alfredo Abel, Head of the Information, Education and Communications Center of the National Drug Council and a faculty member of the Ponticia Universidad Católica Madre y Maestra (PUCMM) will coordinate collaborations between their respective organizations. To accomplish the above goals the PI will organize meetings of all the participants (including OAS representatives and international collaborators) as needed (at the least once a year). Conference calls will take place

biweekly, and individual visits of researchers from one institution to collaborating institutions and agencies will take place at least once every semester. As early as possible in the project, a test bed MEM process will be implemented to provide a concrete case study that can serve as basis for the collaboration among all the participants.

The plan of work is as follows:

Year 1: Address research issues that arise in the design and implementation of stages of a prototype implementation of a MEM process like the one shown in Figure 1. The candidate processes are the example processes shown in Table 1. We will build upon digital government software developed with partial support of the OAS for the Mexican State of Colima. Currently, government functions that involve document creation and retrieval by citizens of Colima use this software (e.g. death and birth certificates, driver and property licenses, etc). Past and current development of this software is supervised by Dr. Domingo Zuniga, one of the attendees of the Transnational Digital Government workshop in Belize where he demonstrated the software. He has made the software available to the PIs and countries involved in this project. The prototype stages (to be integrated with the existing software wherever necessary and possible) include the following:

a. A spoken language interface for data collection in Belize and in the Dominican Republic
b. A portal for Internet-based registration and publication of data and services of use to MEM processes; this portal will contain a tool kit for MEM-process implementation
c. Software components and a specification language for implementation of distributed active objects and ETR-based data processing, and e-services
d. Machine translation software
e. Instrumented proxy servers for network performance optimization
f. Initial models of network behavior for online prediction of user and network behavior

Year 2: The stages designed and implemented during Year 1 will be interconnected so to build a geographically distributed implementation of the entire process. The resulting test bed will be used for initial functional evaluation, performance testing and analysis. Improvements to the techniques used in Year 1 will be investigated to address identified problems.

Year 3: The results of Year 2 will be integrated and tested at the agencies responsible for implementing MEM-processes in the participating countries. Identified problems will be addressed.

Evaluation of proposed work: Throughout the project strong emphasis will be placed on continuous evaluation of the suitability of the approaches under investigation and the attained results to the specific needs of MEM processes. The first level of evaluation will occur across the different areas of research by the PI and co-PIs. As needed, appropriate test scenarios will be constructed to detect inconsistencies and mismatches between solutions that can compromise performance or viability of integration across work done at different institutions. All participants will be required to provide evidence of this type of evaluation through joint design reviews, experiments or both. By the end of the second year of the project we will have an initial prototype of a system that supports a MEM-like process across the participating institutions. This system will be used to evaluate key aspects of the research done thus far, in accordance with stated goals and an evaluation strategy decided by the participants. Tests will address functionality, security, performance and correctness of the implementations.

Throughout the project, input and feedback from professionals of the OAS and participating agencies will be solicited on different aspects of the research. After the second year of the project and the successful deployment of the academic test bed field tests will be carried out. All the PIs will be required to work with users (through the OAS and colleagues in each country) to design evaluation scenarios that test the results of the proposed research.

C.7 - Budget explanation

Funding of researchers at the participating universities in the US: funds will be used to cover the participation of faculty, senior research scientists and graduate students; also requested are funds for traveling required to attend annual meetings with all project participants, and to disseminate results at technical conferences; budget also includes funds for communication charges associated with biweekly teleconferencing calls; equipment funds are intended for personal computers and portable devices needed to create a representative test bed for a networked information system for transnational digital government.

Funding of OAS participants: funds will be used to cover personnel time spent in transferring and testing research results into MEM operational settings and a OAS portal, interfacing with officers of ministries and agencies of the participating countries; equipment funds for the OAS component of the research test bed are also requested.

Funding of participants from universities in Belize and Dominican Republic: limited funding (30 k/year) is requested to cover participation or faculty and students in the international component of this project, and to provide minimal equipment resources for deploying components of realistic test beds in these countries.

