ADULT LINKAGES PROJECT

PROJECT PROPOSAL

Submitted to

Chief Information Office
Information Technology Fund

by

Chief Administrative Office
Children and Family Services
Community and Senior Services
Health Services
Mental Health
Probation
Public Health
Public Social Services
Sheriff

September 29, 2006

TABLE OF CONTENTS

INFORMATION TECHNOLOGY FUND
PROJECT PROPOSAL

PROJECT TITLE
ADULT LINKAGES PROJECT

PROJECT LEADER
Name: Halil Toros
Phone: (213) 974-5162
Fax: (213) 620-7131
Email: htoros@cao.lacounty.gov

BACKGROUND

The County of Los Angeles currently serves roughly 65,000 participants in its General Relief (GR) program. In a given year, 120,000 unique participants enter the GR program, and more than 200,000 participants enter the program over three years. The Department of Public Social Services (DPSS) spends an average of $200 per GR participant per month. Although a substantial proportion of General Relief participants receive multiple services from separate county departments, the number served by different departments, as well as the amount spent to assist these needy adults, remain unknown. This lack of knowledge is exacerbated by the fact that more than one-third of the County's GR population is homeless. The project proposed here is designed to provide new information on the service utilization patterns of indigent adults and thereby lay the groundwork for cross-departmental service coordination and integration.

THE PROBLEM AND ITS SOLUTION

A substantial proportion of General Relief participants receive multiple services from separate county agencies, and there is currently no systematic coordination in the delivery of these services. As a result, many clients fail to consistently receive the services they need because they are ill-equipped to follow the rules and case management plans upon which receipt of services depend. A systematic coordination and integration of service delivery would help GR participants to receive these services in a timely, effective and efficient way.

This project will describe, for the first time, the extent to which GR recipients use services from departments other than DPSS. This information is vital because county dollars contribute substantially to programs serving indigent adults, and there is currently no central source from which to draw data on service utilization and estimated costs. The project will also serve as a model for exploring new ways to assist other target populations in the future, such as the elderly and children, as well as specific groups served by other departments, such as children in foster care, homeless families, and mental health patients. Participating departments will find the new information to be of significant value in planning future services.

DESCRIPTION OF THE PROJECT

The project proposed here is to be centered on participants in DPSS' General Relief (GR) program. The project creates a baseline database for an entry cohort of GR participants, as well as a record linkage methodology connecting the target population of indigent adults to the multiple departments from which they receive additional services. Linked data files will be loaded to a data warehouse. Separate county departments will have access to this integrated and centralized information on overlapping clients through a web-enabled technology. The present proposal covers the technology component of the project, which involves the acquisition of necessary software licenses, hardware equipment and consulting services.

DPSS is the lead agency providing basic services to GR participants. For this interagency collaboration project, DPSS will work in partnership with other county departments, each of which also provides services to GR participants. This inter-departmental collaboration will build a data warehouse containing systematic information on GR participants and additional data on services provided to these participants by Probation, the Department of Mental Health, the Department of Health Services, the Department of Public Health, the Department of Children and Family Services, the Sheriff, and the Department of Community and Senior Services.

The project is composed of several components as follows:

