ANE MIS/GIS PILOT SUPPORT PROGRAM

INTEROPERABILITY REPORT

February 12, 2007

This publication was prepared by Booz Allen Hamilton for the United States Agency for International Development.

Table of Contents

1 Executive Summary 3

1.1 Background 3

1.2 Summary of Key Findings 3

1.3 Summary of Recommendations 4

2 Interoperability Task in Detail 7

2.1. Task requirements, goals, objectives 7

2.2. Interoperability as implemented in pilot 8

Making Data Interoperate 8

Developing an interoperable MIS/GIS 9

2.3. Recommendations 12

1 Executive Summary

1.1 Background

This report summarizes the interoperability task of the U.S. Agency for International Development Asia and the Near East Bureau (USAID ANE) Management Information System/Geographic Information System (MIS/GIS) Pilot Support Project. This report provides the key findings and recommendations for developing a system capable of interoperating with Mission field systems as well as other Agency systems such as the Global Acquisition System (GLAS), Phoenix, and the Foreign Assistance Coordination and Tracking System (FACTS).

Projects like the ANE MIS/GIS are often conducted in a constantly evolving institutional environment. In addition to the previously mentioned projects running concurrently with the development of the ANE MIS/GIS, there was a significant reorganization of the Agency's taxonomy for planning and reporting foreign assistance. The Office of the Director of Foreign Assistance established a new Strategic Planning Framework that provided a prescribed hierarchy and associated naming convention for planning development projects across the State Department and USAID. These institutional factors influenced *how* the Booz Allen team approached the task and our activities under this task, but did not alter the end goal of providing an interoperable system.

1.2 Summary of Key Findings

The first key finding is that the ANE MIS/GIS occupies a unique niche in the USAID information ecology in that it is the bridge between operationally-oriented field systems and the strategic- and planning-level systems of Washington headquarters. No other existing system is equipped to seamlessly aggregate and summarize data from the activity level up through the Program or Functional Objective level. This bridging capacity is essential to enabling Washington headquarters staff to track the allocation of resources in the field, and be capable of effectively and efficiently responding to adhoc queries from overseers on activity and program status.

A second finding is that the majority of data collected and documented at the Mission- and operational-level is insufficiently structured for automating the reporting process and therefore relieving the extraordinary efforts required today to satisfy the many ad hoc queries from Congress and other key stakeholders. Because the emphasis of the MIS/GIS is on reporting using an open and extensible architecture, the Booz Allen team focused from the beginning on finding data and making it exchangeable between systems, i.e., data interoperability. Booz Allen solicited data of all types from several Missions. Most of the data that was provided either did not conform to any standard schema or format and none came with descriptive metadata. In several cases, we received finished reports rather than raw data, and raw data is what is needed for reporting.

The third and final finding is that there is no effective governance mechanism within USAID that exerts the policy and technical oversight necessary to drive Agency-wide information-sharing, including what is needed to implement and administer a multi-tiered, integrated system like the ANE MIS/GIS. There are numerous talented and dedicated individuals who do a very good job of maintaining the existing information infrastructure, but their efforts are not coordinated through a formal governance structure, such as a council or working group. As a result, policy regarding standards, compliance, data structure, technical oversight, and such tend to be overlooked.

1.3 Summary of Recommendations

Based on the findings, we offer several interrelated recommendations for USAID that will better enable interoperability:

- Establish and charter an information **governance body**
- Implement a formal, structured, and standardized **data model** and **reference architecture**
- Implement **data collection** methods based on MIS/GIS architecture
- Adopt recommended **standards** for agency-wide usage
- Implement system **compliance testing**

Governance Body: A formal governance body for agency-wide information should be established to provide a focal point for the development of Agency-wide information policy. (Informal and ad hoc working groups of subject matter experts already meet to address specific issues but do not enjoy a significant mandate to lead policy development.) This group should be drawn from information technology experts and practitioners, subject matter experts familiar with Agency operations, and others with a particular relevant expertise or interest. This group should also work closely with the Chief Information Officer (CIO) for consistent application of policies and procedures.

