The Enterprise Melting Pot

Everyone’s a SAS User – They Just Don’t Know It
Gail Davidson, KForce Government Solutions, Fairfax, Virginia
Guinevere Mills, U.S. Census Bureau, Suitland, Maryland

ABSTRACT
This paper describes the business intelligence effort for the Decennial Census through implementation of SAS Business Intelligence. The intended audience for this paper is organizations that are just beginning to implement common enterprise-wide computing that want insight into the decision process. The paper highlights what worked for us, how we made the decisions we did, and how you can benefit from our lessons learned. The environment consists of separate stovepipe applications; even our data warehouse was a stovepipe. All the usual problems associated with stovepipe systems development exist at the Census Bureau such as data duplication, conflicting data definitions, and turf wars, etc…. To dispel these problems, we started with the concept of an enterprise data warehouse and making that data available to only the right and authorized users. The stovepipe systems were tossed into the melting pot and evolved into an enterprise data warehouse. We got user buy-in through introducing SAS Business Intelligence reporting capabilities and SAS Add-in for Microsoft Office; getting the users excited about what they will now be able to accomplish on their own thereby surreptitiously creating SAS users. This paper talks about the processes used to get the user community buy-in and the experimentation involved in determining the optimal hardware and software configuration.

INTRODUCTION
"Though this be madness, yet there is method in't.", Shakespeare's Hamlet (II, ii, 206).

That quote from Hamlet is appropriate to our efforts to introduce enterprise business intelligence in our organization. When we began, most stakeholders, if not openly opposed to a common enterprise data warehouse, certainly were not rooting for our success and some thought we were quite mad. This paper examines the methodology we employed to get this project off the ground, from a technical as well as political point of view. A project of this size and scope cannot be address without addressing both.

VISION
Defining a single vision of our system has proven to be difficult. The processing and analysis culture at the Census Bureau is decentralized but moving very slowly toward centralization. And given human nature, some on our team tend toward a conservative approach while others are all for pushing the envelope. We have settled on an ambitious vision while deciding on a conservative methodology to get there. Our ultimate vision is for a common data repository and having all Decennial Census applications share one set of data. The respective area of interest would still develop their own applications however they would share the data with other applications – one version of data shared across all application systems. The SAS Business Intelligence platform provides the technological base for implementing this vision.

PROJECT SCOPE
WHERE WE CAME FROM
Our team began with the charter to develop and maintain the Cost & Progress System for the Decennial Census. This application tracks progress and costs for the facets of conducting the Decennial Census. The Cost & Progress is the definitive source of data for Decennial management and oversight. However, it is just one application, many other systems create the same types of informational reports albeit with a narrower focus. Hence, there exist many versions of cost and progress information. Our job is to reconcile the sources and present one version of the truth.

The previous Cost and Progress application was a thin-client, web-based system that delivered daily or weekly reports on cost, and where appropriate progress. The application was built using SAS/IntrNet software and related web-publishing technologies. The application’s front-end was menu driven interface that directed the user to operation specific reports. The cost and progress data was refreshed daily from various source systems and stored in a cumulative data repository. Then HTML reports were generated using SAS and ODS. These HTML reports were
then selected and viewed by our users through the SAS Intranet front-end.

WHERE WE WANT TO GO
Our basic charter for the 2008 Dress Rehearsal/2010 Cost and Progress system remains to primarily provide cost and progress information to Decennial management. Our scope has recently been expanded to include the 2000 Data Warehouse and 2010 Planning Database. With these additional systems and the expansion of our charter, the system we are developing with the SAS Business Intelligence architecture has been renamed from the Cost and Progress System to the broadened Decennial Business Intelligence and Analysis system. Integrating these new applications in a common repository will provide a holistic business view of the Decennial Enterprise. The SAS Business Intelligence platform gives us the tools to accomplish the task at hand and to empower the users such that they aren't dependent on our shop for their reporting needs.