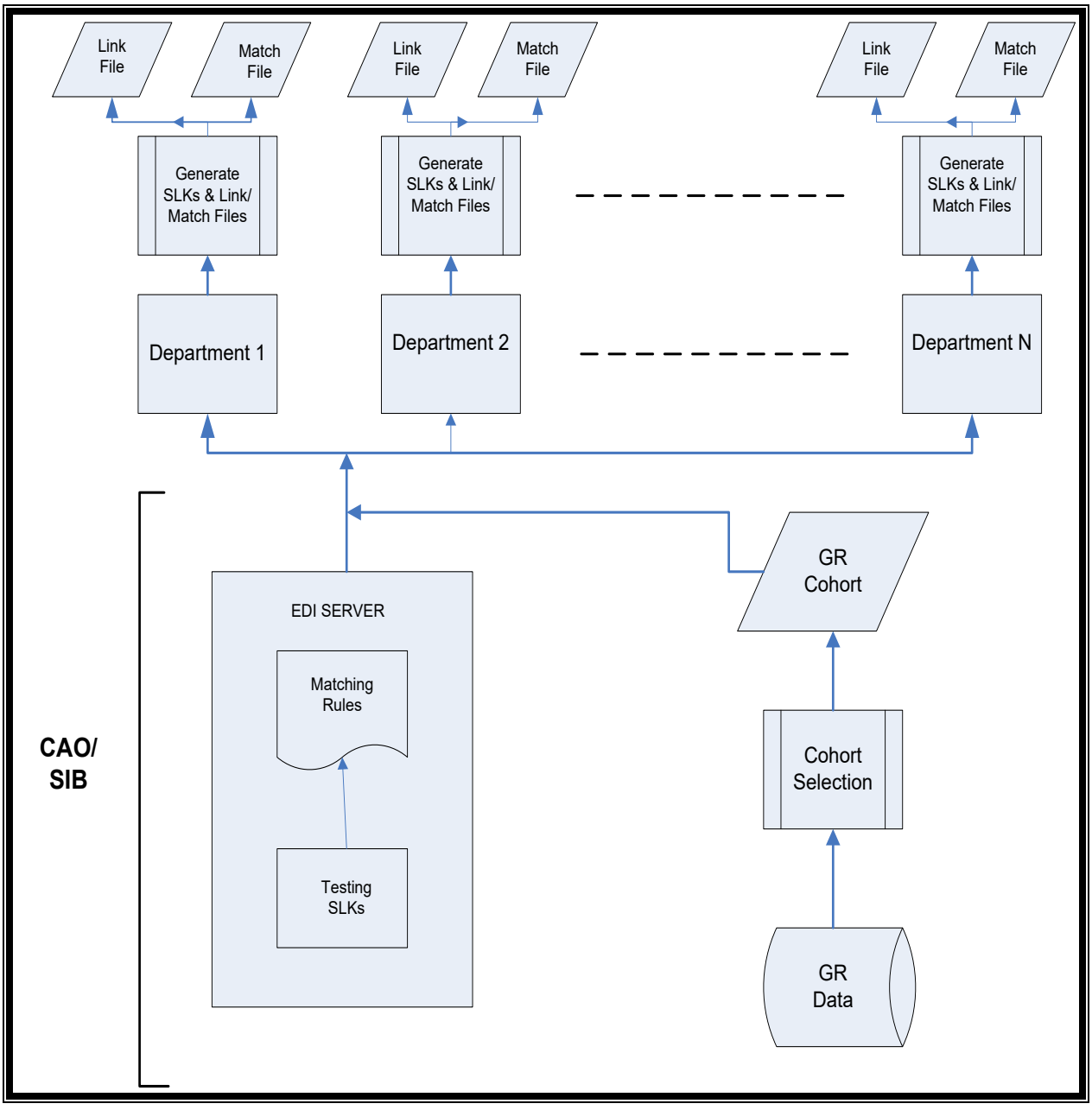
1. **Development of Record-Linkage Methodology:**

The Chief Administrative Office's Service Integration Branch (SIB) will develop the record-linkage methodology linking GR participants across departments. The development will be done using the SAS Enterprise Data Integration (EDI) Server, which features a data quality solution utilizing the Dataflux technology. This data quality solution not only accomplishes extensive data profiling and quality enhancements that can be used to understand, analyze, correct, reconcile and consolidate data, but also features a powerful match engine that can be used to link and integrate data across discrete sources.