The thrust of this group should be to develop policies for the Agency's information resources. For example, a reference architecture for information systems could provide guidance for the development of new field systems that could ensure their interoperability and reuse lessons learned, thereby saving costs on development and maintenance. The governance body should also assume responsibility for the adoption, development, and maintenance of Agency-wide information standards and issue compliance testing for all systems to ensure they interoperate with the Agency's information infrastructure.

There is a cost associated with a dedicated governance body in the form of time and resources necessary to provide oversight to the Agency's programs. However, this cost is more than offset

by efficiencies gained through coordinated data and information policies, standardized business rules and terminology, and an enterprise (as opposed to *ad hoc*) approach to system development and deployment.

The benefits conferred by a governance body include:

- A forum to focus attention on issues related to the use of information technology
- A process for advancing developments in standards, information processing, and information technologies
- A community of interest focused on mission objectives
- Mature standards and business rules for information handling

Data Model and Reference Architecture:

ANE should consider adopting and implementing the ANE Feature Model and Business model demonstrated in the MIS/GIS. To provide objective, data-driven reporting, it is necessary to have a data structure with a very fine degree of resolution from which data can be distilled into a strategic-level report. This means a data model that not only complies with FY07 Strategic Planning Guidance, but is capable of tracking the project level concerns of ANE's Missions. The ANE Feature Model is a formalized structure that provides the necessary ground-level resolution, making it well-suited for the widely varied conditions found across ANE's Missions. Moreover, the MIS/GIS data models are highly consistent with the EIS data model and the FACTS system, making it possible to leverage work that has already been done by USAID.

Also, ANE should adopt a reference architecture to guide field system development and improvements. A reference architecture is a conceptual model of how a system should work. It describes what components the overall system is made of, but does not dictate how they are integrated, or how they are to be implemented. The appeal of a reference architecture is that, for an agency with widely decentralized operations under widely varying operating conditions, implementation can be tailored to local conditions as long as the system passes information in accordance with Agency protocols. This would seem to fit the need for Missions to be capable of establishing systems tailored to their particular needs, but at the same time ensuring interoperability across regions and programs. Furthermore, given the need to integrate existing corporate systems with heterogeneous field systems, we strongly recommend USAID develop a reference architecture in the style of a Services Oriented Architecture (SOA).[1]

The benefits of a data model and reference architecture are:

- Ensures the technology serves Mission goals for their tailored needs;

[1] See USAID ANE MIS/GIS Demonstration Report for a discussion on SOA architectures.

- Provides an extensible and replicable approach across Missions ensuring interoperability across regions and programs;
- Supports data tracking at a fine resolution, which is necessary for relevant reporting; and
- Promotes accurate aggregation of operational data for high level reports.

Data Collection: There is no escaping the fact that good reporting begins with good data. Good data is compiled from the lowest levels of granularity and aggregated where it makes conceptual sense so that it can be rolled effectively into a strategic-level report for managers and policy-makers. The question then becomes how to acquire the best data. The ANE MIS/GIS provides a tool set for use at the operational level that supports reporting within the framework established in the FY07 Operational Planning Guidance. This data entry tool enables Missions and implementing partners to link features (roads, schools, hospitals, etc.) with activities and instruments and in turn, link those to program sub-elements, which in turn can be linked to program areas, and so on. This provides staff at operational levels with the detail they need, while providing an accurate, data driven summary report to headquarters.

Benefits of the ANE MIS/GIS for data collection include:

- Providing tools tailored to Mission needs benefits users in the field;
- Standardizing data collection at the feature-level ensures that results reporting is data driven; and
- Standardized tool sets could reduce the burden of data calls and specialized queries to Missions from Washington.

Standards: The Internet has greatly increased the use of modular and loosely coupled software. In this environment, standards are becoming absolutely essential for managing the complex inter-relationships of all the moving parts of an Internet-based system and thus making them interoperable. Standards make these relationships understandable, and therefore more reliable and cost-effective.[2] This removes risk for the implementing organization and also helps build organizational resilience by providing an extensible model for re-tooling when organizational priorities change.