Users access the applications and data via a single sign-on web interface, the SAS Information Delivery Portal. Within the Information Delivery Portal, a normalized dashboard home page has been developed. The Information Delivery Portal also enables the user to organize personal portal pages so they contain only the information needed, in the format that makes the most sense to the user. The applications provide pre-defined reports with graphs. In addition, the users are empowered to create their own reports and graphs by providing tools that make it easy to find, interact with, create and share reports based on a data warehouse described in business terms. The SAS Add-In for Microsoft Office enables users to harness the power of SAS analytics and to access SAS data sources from within Microsoft Word, Excel and PowerPoint. The Decennial Business Intelligence and Analysis system includes SAS maps, and geospatial mapping capability as one of the automated graphics features. The depth to which maps and reports will be drillable depends on the level of detail available in the data collected.

This system allows for central management of metadata resulting in transparent (but protected) data access to all mapped data across the enterprise, in effect, opening the atomic and pre-summarized data to in-depth analysis.

HOW ARE WE GETTING THERE
The decision to proceed with implementing a SAS Business Intelligence solution was based on a confluence of technology and process maturation. It was predicated by an overwhelming need for a common enterprise data warehouse and ad hoc reporting capability. The inception of common enterprise data warehouse required that we find a tool set that could manage the metadata and provide a business view of the data to end users. The SAS Business Intelligence platform was the singular solution that met all those requirements.

At the outset, there was very little experience and knowledge at the Census Bureau of SAS Business Intelligence. No one at the Bureau had ever set up an enterprise level SAS Business Intelligence system. Once the decision was made to proceed with the SAS Business Intelligence platform, the staff went through an intense period of training, taking all the BI courses offered by the SAS Institute. The decision to proceed with SAS Business Intelligence was made two years before the 2008 Census Dress Rehearsal and we had the time to take formal training. We implemented two prototype systems to learn and then sell the solution to our stakeholders. Moreover, we depended heavily on SAS Institute Center of Excellence and SAS Technical Support during the initial planning and system installation.

Some roadblocks to implementation are proving more stubborn than tackling the learning curve. These are turf wars and fear of micro-management. As other departments hire contractors to develop reporting capability there is conflict over duplication of effort and the level of detail appropriate for each reporting system resulting in multiple versions of the truth. This has caused several turf wars over who gets what data and the granularity of that data. The battles rage from upper management down to the individual staff members. The fear of micro-management is often part of the turf wars and deals with the level of data and timing, when should the data be available outside the staff that produces the data. We are addressing this problem by training the stakeholders in the security features offered with the SAS Business Intelligence platform. Additionally, we proceeded with a marketing effort aimed at the stakeholders designed to demonstrate the SAS Business Intelligence capabilities alleviating their fears and getting them excited about what they'll be able to do on their own. This marketing effort is ongoing.

Other roadblocks are standard project management issues. As the marketing effort proceeds, more applications were jumping on the SAS Business Intelligence bandwagon. We are a small staff and resources are now stretched to the breaking point. We are at a critical juncture where results are expected but expansion is required to reach the vision of the project.

HOW WE GOT STAKEHOLDER BUY-IN
EDUCATION
The first step to getting buy-in was educating ourselves on SAS Business Intelligence capabilities and getting management to see that this was the solution to meeting the organizational needs. Our education consisted of...
reading white papers, talking to other divisions and government agencies as well as working with SAS Institute. Two other divisions at the Census Bureau, Housing and Household Economic Statistics Division (HHES) and System Support Division (SSD) had SAS Business Intelligence research projects. Their help and advice in developing our prototype systems were invaluable.

PAPERS AND RESEARCH
In order to solidify our understanding and refine our requirements for the SAS Business Intelligence, the DBIA staff wrote documents and white papers on SAS Business Intelligence capabilities, how the different tiers/products worked together, how it could be implemented at the Bureau, and requirements for the two prototypes. We had several meetings to review our documents to foster understanding of what we were trying to do. We then met with our System Integration staff; they are the ones that are responsible for installing the hardware and software, and the System Support Division; the de facto SAS Business Intelligence advisors at the Bureau. We communicated what we thought were the requirements, which were then refined with their input. These two external divisions became critical members of the SAS Business Intelligence implementation team. The learning curve on their part and ours was steep.