C.8 – Broader impact

Importance of the proposed work: The potential impact of the proposed work is enormous. The successful deployment of an automatic MEM process that is viable and sustainable across the participating countries will lead to a model that can be used by all countries in the Caribbean region. This same model (or components of it) could be subsequently extended to the entire Western Hemisphere and the AMIGA infrastructure. The OAS has both the expertise on international affairs, and the mandate from its member nations, to facilitate the deployment of the results of these projects into other regions of the Americas.

The manifestations of the drug problem in all aspects of society are a daunting social problem throughout the world. For example, the strong links between drug consumption and the spread of AIDS have tremendous global and regional implications. The proposed work will have a major impact on the ability of nations to use information technology to track manifestations and devise solutions of problems of vital importance to their future.

The multifaceted research required by this project will also have impact beyond the specific transnational processes it targets. For example, speech-based dialogue-oriented interfaces will enlarge the number of citizens who can interact with computer-based systems and access online information. The techniques proposed for providing and accessing services over the Internet will provide citizens with information and capabilities that are independent of geographical boundaries and distances. In the long run these technologies will greatly contribute to the elimination of the so-called digital divide between citizens and nations with different resources. This will be essential to the elimination of the weakest links of transnational information systems used to address global problems.

Educational impact of proposed work: The interdisciplinary nature of the work will provide unique opportunities for students at both the undergraduate and graduate levels to carry out work that cuts across cultures and the research topics of this proposal. The interactions between the US universities and those of Belize and the Dominican Republic will be the first steps towards forms of cooperation These interactions will include short courses, student and faculty visits, student exchanges and other collaborative arrangements. For example, the University of Colorado will give a short course to all interested collaborators on advanced dialogue systems, which will train researchers to develop working systems. The interactions between universities and agencies will provide opportunities for students to understand aspects of government and international organization activities that will both educate and motivate them to pursue careers in digital government.

Another expected consequence of the proposed work is an understanding of transnational aspects of digital government, which have not yet been brought to the attention of the information technology research community. Published results of the proposed work will contribute a body of knowledge that can be used by other researchers and educators in developing technology and trained professionals for transnational digital government.

C.9 – Conclusions

We propose pioneering information technology research that addresses problems faced in the implementation of transnational digital government processes. We focus on a specific class of process – the Multilateral Evaluation Mechanism, a tool to measure the anti-drug efforts of thirty-four countries– that plays a key role in dealing with manifestations of the illicit drug problem in the Western Hemisphere. We formed a partnership that includes agencies in three different countries and researchers from six universities. This partnership is the basis for a close interaction between research activities and the problems faced in the MEM implementation. The US researchers will focus mostly on the research issues described in this proposal, the OAS and the faculty from the Universities of Belize and the Dominican Republic will play a key role in transferring the research results to the operations of agencies of each country, and the participating agencies will provide insights into the important problems to be addressed, and validate and evaluate the proposed solutions through deployment of MEM processes. The project has the enthusiastic commitment of the leadership of the OAS and government ministries (supporting letters have been provided to the Dr. Larry Brandt, Program Manager of the NSF Digital Government program). The relatively small size of the participating countries provides a manageable digital government context where the research team can effectively interact with all branches of government as necessary. We believe that we have a unique setup where our work can focus on fundamental problems of transnational digital government and our solutions can be effectively tested. It is exciting to know that our success in the proposed work can lead to the

deployment of transnational digital government systems across the many countries of the Caribbean region that are similar to Belize and the Dominican Republic. Later, it could possibly extend to other countries of the Americas.

D. References

[1] Cole, R.A., Fortes, J. A. B. and Klinger, A., "Workshop on International Collaboration in Computer Science and Engineering," Stevenson, Washington, October 9-11, 1997, http://cslu.cse.ogi.edu/nsf/wiccs97/report.html

[2] N. H. Kapadia and J. A. B. Fortes, "PUNCH: An Architecture for Web-Enabled Wide-Area Network-Computing," special issue on High Performance Distributed Computing of *Cluster Computing*: *The Journal of Networks, Software Tools and Applications*, Vol. 2, No. 2, September 1999, pp. 153-164.