The data governance body should have two primary responsibilities with regard to standards. The first is to be responsible for developing agency-wide information standards. Agency-wide

[2] The USAID ANE MIS/GIS Standards Reference document discusses the rationale for standards, as well as applications and implementation of information standards, particularly for geospatial information. Please refer to that document for a more complete discussion of this subject.

standards could be profiles of other Federal Information Processing Standards (FIPS), Federal Geographic Data Committee (FGDC), American National Standards Institute (ANSI), or International Standards Organization (ISO) standards. Or, they could be drawn from a particular business practice within the Agency. The second responsibility is to adopt, implement, and ensure compliance with industry-wide standards consistent with the information goals of the Agency. These include standards such as World Wide Web Consortium (W3C), Organization for the Advancement of Structured Information Standards (OASIS), the Open Geospatial Consortium (OGC), as well as ANSI and ISO.

The benefits of adopting standards include:

- Formalized system and data knowledge;
- Participation and support from broader standards community; and
- A formal place for corporate knowledge

Compliance testing:

The governing body should apply a set of performance criteria associated with field systems. Systems must be able to collect and store data consistent with the ANE data models and to exchange that data freely with field systems and with enterprise-level systems. For new systems, or system upgrades, criteria for compliance can be incorporated into procurement language.

2 Interoperability Task in Detail

2.1. Task requirements, goals, objectives

Interoperability can be defined as the ability of two computer systems to interact, machine-to-machine, to produce the desired result. Additionally, it can be defined as the ability of two heterogeneous information systems to interact with a minimum of ad hoc intervention.

Consistent with this definition, the task goal is to deliver a system that is interoperable with its constituent parts, and provide repeatable methods for adding new data and related subsystems.

Requirements for the interoperability phase of the Standards and Interoperability task support these goals:

- Implement, to the degree practicable, the standard specifications of the Open Geospatial Consortium (OGC);
- Document steps required to convert existing source data and render it consistent with the MIS/GIS data model;
- Provide linkages between text-based information and geographic data

From an implementation perspective, the requirements serve to drive three major activities:

- Test assumptions about the feature model to facilitate reporting;
- Determine whether ETL (extract-transform-load) tools can be used effectively to facilitate interoperability; and
- As needed, determine whether the MIS/GIS toolset can utilize data in its native state.

2.2. Interoperability as implemented in pilot

Making Data Interoperate

From the outset, the plan for this task was to evaluate methods for moving Mission-level data into the MIS/GIS so that it could then be manipulated to produce reports. To accomplish this, the Booz Allen team requested a wide variety of data from Missions in the ANE Bureau. The team anticipated receiving a wide cross-section of spreadsheets and small databases which could either be converted to the MIS/GIS data model or 'mapped' from the local, mission-level data model to the MIS/GIS data model. Sample data were requested and obtained from Indonesia, West Bank/Gaza, Afghanistan, Egypt, among others. The plan supposed that if data were accessible to the MIS/GIS, the GIS and Business Intelligence (BI) toolset could be used to build reports.

With data in hand, the task is to map existing data into the MIS/GIS database design. More specifically, the data from the Missions must be mapped into the feature model as well as into the revised business model that reflects the new FY2007 Operational Planning Guidance. The data were evaluated for their ability to match up to these models.

The best, most usable data received by the project team included the schools and clinics in Afghanistan; relatively robust data from West Bank/Gaza; and a robust database from Indonesia. Some or all of these data sets were incorporated into the MIS/GIS pilot. In the case of the Afghan school and clinic data, it was necessary to supplement the data by creating additional attributes to support a multidimensional report. In other cases, only the position and feature type (e.g. 'clinic', 'school') and associated position (e.g. latitude/longitude coordinate pair) was used and additional attributes were created to supplement the reporting process. Where necessary to invent data and attributes, it was done in accordance with the MIS/GIS data model.

In short, results of the call for data dictated a course of events different from what was anticipated. Only a small amount of the data was usable for reporting, and in all cases it had to be further embellished to support the MIS/GIS demonstrations. While integrating the data into the MIS/GIS, we faced several challenges:

- **Non-standard format:** The data received by the project team was not in a consistent format (even within the same mission) nor did it conform to any standard schema or format. Each spreadsheet contained a different level of detail.
- **Lack of metadata:** There was no descriptive metadata the team could use to determine what the data purported to describe or activities they supported. Crucial information, such as primary contacts, data creation process, accuracy, currency, data refresh cycle, etc. was not provided.
- **Not really data:** Some data were finished reports in a discursive format, rather than raw data that could be used to support a report. In these cases, the project team was required to 'reverse engineer' by extracting the data in a report and turn it in to usable data, which could be integrated in to the MIS/GIS application schema.