This rules-based matching engine uses a combination of parsing and standardization rules, phonetic matching, and token-based weighting to strip the ambiguity out of source information. After applying thousands of complex rules to each field, the engine outputs a "match key" that, for the purposes of this proposal, can be referred to as the "statistical linkage key" (SLK). The SLK is an accurate representation of all versions of the same data—such as different spellings of a name or an address. The SLK offers pre-defined match rules and the match engine is extremely customizable, allowing users to define their own matching rules that may include any number of fields and conditions. SIB will test alternative matching rules to come up with the most accurate algorithms to link GR participants across departments. After the final matching rules and SLKs are determined using the EDI server, these algorithms will be transported to

participating departments for the purposes of conducting the record-linkage process. A simple flowchart of this step is shown in Figure 1.

Figure 1. Record-Linkage Methodology



2. Data Linkage Process:

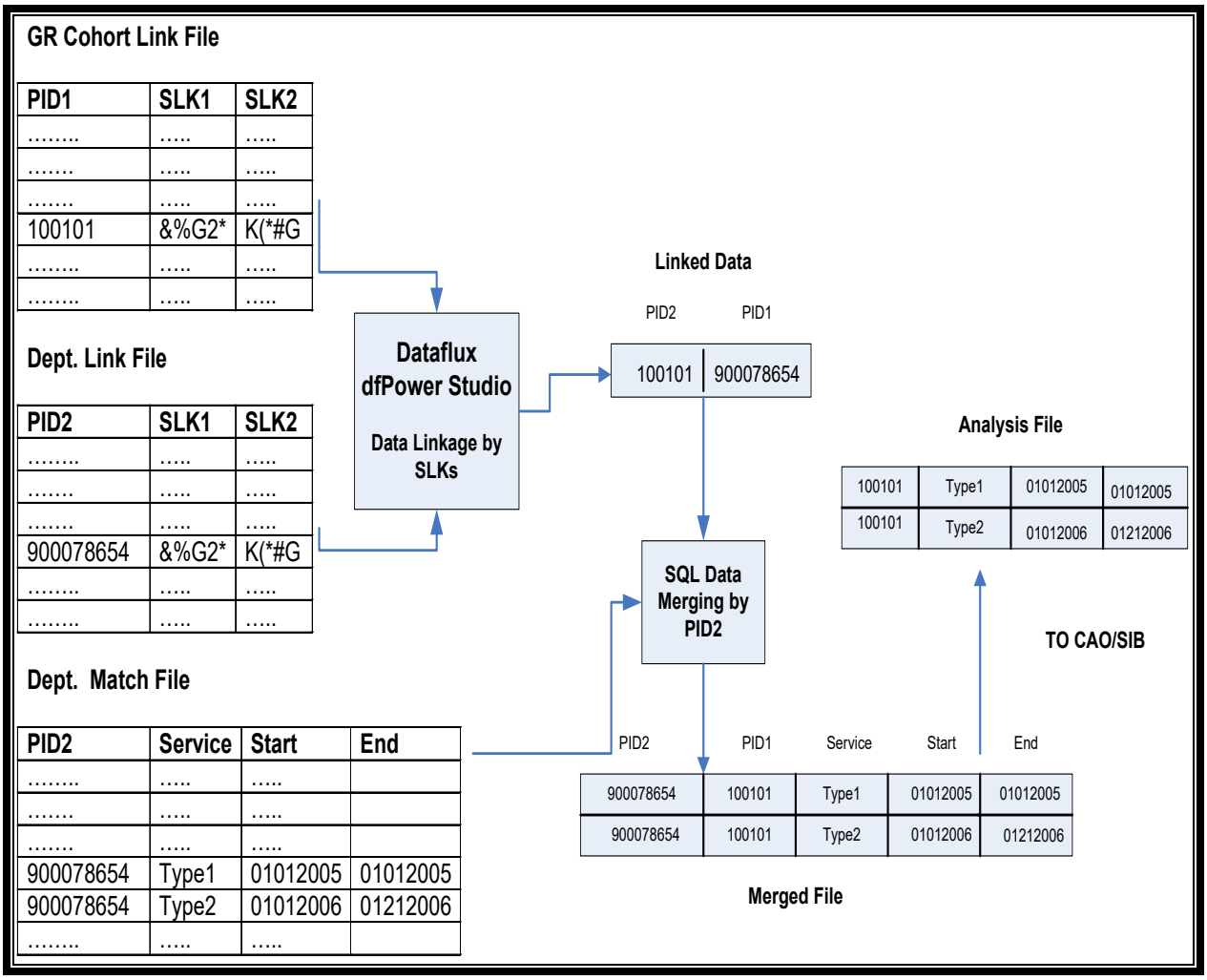
SIB will select a cohort of GR participants who entered the program during a given period of time, such as the first two quarters of 2003. After matching rules are determined and SLKs are generated, this data will be loaded to a Windows-XP workstation to be transported to the departmental sites for the purposes of linking cohort data to departmental data. A flow-chart of this step is presented in Figure 2.