The first prototype implemented was and abject failure, it never worked correctly. We did, however, learn from the experience and the second prototype was quite successful.

DEMOS - MARKETING
Once we had our second prototype set up, we started demonstrating the system to show how to surface data in the Information Delivery Portal, create ad-hoc reports/graphs with Web Report Studio, view reports/graphs already created with Stored Processes and Web Report Studio, working with Microsoft Excel, Word, as well as the mapping capabilities of SAS Business Intelligence and ArcGIS server integration. At each demo we got more stakeholder buy-in and requests for more demos to other staffs/divisions and upper management. Also because of our demos we received two new projects, the 2000 Data Warehouse, and the 2010 Planning Database. In summary, the marketing effort for the SAS Business Intelligence project was very successful. Once the stakeholders saw the demo, they were on board. They had never had a system that was so empowering and easy to use.

THE MELTING POT - CONCEPTUAL SYSTEM VIEW
The primary reason for the business intelligence is to empower users to access and analyze their data, delve into, explore and turn the data into intelligence. The enterprise data warehouse (EDW) is the foundation of that effort. The EDW is evolving and melting down the stovepipe applications. Our EDW is based on the government information factory (GIF) architecture, first posited by Bill Inmon, and based on the corporate information factory. We began this effort by profiling the stovepipe data and identifying duplication. This effort is ongoing. The GIF is evolving by breaking down the stovepipes one system at a time, folding the data into the common decennial enterprise data model.

![Figure 1](image_url)

The vast majority of the systems illustrated in figure 1 are source systems into the 2010 C&P data warehouse. Not all the data in those systems will end up in the EDW in the immediate future. For now, only select data is extracted from them and conceptually placed into the 2010 C&P DW. The 2010 C&P DW does not exist and never will exist as a
separate physical system. Its development schedule serendipitously lends it to be the first system melted down into the EDW. Over the next few months the other stovepipe data warehouses will be incorporated into the EDW.

The following list describes the first systems that will be melted down into the EDW:

2000-2006 C&P Data Warehouses – Each of these data warehouse/repositories are structured similarly and contain the same types of data, namely, the cost and progress information for the last Census and the Census tests through 2006.

2000 Census Data Warehouse – This data warehouse contains all the detail data from the 2000 Census such as applicant, payroll, and field operational details.

2010 Planning Database – This database contains geography and response information from the 2000 Census that is used in planning operations and resource allocation for the 2010 Census.

ENTERPRISE DATA WAREHOUSE

One of the biggest hurdles encountered in this effort was getting buy in from the different organizational entities at the Bureau. The EDW was the area where most of the problems associated with turf wars were encountered. The IT culture at the Bureau is decentralized. We still haven’t reached the point with the Bureau’s culture where centralized data management is trusted. A lot of what we do is a duplication of effort at this point. However, when the 2010 Census is over, our charter includes maintenance of the data. At that point, the plan is to fold the operational data into the EDW. That is where the SAS Business Intelligence comes in. It will manage the access and the metadata on the operational data. Then it will be available for the users access it via the Business Intelligence client tools.

To gain user support they need to know they still own the data and they aren’t giving up control of it. Each piece of data in the GIF is assigned an owner (data steward). This is one of the focuses in the marketing effort. The steward decides what is done with the data and who has access. We have worked out procedures for communication between the administrators (us, the DBIA staff), the user community, and the data stewards. This process so far has proved to meet the needs of the users and the stewards. It is responsive thus far, but has not been stressed. We already know there is room for improvement.

HOW THE STOVEPIPES FORMED ONE VIEW – GETTING THE DATA IN

We had to revamp the way we thought about our systems. In the past we concentrated on report delivery and the data was only there to support the reports. The SAS Business Intelligence relieved us of being a report centric and forced us to concentrate on the data. The change in the way we thought of our systems predicated a change in the organization of how the staff is structured to develop and deliver the systems. We split our staff into two different functional areas. One area concentrates on getting the data in. Quality and maximum reuse are the guiding mantras for this part of the staff. The other part of the staff concentrates on gathering user requirements, acting as a bridge to ensure a business view of the data exists, and developing the user interface.