Due to these problems, it was necessary for the project team to construct mock data sets based loosely on the inputs from the missions. The mock data sets would support both the feature model and the new business model based on FY2007 Operational Planning Guidance.

Developing an interoperable MIS/GIS

In contrast to the issues with data, application development posed a set of challenges centered on making several commercial software packages work together. The most important of these challenges was geo-referencing the text and alphanumeric data, and to do it in such as way as to make the transition between data types as seamless as possible to the end user. To do this, the development team took advantage of alliances between software vendors to knit several small modules together and make them accessible through one centralized portal. The vendors of the major software components, including Earthwhere, ESRI, IBM, MapInfo, and Microstrategy, contributed significantly to the integration effort. The bulk of the task was focused on integrating Business Intelligence (BI) tools with GIS and integrating them within a portal structure. For example, MapInfo and Microstrategy support a fairly high degree of integration with each other's products. MapInfo's Location Intelligence Component interacts well with Microstrategy, spatially enabling the reporting functions. The overall approach to building a loosely-coupled system was a sound strategy and has performed well in the MIS/GIS testing environment.

The process of evaluating and generating data, and the challenges of building the reporting application led to several findings that have a bearing on the future implementation and deployment of the MIS/GIS.

Data collection: The inadequacy of the Mission-level data to provide sufficient detail in a structured format is regarded as a critical finding for this project and one that deserves continued attention. The uniqueness of the MIS/GIS rests on the ability to connect numerous Mission-level activities with strategic-level processes carried out at the headquarter-level while at the same time shielding headquarters from the *minutiae* that attends operational-level activities. Integrating

information from field systems is essential to providing that connection. The development of the MIS/GIS has shown that baseline data needs are not being met by the current constellation of field systems.

The experience of the project team is that at the Mission level, there are hundreds of systems, each with a unique scope and focus, and deployed in a very wide range of operating environments. The wide variety of data formats received from ANE missions for this project suggest that most of these systems are not configured to collect and share data in a manner consistent with MIS/GIS project objectives, or any other system. For example, the extensive reliance on spreadsheets and desktop databases suggests that most field systems are not collecting data in a formally structured relational database management system. This greatly inhibits USAID's ability to exchange existing data or even make it available within the local mission.

The challenge will be to develop design principles that account for the wide ranges of operating conditions, such as intermittent electrical power, spotty Internet access, and low levels of computing power, while at the same time collecting data sufficient to support operations as well as strategic-level reporting needs.

This finding has important implications for the broader USAID information environment. If data cannot move freely between the constellation of Agency systems and be aggregated between the various systems, then Agency-level systems will place a severe burden on Bureaus and Missions to manually provide appropriate data inputs. Similarly, users need the ability to drill down from aggregated and generalized results to Mission-level detail. The MIS/GIS system is a significant step in building the necessary infrastructure to support these specific-to-general and general-to-specific views of USAID's information. However, the amount of data Booz Allen had to mock up to support these views indicates that data governance should be a high priority if USAID's investments in information technology are to be realized.

Metadata: Metadata is information that describes aspects of a data set such as accuracy, currency, creation and update processes, and maintaining organization. This information is used by parties exchanging data in order to understand what data are being recorded, field names, data types, and degree of fit for a particular business purpose. Without descriptive metadata, users that receive data either have to contact the originating agency to provide a description or guess at the meaning of the data.

None of the data obtained from USAID missions was accompanied by any descriptive metadata. In the course of developing the MIS/GIS, the lack of metadata reduced the value of data obtained from Missions and increased the costs associated with rendering it useful. Also, the lack of even minimal metadata poses a risk to USAID in that their data resources are not part of the corporate knowledge base. If the individuals with specialized knowledge of the data leave the organization, their specialized knowledge and expertise leaves with them.