The cohort data will only include a random project id and the SLKs to match the id to the departmental data files. The workstation will also include Dataflux dfPower Studio client software to match data at the departmental site. This software uses the same matching engine deployed in the SAS EDI server so that the same matching rules will be applied to both source (GR cohort) files and reference (departmental) files. Consequently, the same SLKs will be produced for multiple records of a given client served by different departments.

The matching of GR cohort files to the departmental link files will generate files that include two data fields—cohort id and the departmental id. Using SQL script, these two-field files will be merged by the departmental match files to append selected analysis fields, such as service dates and service types. The departmental ids will be dropped and the final analysis files will be delivered to SIB with the departmental service fields.

As shown in Figure 2, the record-linkage will be accomplished without using any identifiable information on the clients. SLKs are composed of special characters which are not recognizable. Strict data confidentiality requirements of county departments will therefore be met through this technology. Since the identity of individuals is not important and not retained in the project, departments do not see the identities of cross-agency clients. Linked data are not returned to source departments and the departments will not be able to acquire further information on linked participants for client management purposes. Moreover, data linkage will take place at departmental sites so that departments do not release their databases to outside agencies. Finally, a limited number of SIB researchers who are independent of participating agencies will be the only people with access to individual-level data on GR participants originally selected by SIB.

Figure 2. Data Linkage Process



3. Data Warehousing Deployment:

After receiving all analysis files from participating departments, SIB will load these files into a data warehouse platform that will be developed using the SAS technology. In addition to the data quality solution noted earlier, the SAS EDI Server also includes an ETL studio to develop a complete data warehousing apparatus. EDI Server provides an interactive data integration environment to connect, extract, transform and load multiple data sources. EDI server has a superior metadata environment and documentation properties, and it uses the powerful SAS language for data transformations and enrichment. ETL studio is seamlessly integrated with the data quality/matching solution and this functionality and design enables the rapid generation of data warehouses and data marts.

A flow-chart of the data warehouse development is provided in Figure 3. The development of a data warehousing platform will include the following steps:

- Business requirements analysis
- Data warehouse design
- Installation and configuration of the server
- Data integration and ETL development
- Testing and validation
- Migration to production

