

Paper 478-2013

SAS Enterprise Business Intelligence Deployment Projects in the Federal Sector

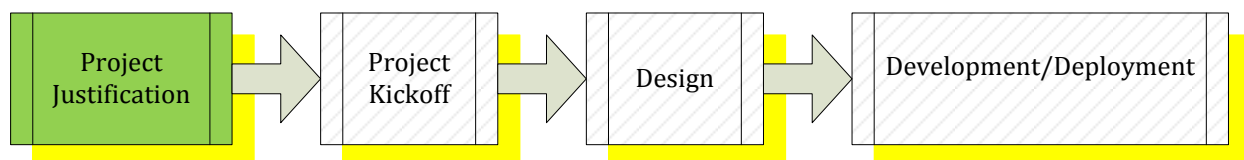
Jennifer Parks, CSC, Falls Church, VA

ABSTRACT

Systems engineering life cycles (SELC) in the federal sector embody a high level of complexity due to legislative mandates, agency policies, and contract specifications layered over industry best practices all of which must be taken into consideration when designing and deploying a system release. Additional complexity stems from the unique nature of ad-hoc predictive analytic systems that are at odds with traditional, unidirectional federal production software deployments to which many federal sector project managers have grown accustomed. This paper offers a high-level roadmap for successful SAS EBI design and deployment projects within the federal sector. It's addressed primarily to project managers and SAS administrators engaged in the SELC process for a SAS EBI system release.

INTRODUCTION

Interest in predictive analytics among those in the federal sector has surged. Federal agencies have jumped aboard the predictive analytics train with increasing frequency over the past half-decade and the launch of the Big Data initiative by the Obama administration has pushed all agencies to uncover business value in the petabytes (one quadrillion bytes!) of data flowing into their offices. Predictive analytic capabilities require the deployment of predictive analytic systems or what we call EBI deployments. Mandates pushing down from the administration, data pouring in from all directions and flashy sales presentations by EBI vendors spur agencies to initiate costly and complex EBI deployments that must be carried through the traditional system engineering life cycle phases. Most prominent agencies are limited to waterfall or linear development cycles which can further hinder the flexibility that would benefit an EBI deployment (EBI deployment project managers should inquire with the agency SELC process management teams as to whether AGILE or iterative development life cycles are an option). For all agencies, these SELC phases, loosely defined, include project justification, project kickoff, design, and development/deployment. Each phase is marked by points of potential failure for our SAS EBI deployment and our goal as the deployment team is to mitigate these risks with a systematic approach to project management tailored to federal sector SAS EBI deployments.

PROJECT JUSTIFICATION**Figure 1: SELC Phase Diagram (Project Justification)**

Every SAS EBI deployment must begin with justification for the deployment before sponsors who have the authority to authorize the project.

DETERMINE BUSINESS GOALS TO BE MET BY EBI DEPLOYMENT

We initiate project justification by posing and answering a few preliminary questions. How do EBI and predictive modeling fit into the goals of the agency or department? What exactly are we going to do with our data? It can be tempting for agency or division leadership to initiate a deployment project based on an exciting demo by a sales team covering the technical possibilities of EBI offerings but product features are never enough to truly justify taxpayer expenditures and ensure our project sponsorship remains secure.

The granddaddy rule of all rules for a successful SAS EBI deployment: project benefits and business value must be clear and demonstrable; there must be a straightforward application of the EBI deployment against agency or division objectives.