[3] N. H. Kapadia, J. A. B. Fortes, and M. S. Lundstrom, "The Purdue University Network-Computing Hubs: Running Unmodified Simulation Tools via the WWW," *ACM Transactions on Modeling and Computer Simulation* special issue on Web-based Modeling and Simulation, 2000.

[4] N. H. Kapadia, R. J. O. Figueiredo, and J. A. B. Fortes, "The PUNCH WWW Portal to Tools for Computer Architecture Education," *IEEE MICRO* special issue on Computer Architecture Education, June 2000.

[5] Su, S. Y. W. and Lam, H., "Iknet: Scalable Infrastructure for Achieving Internet-based Knowledge Network," Proceedings of *International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet*, l'Aquila, Rome, Italy, July 31-Aug. 6, 2000. 80 pp. (full length paper in CD).

[6] Su, S. Y. W., Lam, H., Lee, M. S., Bai, S, and Shen, M., "An Information Infrastructure and E-services for Supporting Internet-based Scalable E-business Enterprises", paper submitted to the *Enterprise Distributed Object Computing Conference (EDOC 2001)*, Seattle, Washington, Sept. 4-7.

[7] M. Yajnik, S.B. Moon, J. Kurose, D. Towsley. "Measurement and modeling of the temporal dependence in packet loss," *Proc. INFOCOM'99*, March 1999.

[8] J. Padhye, V. Firoiu, D. Towsley, J. Kurose. ``Modeling TCP throughput: a simple model and its empirical validation." *IEEE/ACM Trans. on Networking*, **8**(2), 133--145, April 2000.

[9] J. Nonnenmacher, E.W. Biersack, D. Towsley. ``Parity-based loss recovery for reliable multicast transmission," *IEEE/ACM Trans. on Networking*, **6**(4), 349--361, August 1998.

[10] Cole, R., D. W. Massaro, J. de Villiers, B. Rundle, K. Shobaki, J. Wouters, M. Cohen, J. Beskow, P. Stone, P. Connors, A. Tarachow, and D. Solcher, " New tools for interactive speech and language training: Using animated conversational agents in the classrooms of profoundly deaf children," In Proceedings of ESCA/SOCRATES Workshop on Method and Tool Innovations for Speech Science Education, London, UK, Apr 1999.

[11] Cole, R., T. Carmell, P. Connors, M. Macon, J. Wouters, J. de Villiers, A.Tarachow, D. Massaro, M. Cohen, J. Beskow, J. Yang, U. Meier, A. Waibel, P. Stone, G. Fortier, A. Davis, and C. Soland. "Intelligent animated agents for interactive language training". In STiLL: ESCA Workshop on Speech Technology in Language Learning, Stockholm, Sweden, May 1998.

[12] Cole, R., S. Sutton, Y. Yan, P. Vermeulen, M. Fanty, "Accessible technology for interactive systems: A new approach to spoken language research," Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Seattle, WA, May, 1998.

[13] Sutton, S., Cole, R., et al. "Universal Speech Tools: the CSLU Toolkit," In Proceedings of the International Conference on Spoken Language Processing (ICSLP), pages 3221-3224, Sydney, Australia, November 1998.

[14] Pellom, B., Ward, W., and Pradhan, S. "The CU Communicator: An Architecture for dialogue Systems," Proceedings of International Conference on Spoken Language Processing, Beijing, China, October 2000.

[15] Cole, R., Novick, D., Fanty, M., Vermeulen, P., & Sutton, S. Experiments with a Spoken Dialogue System for

Taking the U.S. Census. *Speech Communication*, 1998.

[16] Dayal, U., et al. "The HiPAC Project: Combining Active Databases and Timing Constraints," ACM SIGMOD Record, 17(1), March 1988.

[17] Stonebraker, M. Hanson, E.N. and Potamianos, S., "The POSTGRES Rule Manager," IEEE Transactions on Software Engineering, 14(7):897-907, July 1988.