Effort associated with software integration: To integrate commercial off-the-shelf software (COTS) BI tools with GIS applications requires custom code be generated for each report type. Although the GIS software from MapInfo was specifically developed to interoperate with the BI

tool, MicroStrategy, because every report type is unique, an additional level of effort is required to support the drill-up and drill-down feature within the map application for each report. This is because a GIS looks for geo-referenced information in a specific format in order to render a map. So, for each report type, the Booz Allen team wrote XML code to pass data from MicroStrategy to MapInfo in the format needed by MapInfo. In the next release of MapInfo's Location Intelligence software, they state it to be a more "out-of-the-box" ready solution. If true, the challenges of the BI/GIS integration encountered in the MIS/GIS demonstration may be mitigated to some extent when using MapInfo as the GIS component.

In addition to the BI/GIS integration, the Booz Allen team found it necessary to write software code for the geospatial visualization tool. Generally, the user interface for GIS applications that support advanced operations, such as being able to select a series of points (E.g. schools) and show the underlying information about those points (E.g. name of school, number of students attending, number of students by gender etc.) appear complex to the novice user. To reduce the complexity of a standard GIS interface, the Booz Allen team determined it would be best to write a custom user interface. This required writing code between the map server and the application server in order to render the user interface and display the data.

Both of these findings related to COTS software integration, have a direct effect on the total cost of ownership for software. To the extent possible, these costs should be accounted for during the early planning phases of the project. The MIS/GIS Demonstration provides a source for understanding where custom code will have to be generated, and a framework for scoping how much custom code could be required.

Integration through the portal: The portal serves as the front door, literally and figuratively, to the MIS/GIS. Software vendors typically provide an application programming interface (API) that allows developers and system integrators to access the business logic of the software.

Although each portal consists of proprietary software, each with its own API, using the portal as the focal point to organize the workflows proved to be a very successful approach. The portal provides a place to integrate different software modules, thereby taking advantage of their inherent reusability and portability. This allows for rapid extension and customization of the MIS/GIS as the organization matures and conditions change. It also provides for a simple and intuitive interface that minimizes the demand for special knowledge on the part of the end user.

One key advantage of the portal approach was that software components could be loosely coupled and integrated with relative ease. This means the user interface can be accessed through the portal while the application and associated business logic can be hosted elsewhere. The portal applications used in the MIS/GIS (e.g. BEA, IBM Websphere, etc.) all utilize proprietary API's, and as long as the development team uses the same API, integration in the portal is relatively

straightforward. Although each of the portals have proprietary features, all follow the JSR 168 specification[3] and behave in a manner consistent with the industry specification.

2.3. Recommendations

Our recommendations are inter-related and focused on data resources and a process for governing the free flow of that data. As with many organizations, data and information is the currency of USAID's operations. The recommendations offered here focus on a structured group within ANE whose charter includes governing information processes, a template for system architecture, tools for collecting self-describing data, and system compliance. These recommendations, if implemented, will boost the ANE Bureau's data handling capabilities and enable the Agency to distribute more usable information across the organization.

The recommendations are:

- Establish and implement a governance process for all information systems
- Develop a standardized reference architecture and data model for field systems
- Provide a data entry application for field systems that do not otherwise capture the baseline data needed for reporting
- Require compliance testing for new field systems and subsequent improvements to, and expansion of existing field systems

Establish and implement a governance process for all information systems. Attention across USAID has focused on the Executive Information System (EIS), the Enterprise Data Model, FACTS, GLAS and the ANE MIS/GIS. An oversight body is needed that can offer Agency-wide coordination for information technology and knowledge investments at all levels of the organization. This oversight body needs a charter describing the scope and purview of the governance body, and should be drawn from subject matter experts, technical experts, project managers, and other individuals with either a specific interest or skill relevant to the working group.

The duties of the governance body should include

- Establishment and implementation of Agency-wide information standards;

[3] Java Community Process Program. *Java Specification Requests*. Retrieved at: http://jcp.org/en/jsr/detail?id=168

- Development or adoption of a reference architecture for a federated system of systems, preferably in the style of a SOA;
- Adoption of formal, standards-based data models and data standards;
- Adoption of documentation requirements, including metadata; and
- Compliance testing for these criteria.

There is a cost associated with a dedicated governance body in the form of time and resources necessary to provide oversight to the Agency's programs. However, this cost is more than offset by efficiencies gained through coordinated data and information policies, standardized business rules and terminology, and an enterprise (as opposed to *ad hoc*) approach to system development and deployment.