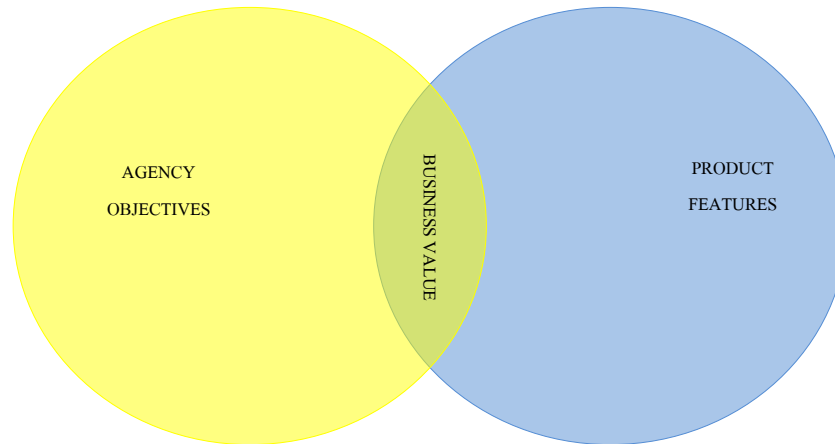


Figure 2: Defining Business Value

PRODUCT JUSTIFICATION

Once we have clearly established the business value of an EBI deployment to serve an agency or division, our next step is to select and justify the appropriate SAS products, subject to our budgetary constraints. Specific EBI client tools to be deployed in the server-client configuration will be flushed out during the user requirements phase, but the server platform needs to be established at this point. EBI modeling and predictive analysis across a variety of subject matter areas and for multiple purposes is typically accomplished with the core [SAS EBI Platform](#). Advanced analytics concentrated in core subject matter areas or toward a single purpose may be better served with a SAS Solutions product. These specialized solutions are more costly in their implementation due to licensing structures but come prepped with useful models weighted toward the specific task. For example, SAS Fraud Framework is targeted toward fraud detection. Agencies or divisions selecting specialized SAS Solutions must also budget additional full time employee (FTE) resource dollars as consultants and contractors trained in these packages come at a premium.

In addition to the core analytics platform, EBI deployment teams need to review data management applications for data set storage and extraction, transformation, and loading (ETL). SAS EBI Platform is often deployed in conjunction with the [SAS Enterprise Data Integration \(EDI\) Platform](#). Special attention to security standards (derived from the agency's Security Policy and relevant NIST specifications) must be taken during product justification as such standards may suggest encryption requirements.

For SAS EBI deployment teams planning to outsource the installation and configuration of SAS EBI to SAS Institute, note that SAS does not typically supply SELC deliverables beyond a SAS Plan file (summarizes the distribution of SAS licensed servers across machines in your environment) and the SAS install summary (useful as input for a version description document). For such arrangements, we will still need a knowledgeable SAS subject matter expert (SME) on hand as a key participant throughout SELC process.

PROJECT KICKOFF: PLANNING AND REQUIREMENTS

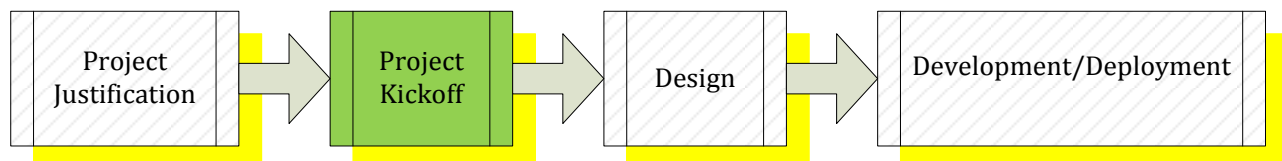


Figure 3: SELC Phase Diagram (Project Kickoff)

Consider the preceding steps we've completed as akin to the pre-boarding portion of an airline flight; most of the hard work is still ahead of us and we kick it all off with the (appropriately dubbed) project kickoff meeting.

SCHEDULING

In preparation for the kickoff meeting with project stakeholders, the SAS EBI deployment project schedule draft should be constructed (at some agencies the schedule will be produced after the kickoff). SAS administrators provide strategic input into the schedule (and accompanying project plan if one has been required) to assist the project manager in properly estimating installation, configuration and system testing timelines. If the SAS EBI release is to incorporate managed data structures and reports, SAS data modelers and/or programmers provide guidance on development timing for ETL tasks and the initial report population.

STRATEGIC PLANNING

In addition to the SAS EBI deployment project schedule there are a number of additional planning documents that are required as part of an agency's SELC process. Typically these documents must be delivered and reviewed by a gate review team before the deployment project is permitted to pass to the next phase in the project. Examples of planning documents include project tailoring plans, risk management plans and disaster recovery plans. When completing the remainder of the planning deliverables, we must adhere to agency lifecycle process requirements, data center and security/privacy requirements such as those surrounding personally identifiable information (PII) and ensure our stakeholders are onboard with the deployment plans.

FUNCTIONAL AND USER REQUIREMENTS FOR SAS EBI

Gathering requirements is an important step in successful SAS EBI deployment projects and should not be overlooked. SAS administrators are typically responsible for working with the requirements team to map user and functional requirements to specific SAS product offerings. Feature by feature we need to comparatively assure that the SAS Platform we selected is suited to our requirements. Further, we must now select the specific EBI client tools necessary to meet user requirements (while continuing to honor budget constraints).