IMPLEMENTATION

PROTOTYPE 1

The first prototype was a learning exercise for us. Operationally, it was a failure. But as a learning project, it could be considered a huge success. This prototype project’s goal was to get a SAS Business Intelligence system up quickly, learn from the installation, gather performance information, and learn the client tools. This prototype was only operational for two months.

All parts of the BI architecture were implemented on one Sun V890 system. The V890 was configured with:

- 4 Sparc9 processors at 1350MHz
- 8 GB memory
- Disks
  - 146GB mirrored (2x146GB)
  - 410GB raid5 (4*146GB)
- Solaris 10
- Tomcat
- Xythos WebDav
- Oracle 10g with UTF8 character set
- SAS version 9.13 service pack 3

Figure 2 illustrates the planned system context. The single V890 server held the logical data server, SAS Foundation and Mid-tier servers. The Xythos WebDav store was a schema in the C&P database.

This prototype was not installed correctly. Xythos never worked. We never got to the point where ArcGIS server was
installed. But it was minimally operational. We were able to create metadata and use the Foundation servers. We used this prototype to learn the client tools.

**Figure 2: Prototype 1**

**PROTOTYPE 2**

This prototype was much more successful. It was configured on four separate Sun V890 servers. The hardware configuration for each V890 was the same as the V890 used in prototype 1; the difference with this prototype being, the introduction of WebSphere and spreading the processing over multiple nodes. It took three months to get this prototype fully operational. The focus of this prototype was optimal performance. Most of the problems encountered were from incorrect installations. As a general rule, the SAS products installed flawlessly. It was the installation of the third party products and integrating them with the SAS Business Intelligence servers and applications that was problematic. This was this prototype that we used for the various demos.

One of the most requested features by our users was the inclusion of maps for spatial data analysis. With the inclusion of the ArcGIS server integration with the SAS Business Intelligence we were able to fulfill their request. However, the ArcGIS server was one of the most difficult pieces to fit into the configuration. The Census Bureau was the first customer site in the world to configure ArcGIS server on the Solaris platform and integrate it into the SAS Business Intelligence platform.

**Figure 3: Prototype 2**

Due to problems experienced with the testing software suite interacting with the SAS Business Intelligence client applications the results of the stress and performance testing were limited. We only have anecdotal performance observations. Testing on prototype 2 has resulted in inconsistent results; however, in one test the number of virtual users was pushed up to 240. In each test the mid-tier server begins to hang because of errors generated by the testing software. The number of virtual users varies based on the testing scripts and how many errors they generate.

**PROTOTYPE 2 LESSON LEARNED**

- The roots of the problems with installation were due to not letting the installation scripts do the work for us. Let the installation programs do the work. Use the installation created scripts.
- Get Xythos working with Tomcat first, and then bring WebSphere into the mix.
- Make sure you’ve got the latest drivers and hot fixes.
- WebSphere uses a self contained JRE, not the JRE required by SAS, 1.4.2. The SAS installation navigator will install and configure the correct JRE.
- ArcGIS server installation and set up was not well documented. It requires an additional identity server on UNIX still uses Microsoft directory notation.
- With metadata administration there were a lot of little issues. Some were part of our learning curve but some were just little gotchas. For example, we set up and Oracle library, by chance used the same libref that was used in a different pre-assigned SAS base library. Everything looked as if it was set up correctly to access the Oracle tables. However, the import wizard used the libref of the pre-assigned library and only listed those tables, not the Oracle tables.
- Naming standards enforcement is imperative. We spent a lot of time and effort developing naming standards to ensure every name in the metadata was unique and identified what it was.
- Authorizations were difficult to comprehend, although it’s very well documented in the latest BI admin guide, 5th edition. However, we were six months into our prototype before that was published.
- The documentation was a little rough when we began building the prototypes. It’s much better now. Check sas.com often for documentation.
- Define data sources in the custom repository. Define targets in Foundation repository because information maps can only be surfaced to Web Report Studio from the Foundation repository and Foundation maps can only be made from Foundation defined data.