[18] Chakravarthy S., E. Anwar, L. Maugis, and D. Mishra, "Design of Sentinel: An Object-Oriented DBMS with Event-Based Rules. Information and Software Technology," 39(9):555--568, London, Sept. 1994.

[19] Widom, J. (ed.), Active Database Systems: Triggers and Rules for Advanced Database Processing, Morgan Kaufmann, San Francisco, California, 1996.

[20] Spencer, B., "Business Rules vs. Database Rules. A Aosition Statement," Object-Oriented Technology, ECOOP'98 Workshops, Demos, and Posters, Proceedings, 1998, p. xxii+573, 200-1.

[21] Ulrich, W.M., "Knowledge Mining: Business Rule Extraction and Reuse," Cutter IT Journal, Nov. 1999, vol.12, no.11, 21-26.

[22] Seiler, H., "Managed Business Rules: A Repository-based Approach," PC AI, July-Aug. 1999, vol.13, no.4, p. 16-19.

[23] Haley, P., "Agile business rule processing," PC AI, March-April 1999, vol.13, no.2, p. 35-9.

[24] Plotkin, D., "Business rules everywhere," Intelligent Enterprise, March 1999, vol.2, no.4, p. 36-44.

[25] Lam, H. and Su, S.Y.W., "Component Interoperability in a Virtual Enterprise Using Events/Triggers/Rules," Proc. of OOPSLA '98 Workshop on Objects, Components, and Virtual Enterprise, Vancouver, BC, Canada, Oct. 18-22, 1998, pp. 47-53.

[26] Lee, M., "Event and Rule Services for Achieving a Web-based Knowledge Network," Ph.D. Dissertation, Department of Computer and Information Science and Engineering, University of Florida, April, 2000; http://www.cise.ufl.edu/tech-reports/tech-reports/tr00-abstracts.shtml, TR 002.

[27] Lee, M., Su, S. Y. W. and Lam, H., "A Web-based Knowledge Network for Supporting Emerging Internet Applications," paper submitted to the World Wide Web Journal, 2000.

[28] Workflow Management Coalition, "Interface 1: Process Definition Interchange Process Model," 1998, http://www.aiim.org/wfmc/mainframe.htm.

[29] Su, S. Y. W. and Lam, H., "Iknet: Scalable Infrastructure for Achieving Internet-based Knowledge Network," Proceedings of *International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet*, l'Aquila, Rome, Italy, July 31-Aug. 6, 2000. 80 pp. (full length paper in CD).

[30] Lee, M., Su, S. Y. W., and Lam, H., "Event and Rule Services for Achieving a Web-based Knowledge Network," to appear in the Proceedings of the First Asia-Pacific Conference on Web Intelligence (WI-2001), Maebashi City, Japan, Oct 23-26, 2001.

[31] Ellison, C., "The Nature of a Usable pki," Computer Networks, (31):823--830, November 1999.

[32] Perlman, R. "An Overview of PKI Trust Models", IEEE Network, 13(6):38--43, November 1999.

[33] Chokhani, S, "Toward a National Public Key Infrastructure," IEEE Communications Magazine, Volume: 32 Issue: 9 , Sept. 1994 Page(s): 70 -74

[34] Nyanchama, M., and Osborn, S., "The Role Graph Model and Conflict of Interest", ACM Transactions on Information and System Security, 2(1), February,1999, pp. 3-33.

[35] Ferraiolo, D. F., Barkley, J. F. and Kuhn, D. R., " A role based access control model and reference implementation within a corporate intranet", ACM Transactions on Information and System Security, 2(1), February 1999.

[36] Sandhu, R., Ferraiolo, D., and Kuhn, R., "The NIST Model for Role-Based Access Control: Towards A Unified Standard", In Proc. ACM RBAC- 2000.

[37] Blaze, M., Feigenbaum, J., Lacy, J., "Decentralized Trust Management", Proceedings of 1996 IEEE Symposium on Security and Privacy, May, 1996.

[38] Herzberg, A., Moss, Y. and Mihaeli, J., "Access Control Meets Public Key Infrastructure," IEEE Symposium on Security and Privacy 2000.