We must remember not only to address user process requirements but also user access requirements. At many agencies, the division responsible for firewall maintenance and security is wholly separate from and not subject to the authority of the division sponsoring the SAS EBI deployment project. Unless we document user access requirements and adequately plan access strategy accordingly we may deploy a SAS EBI release that is inaccessible to our end users because of firewall or desktop security restrictions. Will our users be connecting to the SAS EBI environment from their desktops? What about remote access via VPN (it is essential to benchmark VPN speed in such cases because it will directly impact EBI client tool usability, particularly for clients that transfer large amounts of data to and from the SAS EBI servers)? Will the release feature a web interface and if so will it be on the agency intranet or available on the public facing internet?

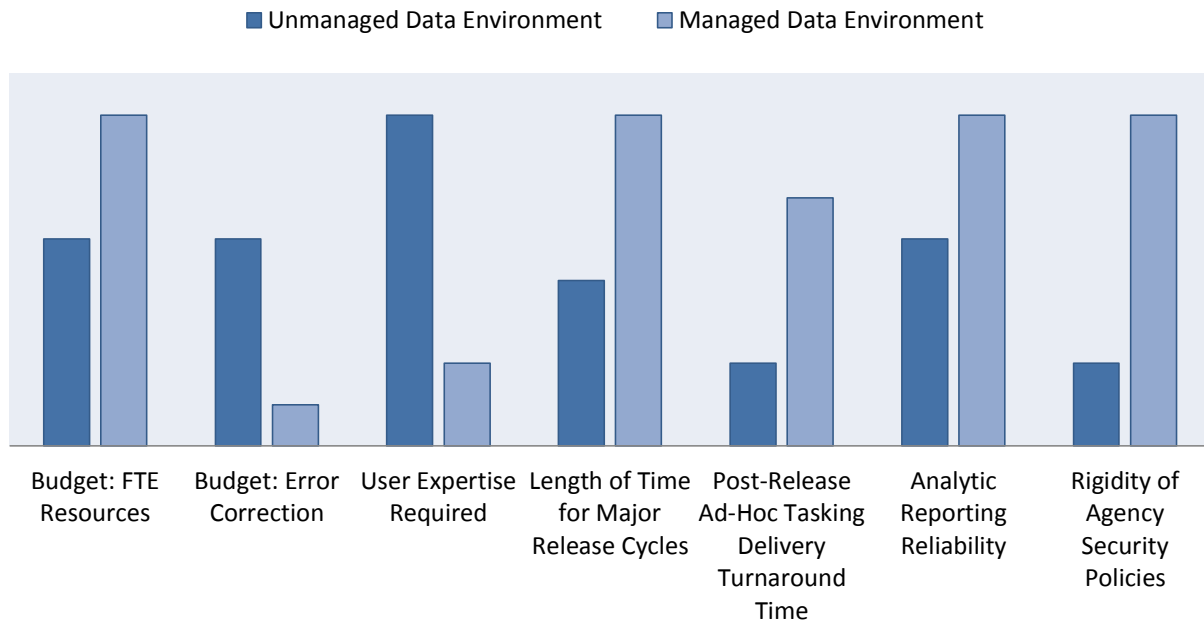
DATA MANAGEMENT DECISIONS

Before detailed system design can begin, the SAS EBI deployment team must make the crucial data management decision: managed or unmanaged data environment. There are clear but contrasting benefits for each approach and our final decision will be informed by:

- Project budget
 - Managed data environments require additional FTE resources (dedicated to data management tasks) on the project team.
 - Unmanaged data environments may result in duplicative, ineffective, or inaccurate data management task work by end users which not only represents a resource cost in lost time for the agency or division but may prompt additional costs to correct or revise errors in reporting.
- User expertise level
 - Effective data management (including cleansing and other ETL tasks) requires a minimum level of technical competency. Without the appropriate level of expertise in data management using SAS and

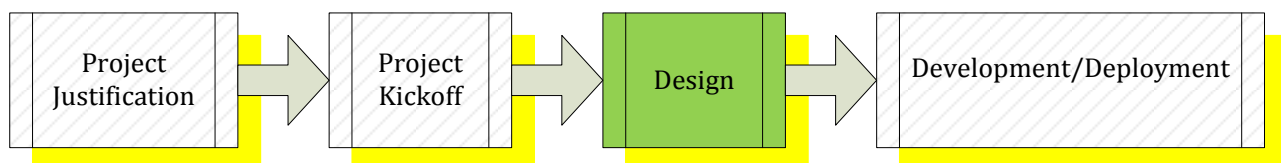
third party tools such as schedulers, as well as an understanding of the subject matter area the data pertains to, end users cannot work data management tasks to completion.