Develop a standardized reference architecture and data model for field systems

A standards-based reference architecture is an abstract and idealized template for an information system. The purpose of a reference architecture is not strict guidance for implementation, but rather as open-ended template to facilitate the planning and implementation of information systems. Adopting a reference architecture would give latitude to missions that must adapt to a wide variety of local conditions, while enabling them to still conform to agency-wide standards for information collection, storage, retrieval, and reporting.

There are several examples of standards-based reference architectures relevant to USAID. These include the OASIS Reference Model for Services Oriented Architecture (SOA-RM)[4], the ISO Reference Model for Open Distributed Processing (RM-ODP)[5], and the Open Geospatial Consortium (OGC) Reference Model.[6]

The benefits of using one or more of these reference models is that USAID will adopt a formal model not just for implementing a technology, but making the technology serve Agency goals at headquarters and in the field.

---

[4] Organization for the Advancement of Structured Information Standards (OASIS). *Reference Model for Service Oriented Architecture 1.0.* Committee Specification 1, 2 August 2006. Retrieved at: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=soa-rm

[5] International Organisation for Standards (ISO). *Reference Model for Open Distributed Processing*. ITU-T Rec. X.901 | ISO/IEC 10746-1 to ITU-T Rec. X.904 | ISO/IEC 10746-4, Retrieved at: http://www.enterprise-architecture.info/Images/Documents/RM-ODP2.pdf

[6] Open Geospatial Consortium. *OGC Reference Model.* OGC 03-040. Retrieved at: http://portal.opengeospatial.org/files/?artifact_id=3836

With regard to data models, ANE should adopt and implement the MIS/GIS Feature Model and Business model. These models provide a formal structure with the degree of granularity necessary to support mission operations and project management, and can be freely aggregated and summarized to provide strategic-level reporting. The ANE Feature Model is a formalized structure that accommodates nearly any activity such as school, road, or clinic construction; or training and outreach events making it well-suited for depicting the development work of ANE's Missions. Moreover, the ANE data models require little harmonization to be consistent with the EIS data model.

Provide a data entry application for field systems that do not otherwise capture the baseline data needed for reporting

.As stated elsewhere, ANE needs to take an approach to information systems that emphasizes creating data of value at the lowest operational levels of the organization and preserves the value of data as it is passed through the organization upward. Existing field systems meet the needs of the community charged with implementing the work of the Mission, and therefore are less well suited to supporting the many ways information must be reported. Standardized toolsets, specifically to collect operational, activity level data, must be implemented in order to meet baseline data needs for an agile reporting environment.

As discussed in the project findings, most existing field systems do not collect and expose data in ways that can be aggregated for reporting purposes. The ANE Feature Model described in the USAID ANE MIS/GIS System Design document dated April 24, 2006 provides a standardized structure, or framework, for mission data at the lowest levels of granularity. This granularity is essential because it builds in the flexibility analysts and managers need to develop accurate reports for any adhoc request. With this data structure in place, the need is to provide tools to end users to populate the databases.

Field systems also need to have tools to generate metadata so as data is exchanged between different parts of the organization, users have the means to evaluate the data. The GIS toolsets used in the development of the MIS/GIS have built-in metadata editors that simplify the job of compiling metadata, especially the more arcane geodetic information. These tools need to be extended so that they are available for all users and for all data collected. Of equal importance, USAID also needs to mandate the creation of metadata at all levels of data creation and maintenance, for implementing partners and for Agency personnel.

Require compliance testing for new field systems and subsequent improvements to, and expansions of existing field systems.

Requiring new and existing field systems to comply with these recommendations ensures the gradual improvement in ANE's information base and reporting capabilities. The governing body needs a set of performance metrics associated with field systems to make sure that the system:

- Meets operational needs,
- Supports the free exchange of data,
- Complies with ANE standards,
- Supports the requisite reporting capability

A compliance matrix can be generated from the reference architecture and from business rules and should be an activity undertaken by the governance body or working group. Field systems should be tested for compliance with:

- The MIS/GIS Feature Data Model
- The ANE Business Data Model
- Metadata development
- Data export to federated ANE systems
- Import of business data from Phoenix, FACTS, etc.

Compliance ensures complete data portability from top to bottom, thereby enabling strategic reporting.