[39] N. H. Kapadia and J. A. B. Fortes, "PUNCH: An Architecture for Web-Enabled Wide-Area Network-Computing," special issue on High Performance Distributed Computing of *Cluster Computing*: *The Journal of Networks, Software Tools and Applications*, Vol. 2, No. 2, September 1999, pp. 153-164.

[40] N. H. Kapadia, J. A. B. Fortes, and M. S. Lundstrom, "The Purdue University Network-Computing Hubs: Running Unmodified Simulation Tools via the WWW," *ACM Transactions on Modeling and Computer Simulation* special issue on Web-based Modeling and Simulation, 2000.

[41] N. H. Kapadia, R. J. O. Figueiredo, and J. A. B. Fortes, "The PUNCH WWW Portal to Tools for Computer Architecture Education," *IEEE MICRO* special issue on Computer Architecture Education, June 2000.

[42] N. H. Kapadia and J. A. B. Fortes, "PUNCH: An Architecture for Web-Enabled Wide-Area Network-Computing," special issue on High Performance Distributed Computing of *Cluster Computing*: *The Journal of Networks, Software Tools and Applications*, Vol. 2, No. 2, September 1999, pp. 153-164.

[43] S. Adabala, N. H. Kapadia, and Jose' A. B. Fortes , Performance and Interoperability Issues in Incorporating Cluster Management Systems within a Wide-Area Network Computing Environment, Supercomputing 2000: High Performance Networking and Computing. November 2000, Dallas, Texas.

[44] Renato J. Figueiredo, Nirav H. Kapadia and Jose A. B. Fortes, "The PUNCH Virtual File System: Seamless Access to Decentralized Storage Services in a Computational Grid,"*Tenth* IEEE International Symposium on High Performance Distributed Computing. August 2001.

[45] Object Management Group "CORBA services: Common Object Service Specification," Technical Report, Object Management Group, July 1998.

[46] Microsoft DCOM, http://www.microsoft.com/com/tech/dcom.asp.

[47] http://java.sun.com/products/ejb/docs.html.

[48] Object Management Group, "UML 1.1 Specification," http://www.omg.org, Nov. 19, 1997.

[49]`Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., and Lorensen, W., Object Oriented Modeling and Design, Prentice-Hall, 1991.

[50] Subcommittee 4 of ISO Technical Committee 182(1992), "Product Data Representation and Exchange – Part 11: The EXPRESS Language Reference Manual," ISO Document, ISO DIS 10303-11, August 1992.

[51] Cattell, R. G. G. (ed.), "The Object Database Standard, ODMG-93: Release 1.2," Morgan Kaufmann Publishers, San Francisco, CA, 1996.

[52] Object Management Group "CORBA services: Common Object Service Specification," Technical Report, Object Management Group, July 1998.

[53]Data Access Technologies, "CDL: Business Object Component Architecture Encyclopedia," http://www.dataaccess.com/dat, May 21, 1998.

[54] Su, S.Y.W., Lam, H., Arroyo, J., Yu, T. F., and Yang, Z., "An Extensible Knowledge Base Management System for Supporting Rule-based Interoperability among Heterogeneous Systems," Proc. of the Conf. on Information and Knowledge Management, Baltimore, MD., Nov. 28-Dec. 2, 1995, pp. 1-10.

[55] Su, S.Y.W., Lam, H., Yu, Tsae-Feng, Arroyo-Figueroa, Javier, Yang, Zhidong, Lee, Sooha, "NCL: A Common Language for Achieving Rule-based Interoperability among Heterogeneous Systems," *Journal of Intelligent Information Systems* (JIIS), Special Issue on Intelligent Integration of Information, 6, 1996, pp. 171-198.

[56] Su, S. Y. W., Liu, Y. Z., Meng, J., Lee, M. S., Lam, H., "Distributed and Concurrent Processing of Business Object Documents in Support of e-Enterprise Integration," 4th International Enterprise Distributed Object Computing Conference (EDOC 2000), Makuhari, Japan, Sept. 25-28, 2000, pp. 107-111.