- Time constraints
 - SAS EBI deployment projects under tight release delivery timelines cannot adequately incorporate data design and deployment into the initial release cycle.
 - Post release, ad-hoc tasks submitted directly by end users in an unmanaged data environment typically allow for substantially faster delivery than those run through formal, managed development processes.
- Commitment to centralized (enterprise) data management and reporting reliability across the agency or division
 - Managed data environments provide a greater level of confidence in the analytic and reporting reliability of a SAS EBI deployment with centralized inventory and documentation of raw data sources, data structures within the SAS EBI environment, SAS code and formats, and reporting output. This centralized inventory reduces or eliminates conflicting business rules, data structures, and reports.
- Agency security policies
 - Agency security policies will have direct impact on data management decisions.
 - Agencies employing rigid managed data environment policies do not permit SAS administrators or programmers on the project team or end users to conduct bi-directional interactions with production systems post release. In such agencies EBI tasks such as ad-hoc reporting, defining libraries, metadata/file system permission adjustments and ETL modifications outside of the formal release cycle are out of bounds.
 - Agencies employing flexible managed data environment policies may allow the project team (but not end users) to conduct bi-directional interactions with production systems post release.
 - Agencies permitting unmanaged data environments in the production domain provide for the most freedom in EBI tasking.

Data Management Decision Factors**Figure 4: Data Management Decision Factors**

Where security policies allow, project teams may also opt to deploy a hybridized data management environment where ETL and data structures (including information maps) are centrally deployed and managed by the project team but SAS analytic tasks and reporting are unmanaged.

Regardless of the SAS EBI deployment project team's decision to proceed with a managed or unmanaged data structure, either approach requires SAS SMEs on the project team post deployment. For managed data environments these SAS SMEs will be responsible for the ongoing centralized administration and data processing. In unmanaged data environments SAS SMEs will be responsible for user assistance and the minimum administrative tasks (such as managing permissions and defining libraries for external data sources) inherent in SAS EBI analytics and reporting.

DESIGN**Figure 5: SELC Phase Diagram (Design)**

There are a number of key design phase tasks and deliverables that need to be completed as part of our SAS EBI deployment, regardless of whether we using an in-house or SAS Institute contracted installation and configuration team.