PRODUCTION
We are currently in the process of building our production environment that will be used during the 2008 Census Dress Rehearsal. This process is going much smoother with the lessons learned from the first 2 prototypes. It is configured very similarly to the second prototype with the exception of adding clustering on the metadata server and mid-tier server to provide high availability. Additionally, the production environment is attached to SAN storage facilitating the failover. The metadata server and mid-tier servers are installed on a 2-node Veritas cluster. Each node is the primary host for one, and is configured to failover and host the other server in case of system failure.

Figure 4: Production

EVERYONE’S A SAS USER – THEY JUST DON’T KNOW IT
This section of the paper contains a short description of the SAS Business Intelligence client tools and how we are using them to implement our application.

INFORMATION DELIVERY PORTAL
We use the Portal as a single point of entry for our users into the SAS Business Intelligence and EDW. There’s lots of native functionality in the Portal: publishing, data exploration…. What’s not native can easily be linked to. We provide links in the Portal to the Cost and Progress Dashboard which is a graphical overview of the 2008 Dress Rehearsal Operations created via stored procedures, predefined Web Report Studio reports, and the 2008 Dress Rehearsal schedule which is in Primavera Project Planner.
WEB REPORT STUDIO
Empowers the user to explore data and create their own reports and graphs. Information maps are set up to provide the business view of the data to the users. The information maps also provide an additional layer of security. In addition to the layer of security on the data itself, the metadata security on the data, access to information maps is another layer that can be used through the web applications. However, the metadata to the data must be set up correctly because there are other BI tools that allow access to the data directly (Enterprise Guide, SAS Add-in for Microsoft Office).

STORED PROCEDURES
Web Report Studio is a great tool for building reports but they are easily modified. Sometimes a report has to be created that must stand as is, for reasons such as regulatory compliance for example. Or if there is a special report that Web Report Studio doesn’t have the capability to create such as an annotated graph. Stored processes are the way static content can be delivered to the BI client applications.

SAS ADD-IN FOR MICROSOFT OFFICE
Opens SAS to non-SAS users. Many of our users are more comfortable with Microsoft Office applications. For many analysts Microsoft Excel is their tool of choice. The SAS Add-in for Microsoft Office opens the SAS functionality to these users.

CONCLUSION
The concept of an enterprise data warehouse and the tools used to implement and empower the users need to be introduced slowly. This is not something that can be forced on an organization, without stakeholder buy-in the project is dead in the water. The way we were able to get stakeholder buy-in was by introducing the concepts a little at a time, and each time emphasizing how they would be empowered and the data protected. The user/stakeholder support is critical to the evolution of the project. The marketing, i.e. demos, always need to be ready and targeted to your specific audience. Some of our stakeholders got excited about the new power they would have at their fingertips. Others got excited when we showed them the security features provided by the metadata.

This project is still a work in progress and will be for the next year at a minimum. We are at the stage where we are about to release the first Dress Rehearsal operations to production and we anticipate a lot of fine-tuning in requirements once the users starts getting into the system. The users will begin the mind shift from looking at the data with an attitude of single purpose toward looking at the data as a multi-purpose corporate asset. This system will be the one-stop ad hoc querying application – turning everyone into a SAS user.

ACKNOWLEDGMENTS
This project could not have undertaken without the tireless efforts and support from the project team: Bert Hotton, David Saunders, Bob Anderson, Annette Davis, and David Kiasi-Barnes, Van Nguyen, Debbie Gatuso, Tim Braam. A special thanks to SAS Technical Support and the SAS Center of Excellence. Thank-you all for your support and assistance.

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the author at:

Gail Davidson
Kforce Government Solutions
301-763-3444
gdavidson@kforcegov.com
gail.s.davidson@census.gov

Guinevere Mills
U.S. Census Bureau
301-763-5421
Guinevere.z.mills@census.gov

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
Other brand and product names are trademarks of their respective companies.