[57] http://www.uddi.org/specification.html; http://www.uddi.org/bestpractices.html

[58] http://msdn.microsoft.com/xml/general/wsdl.asp (specification); http://www.devxpert.com/tutors/wsdl/wsdl.asp (introduction)

[59] http://www.develop.com/soap/soapfaq.htm#1 (faq); http://www.w3.org/TR/SOAP/ (Specification); http://www.w3schools.com/soap/default.asp

[60] Su, S.Y.W. and Yu, T. F., "Distributed Information Mediation and Query Processing in a CORBA Environment," *International Symposium on Digital Media Information Base (DMIB '97)*, Nara, Japan, Nov. 26-28, 1997, pp. 120-131.

[61] Semeczko, G., Su, S. Y. W., Yu, T. F., and Wang, F., "Supporting Distributed Query Processing in a Heterogeneous Environment," Proceedings of the *Fifth International Conference on Database Systems for Advanced Applications*, Melbourne, Australia, April 1-4, 1997.

[62] [SU98] Su, S.Y.W., Huang, Y. and Akaboshi, N., "Graph-based Parallel Query Processing and Optimization Strategies for Object-oriented Databases," J. on Distributed and Parallel Databases, Vol. 6, No. 3, 1998.

[63] [LI01] Li, H. and Su, S. Y. W., "Business Object Modeling, Validation and Mediation for Integrating Heterogeneous Application Systems," paper to appear in the Journal of System Integration, 2001.

[64] R.D. Brown", "Example-Based Machine Translation in the Pangloss System", Proceedings of the Sixteenth International Conference on Computation Linguistics," pp. 169-174, 1996.

[65] Carbonell. J. G., Yang, Y,. Frederking, R. E., Brown, R., Geng, Y., and Lee, D., "Translingual Information Retrieval: A Comparative Evaluation", Translingual Information Retrieval, Nagoya, Japan, (Distinguished paper award), 1997.

[66] P. Sheridan and J.P. Ballerini, “Experiments in Multilingual Information Retrieval using the SPIDER System.", 19th Ann Int ACM SIGIR Conf. on Research and Development in Information Retrieval, 1996, pp. 58-65.

[67] Nirenburg, S. Carbonell, J. G., Tomita, M.,and Goodman, K."Knowledge-Based Machine Translation", Morgan Kaufmann Inc, Year=1991, San Mateo, CA

[68] A. Berger, P. Brown, S. Della Pietra, V. Della Pietra, J. Gillett, J. Lafferty, R. Mercer, H. Printz, L. Ures,”The Candide system for machine translation", Proceedings of the ARPA Workshop on Speech and Natural Language, Morgan Kaufman Publishers,1994.

[69] Jaime Carbonell and Jade Goldstein, “MMR and Diversity-Based Reranking for Reordering Documents and Producing Summaries,” Proceedings of the 21st meeting of International ACM SIGIR Conference, Melbourne, Australia, 1998, August, pages 335-336.

[70] Goldstein, J., Mittal, V., Carbonell, J., and Kantrowitz, M., “Multi-Document Summarization by Sentence Extraction,” ANLP/NAACL 2000 Conference, April 2000, pp. 40-48.

[71] M. Yajnik, S.B. Moon, J. Kurose, D. Towsley. “Measurement and modeling of the temporal dependence in packet loss,” *Proc. INFOCOM’99*, March 1999.

[72] S.B. Moon, P. Skelly, D. Towsley. “Estimation and Removal of Clock Skew from Network Delay Measurements” , *Proc. IEEE Infocom'99* (New York, USA, March 1999).

[73] M. Yajnik, J. Kurose, D. Towsley. “Packet Loss Correlation in the MBone Multicast Network”, *Proc. IEEE Global Internet'96* (London, U.K., November 1996).

[74] K. Salamatian, S. Vaton. “Hidden Markov modeling for communication channels,” to appear in *Proc. SIGMETRICS 2001*, June 2001.

[75] S. Sen, J. Rexford, D. Towsley. “Proxy Prefix Caching for Multimedia Streams,” *Proc. INFOCOM’99*, March 1999.