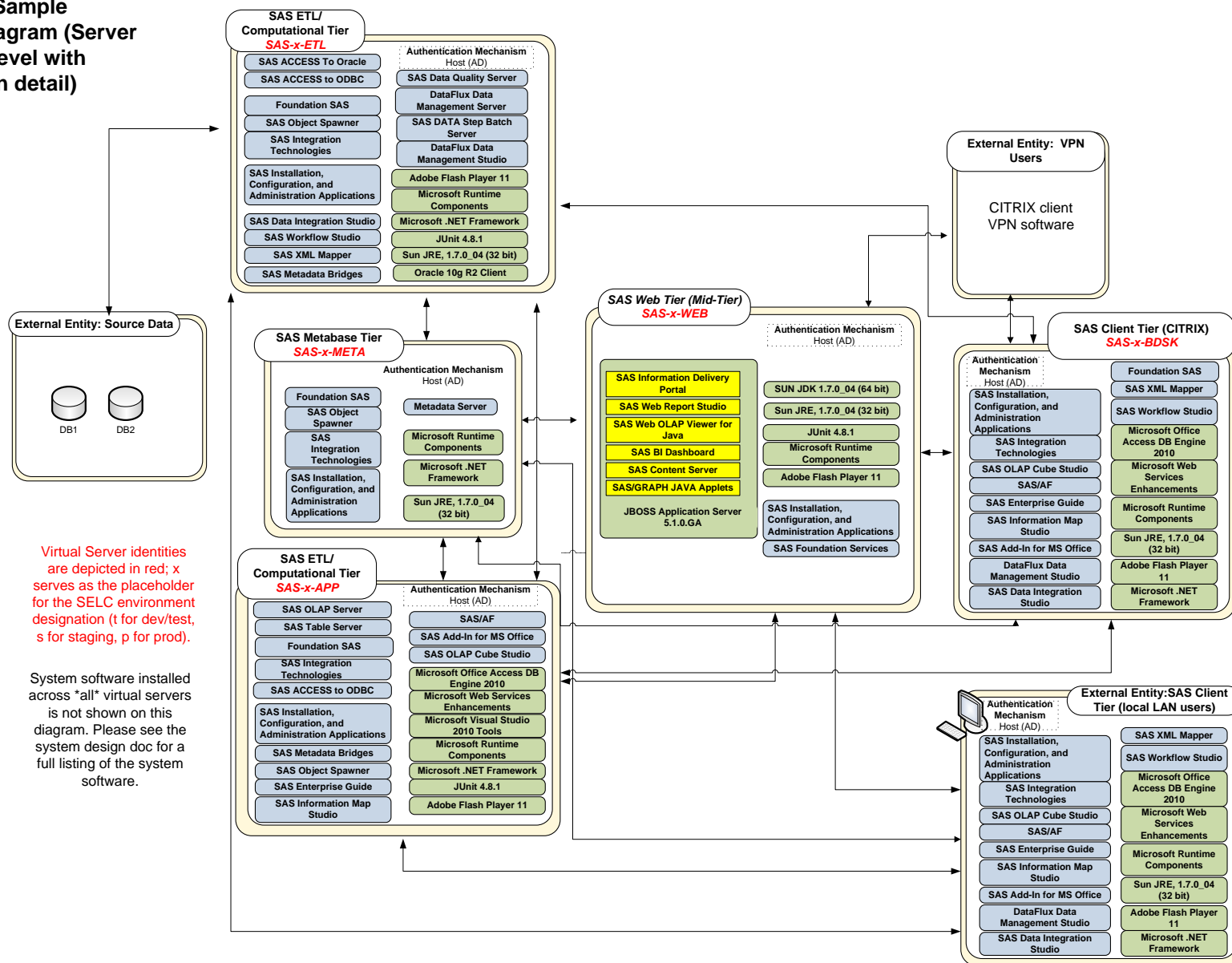
APPLICATION DESIGN (LOGICAL AND SYSTEM)

The logical and system design including application distribution across servers should be crafted by the SAS administrator based on planned patterns of use and constrained by the project budget (for SAS EBI, licensing is usually per CPU core). The SAS enterprise sizing center is a useful resource for determining hardware specifications

to suite usage requirements. Managed data environments will allow for straightforward sizing (base usage + anticipated growth over time) whereas unmanaged data environments force us to speculate with a lesser degree of accuracy when answering many of the questions that ultimately determine sizing recommendations. There are also numerous white papers available detailing lessons learned and best practices for the host level installation and configuration. For both unmanaged and managed data environments, the application design will cement boundaries on system capabilities and is used directly in crafting the SAS Plan file which will in turn be used by the SAS Deployment Wizard during installation and configuration.

Figure 6 on the next page shows a sample design diagram with application detail.

Figure 6: Sample Design Diagram (Server interface level with application detail)



DATA FLOW AND STRUCTURES DESIGN

Beyond the logical and system design documents, plans for the data flow through the application and the data structures maintained within the application must also be documented.

Our data flow documents should encompass both detailed charting of the flow of information from source systems through the various SAS logical servers and applications out to users as well as a high level graphic depiction of system to system data exchange. The latter will serve as input to the System Security Plan (details the system architecture as relevant to security policies and documents how the system complies with such policies) for the SAS EBI release. In addition, for deployments in the federal sector, data flow design tasks include establishing policy agreements governing the connection between our planned SAS EBI system and those systems that we wish to use as remote library sources (such as warehouses or other databases). These are typically referred to as inter-agency (between agencies) or intra-agency (between divisions within our agency) connectivity agreements.

For unmanaged data environments, data structure documentation is minimal. We must set policies on how we will determine when data is stale and subject to archival or deletion. We must also set policies governing user data storage allocation (data directory structures and quotas). Finally, we must remember to craft policies in deference to our security (particularly regarding PII or classified data) and data center requirements.

For managed data environments, we need to incorporate all of the policy planning required of unmanaged data environments described above. Additionally, we must establish business rule policies. These policies should set a precedent for firm business rules that are strictly adhered to and are not subject to discard without due process. There is always a temptation to use the same variable for two different analytic tasks with slight differences in the business rule definition for each. Resist the temptation! Business rule definitions should never conflict nor duplicate one another. Once we've crafted our business rule policies, we can move onto the work of documenting the actual business rules and designs for ETL (including jobs and flows), data marts, lookup tables, SAS formats, OLAP cubes and information maps. Depending on the individual agency or division, reporting structures (code and surface reports) may be defined under data structures or as a separate SELC design task. Finally, for managed data environments we will also need to define a refresh schedule for ETL and reporting.