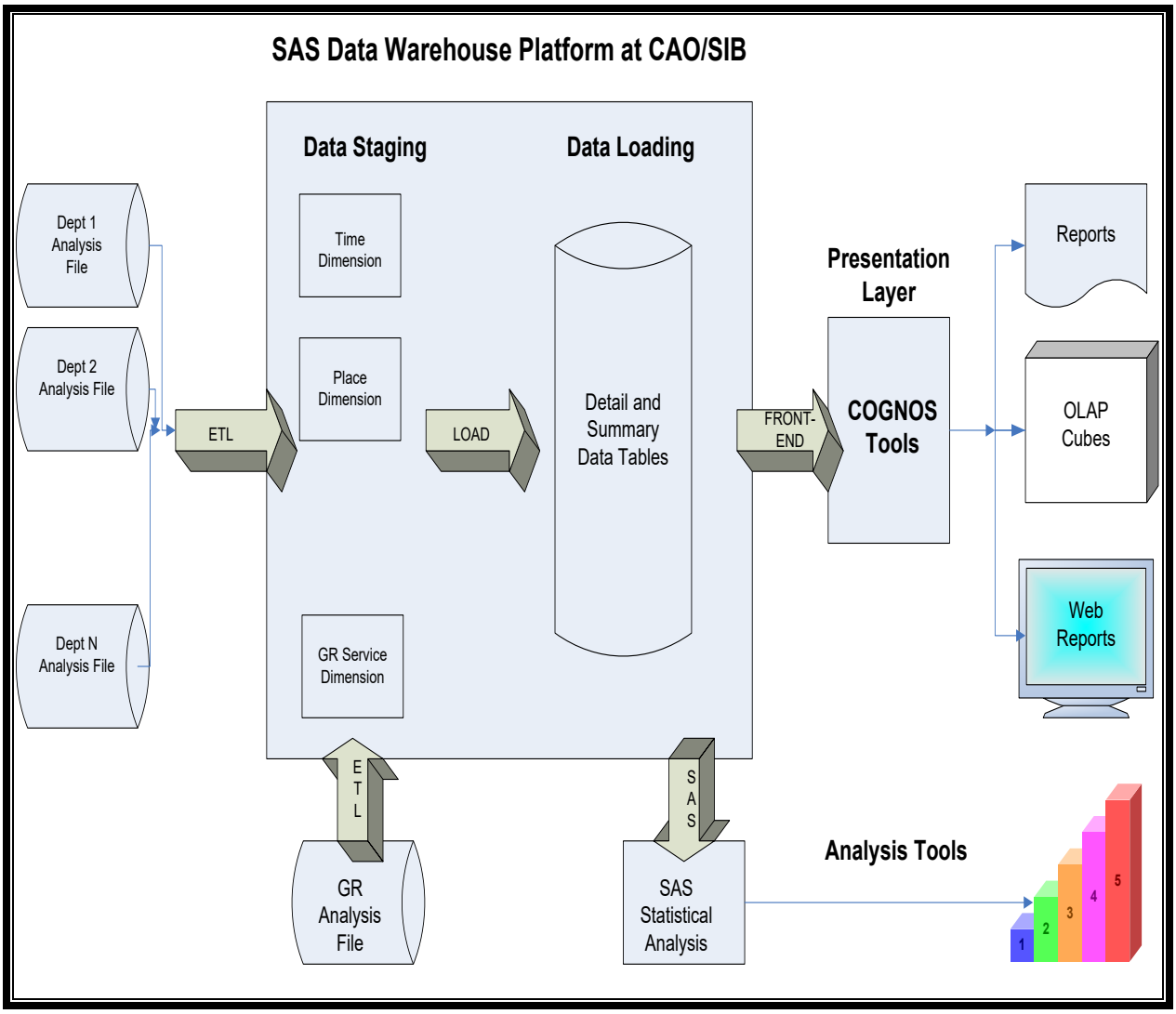
SIB will develop and maintain the data warehouse in accordance with the collaborative agreements and confidentiality and privacy regulations applicable to this data. The database will be housed at SIB. SIB staff will be responsible for designing and developing the infrastructure required to process and consolidate the data. SIB has been using SAS language for many years and has expertise in developing applications with SAS tools. SAS will also provide consulting services to deploy the data warehousing solution for this project.

The centralized data warehouse will be designed and implemented following the Business Intelligence best practices established by the County. Since the data warehouse will be implemented in SAS Data Warehouse technology, it will also provide the functionality required for complex statistical analysis, thereby making it possible to study the service patterns and trends of GR participants across departments.

4. Presentation of Data:

As shown in Figure 3, the presentation layer will be built on Cognos technology with a web-based user interface. Web-enabled Cognos tools will have an access to summary tables stored in the SAS data warehouse through an ODBC connection. Standard or custom-based HTML tables or charts will be easily generated by the end users assigned by participating departments. Web consulting services will be used to build this web interface to access the SAS data by using Cognos software.

Figure 3. Data Warehousing Development



PROJECT BENEFITS

The main goal of this project is to provide an integrated baseline database that links DPSS GR records with records from other departments also providing services to GR participants. The database will have immediate utility for policymakers in the development and planning of strategies to coordinate services provided to indigent adults.