The data flow and structures design documents will provide the basis of estimate for ETL, data modeling, statistical, and programming FTEs that will be needed to deploy and maintain the SAS EBI system.

USER GROUP STRUCTURE AND PERMISSIONS SCHEME

As the data flow and data structures designs are in progress, the SAS administrator serving the SAS EBI deployment project can design the SAS user group structure and permissions scheme. Role based permission groups are highly recommended. SAS administrators should make a detailed study of the permissions required for every functional task within SAS applications or performed by SAS server components. Once a complete understanding of required permissions is achieved, we document the permissions design that needs to be applied against the metadata, the file system, and the WebDAV repository (e.g. SAS Content Server). Out of the box, SAS provides a number of built in application role definitions with template permissions as well as a default access control template (ACT) for use in implementing role based security. The SAS Installation Guide and SAS System Administration Guide published by SAS Institute provide exhaustive details on permissions required by task and per server component.

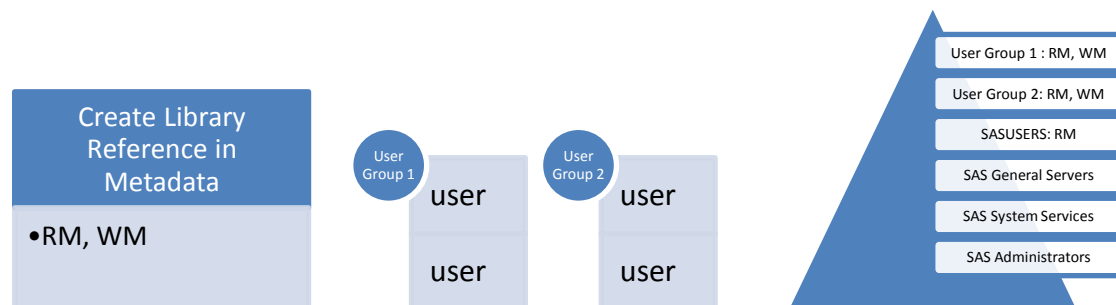


Figure 7: Permissions Scheme

VERIFY APPLICATION DESIGN AGAINST SYSTEM SECURITY PLAN (SSP)

It is absolutely essential to verify that the system design, logical design, data flow and structures design, and user

group structure and permissions scheme comply with the System Security Plan. The SAS EBI deployment team must work closely with the security team authoring the SSP to ensure that the system is not cited during security certification and assessment for non-compliance. It does not matter how magnificent our system is or how rapidly it can deliver analysis with business value if it is never cleared by security to roll into production.

TEST PLANS

We've nearly cleared the design phase. Congratulations! Most of the difficult work of this phase is behind us. It is important now to decide on how we will objectively measure our deployment success against the functional and user requirements. For all EBI systems we must determine what tests should be run to evaluate if the SAS EBI system, including all of the SAS server components, is running efficiently as expected and capable of performing the tasks required of it by our end users. For managed data systems we will need to additionally design tests to verify the logic of our managed code, the estimates of our ETL processing timelines, and the reliability and accuracy of the data housed in our data structures.

DESIGN TRAINING

At minimum brief desktop reference guides specific to our deployment should be planned for release, but other options include a Frequently Asked Questions (FAQ) document and detailed interactive or instructor led trainings to be created by the agency or division. Optionally, if the project budget allows, the agency or division may plan to outsource additional training to SAS Institute. SAS offers web based training as well as live, instructor led training at various locations across the globe.



In planning the training materials we must always remember that we are catering to different types of users including business analysts, statisticians, and advanced SAS programmers. As mentioned previously, users in unmanaged data environments will need more hands on assistance post deployment.

LICENSE PROCUREMENT

The last crucial task in the design phase is to procure the SAS software licenses. Each agency has modest differences with regard to procurement processes that can stretch the procurement from days to months. If the SAS EBI deployment project team has maintained direct and frequent communication with the SAS sales account team assigned by SAS Institute to the agency or division, a temporary thirty or sixty day license can usually be obtained to facilitate the installation and configuration for testing while awaiting final license procurement. This is particularly useful for SAS EBI deployment teams that have opted to install and configure SAS EBI in house (versus outsourcing to SAS Institute).

DEVELOPMENT/DEPLOYMENT

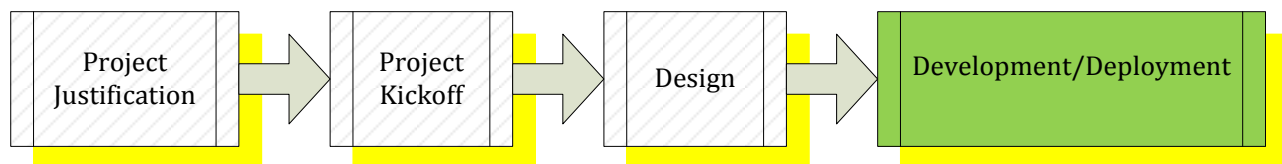


Figure 8: SELC Phase Diagram (Development/Deployment)

One of the widest variations in system release life cycle procedures from agency to agency rests with the allocation of server environments for specific development and deployment phase tasks. Some agencies require separate

environments for baseline development/deployment, system and user testing, and production (dev/test/prod). Others permit tasks to be spread across just two environments (dev-test/prod). Still others require up to four environments (dev/test/staging/prod). For this reason, I've specified the essential tasks for the development/deployment phase without reference to which environment each task should be delegated for completion.

ASSEMBLE YOUR DEPLOYMENT STAFF AND DEPLOY BASELINE APPLICATION

For SAS EBI deployment teams that have planned to outsource the installation and configuration of SAS EBI to SAS Institute, deployment must be postponed until the licensing and installation services are procured. For all other deployment arrangements, deployment phase tasks can begin under an evaluation license once the team has staffed the necessary roles:

- SAS Administrator
- Systems Administrator
- SAS ETL Developer
- SAS Programmer/Analyst with EBI experience
- Statisticians
- Data Modelers
- Database Administrators

Note that many deployment teams prefer to assign the SAS Administrator and Systems Administrator role to the same individual, if security policies permit. Likewise the ETL Developer and EBI Programmer roles can often be satisfied with one capable and experienced SAS programmer. For managed data environments with a large quantity of analytic tasks or a rapid deployment schedule, several programmers may be tasked at once and additional analytic experts such as statisticians and data modelers may be employed. Database administrators may also be required if SAS will be configured to host data sets in a local database such as Oracle.

The baseline deployment should be done in accordance with the system and logical designs authored during the design phase and with guidance from the SAS Institute installation and configuration manuals. It should include only the SAS EBI framework and any custom SAS code. No data or reporting components other than development or testing samples should be loaded into the environment prior to initial system testing.

DOCUMENTATION

Once the baseline deployment is completed, work on the version description document detailing the install, the systems administration and operations manual, and any other relevant deployment documents can begin. These documents will undergo continuous revision over the course of the development/deployment phase.



SYSTEMS TESTING

Full systems and application testing should be completed against the baseline deployment to verify the system functions as expected. Testing completed during this phase follows the testing plan authored during the design phase. Typically, a full security assessment and certification is completed at this juncture as well.

PERFORMANCE BENCHMARKING

Performance testing may be completed in conjunction with the remainder of systems testing as well as repeated as part of user acceptance testing once the system has been populated with data. This testing ensures the system performs as expected with regard to CPU cycles, memory and virtual memory utilization and bandwidth consumption.

SYSTEM POPULATION

Once the baseline deployment has been tested against expected functionality, the system can be fully populated with

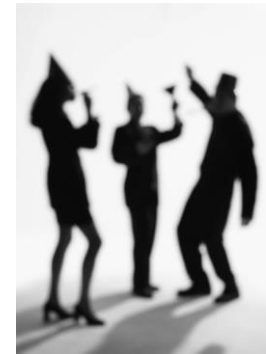
data and report objects (for managed environments) or additional sample data and reports (for unmanaged environments).

USER ACCEPTANCE TESTING

User representatives should be on hand to complete another round of thorough testing against the system to ensure that the system meets all user requirements.

LAUNCH PARTY

To ensure widespread user adoption and evangelism, it is very important to build on the initial momentum and excitement of a new system release. A casual but well planned launch party is an easy way to ramp up enthusiasm for the SAS EBI deployment. It is also a chance for the project team to connect with those we ultimately serve - our end users- one on one and build rapport. We want to provide users with the support they need to succeed using SAS. During the kickoff party, initial planning for a SAS users group within the agency or division can also be offered for consideration and user buy-in.