The baseline database will produce historical data on GR participants to show the patterns and trends for various indicators, such as the number and characteristics of GR clients, types of services delivered, the length and timing of services, the costs of service delivery, patterns of service delivery before, during and after participating in the GR program, patterns of multiple service utilizations, transitions among programs, the geography of service delivery, etc. The Sheriff, for example, would be able to target

more expensive inmates and prevent unnecessary future costs through more effective case planning activities. Similarly, Health Services could target patient profiles in terms of medications used and thereby reduce unnecessary office visits and emergency room usage.

The information produced through this project will be used in the future to design new strategies and programs, such as service coordination pilot projects, by showing which segments of the GR population should be targeted, which agencies should be targeted, and which geographical regions within the County require the most and least amounts of attention. The Department of Probation, for instance, would be able to target its population for specialized supportive services designed to deal with issues related to mental health, substance abuse and domestic violence. In addition, the baseline database will provide valuable feedback to policymakers on trends in service utility, gaps in usage by particular types of clients, the effects of grant interruptions (due to sanctions or time limits) on service utilization and service costs, services utilized before entry into the GR program, etc For example, the database would make it possible to monitor service patterns for youths who emancipate from DCFS and receive GR funds. In connection with this, the database would improve DCFS' capacity to inform emancipated youth about Independent Living Program services, thereby lowering the number of youth applying for GR funds.

It should be emphasized that this project will develop a prototype record-linkage methodology and a data mart that can be transferred to other target populations served by both participating non-participating agencies. For example, the same application may be easily deployed for the analysis of service utilization patterns of DMH patients across other departments. The record linkage methodology will create an integrated/linked database that makes it possible to study service delivery from multiple agencies by resolving the confidentiality requirements that have been a significant barrier to the accomplishment of such a project up to this point. Finally, the project will make the database accessible to all participating agencies through a web-enabled interface, thereby allowing each agency to see the history of service utilization among GR participants.

PROJECT DELIVERABLES

The deliverables for this project are:

- A data linkage methodology that will integrate cross-departmental data while complying with County data confidentiality requirements.
- Analysis files linking GR participants to services delivered by other departments.
- Implementation of the data warehousing solution.
- A web-enabled interface to access the data warehouse using COGNOS tools.
- A final report to be presented to Board of Supervisors studying the service patterns, costs and characteristics of indigent adults served by county departments.
- Technical, security, and operational documentation.

FUNDING REQUESTED

Funds are needed to purchase, customize, and implement the required hardware and software for the indigent adult project, as listed below. This proposal requests funds only for consulting services to be used in the third and fourth components of the project, namely the development of the data warehouse and end-user access to the warehouse.

- SAS EDI Server—SIB already has a license for SAS ETL Studio software and funding is requested for an upgrade to the EDI server;
- Dataflux dfPower Studio Client;
- Windows XP dual-processor work station;
- Laptop computer;
- Consulting services to develop the data warehouse platform (the project's third component);
- Consulting services to web-enable the presentation layer so that COGNOS tools will access the data warehouse (the project's fourth component);
- Training for EDI server

Estimated Cost of System Hardware, Software and Consulting Services	
	1st Year
Software	
SAS EDI Server Upgrade	$94,000
Dataflux dfPower Studio Client License	$61,000
Hardware	
Windows-XP Work Station	$8,000
Windows XP Laptop	$2,000
Consulting Services	
SAS Consulting for Data Warehousing Development (component 3)	$175,000
Web Consulting for COGNOS to SAS Interface (component 4)	$50,000
Training	
SAS Training for EDI Server	$10,000
Total Costs	**$400,000**

MAINTENANCE

The application will be hosted on a UNIX server located at SIB. SIB staff will perform the installation and configuration required for the deployment of this application and will also provide the required training and ongoing support. The annual software maintenance of this application will be covered by SIB and participating departments.