THE “GOTCHA”

While the make-or-break risk junctures of a SAS EBI deployment in the federal sector have been touched on in each relevant preceding section of this paper, they are highlighted as a group below for emphasis and study. If we do not plan for the “gotcha”, our project will fail! SAS EBI deployment project failure means:



- Deployment injunction
- Data integrity questioned
- Battles with systems admin/data center mgt over resources
- User apathy and abandonment because application is too hard, too slow, too confusing

SECURITY VIOLATIONS AND THE GODZILLA USER

The complexity of security regulations, policies, guidelines and restrictions in the federal sector can be paralyzing. SAS EBI deployment teams must pay careful attention at every phase of the life cycle to ensure the deployment is in full compliance. Note that out of the box, SAS standard configurations may not always comply with applicable security policies. For example, some agencies prohibit data storage on web servers with direct client access, even if they are internally facing. Customarily the SAS Content Server repository is stored on the mid-tier (web server with direct client access) and hence in violation of such a security policy. In these agencies the SAS Content Server must be relocated to another machine.

There is also the matter of the Godzilla user in unmanaged data environments. SAS security permissions do not recognize the ‘owner-author’ role for metadata trees and objects. This role traditionally allows all members of a group to share write permissions to a directory to enable document creation, but recognizes each individual’s right to subsequently edit or delete only their own documents in that directory. Lacking this role definition, SAS users given write permission for a shared group location (in order to author new content for the group and edit what they’ve authored) are able to edit or delete all objects in the group location such as reports, code, or information maps, including those for which they are not the author, unless permissions are manually edited on each and every new object after its creation. Such granular repeated permissions editing is extremely impractical in a large user setting. In such circumstances we must make it clear to users that any content placed in group locations are subject to the whims and actions of the entire group and that one bumbling Godzilla user can take out all of the content.

REMOTE USER ACCESS ERRORS AND LATENCY

As referenced in the section detailing the gathering of user requirements, SAS EBI deployment teams must give careful and deliberate thought to planning EBI client tool access to the server across various interface configurations, including via remote access. CITRIX and other virtualized environments may be a good fit for deploying EBI client tools in environments where desktop security, firewall access restrictions, or slow VPN speeds prevent users from effectively utilizing the tools locally.

CONFLICTING OR DUPLICATIVE BUSINESS RULES

Regardless of whether the SAS EBI deployment team is deploying a managed or an unmanaged data environment, conflicting business rules are the evil lurking just around the corner in the dark alley to take us down. For deployments in managed data environments, avoiding business rules that conflict or duplicate one another is a function of an orderly business rule definition document and a willingness to push back (gently!) against client requests to use the same variables with different underlying definitions across different report sets. For deployments in unmanaged data environments our options for mitigation are more limited. We can encourage (again, gently) all user groups to contribute to a centralized business rule definition document and can further encourage them to always refer back to the document to verify their business rules are not in conflict or duplicative.

POOR DATA STRUCTURE DESIGN

Well researched and well written user requirements with an eye for security compliance paired with competent SAS ETL developers and programmers translate into well executed data structure designs. In contrast, poorly formed user requirements or reluctance to press users for implicit requirements or ignorance of potential changes in requirements on the horizon will result in duplicative data structures and pre-summarized hierarchies that don't serve our end users as efficiently as expected. In managed data environments we must come to a deep understanding regarding the nature of our users' objectives and the data itself in order to implement requirements properly. A subject matter expert in the domain of our agency's or division's business can be a valuable resource to have on the deployment team, budget permitting. For unmanaged data environments, we must clearly and assertively guide our agency or division groups who will be using the system to elect their own subject matter experts, requirements managers, and capable SAS developers to populate the deployment.

OUR DATA CUP RUNNETH OVER

Even the most effectively designed systems can denigrate into unproductive chaos if controls on data accumulation and aging are not addressed. How often will data be refreshed? Will refreshes be in place, overwriting the existing data set or will each refresh populate an additional data set? For the latter, how many data sets will we permit concurrently? How much "play" space will we allocate each user? WORK library size? SASUSER library size? Total space allocated to a division or department? Will we allow exceptions to standard quotas for very large multi-variable analytics? Will there be a cap on the number of libraries, reports or Enterprise Guide projects permitted for registration? Systems employing ESRI mapping need to pay particular attention to sizing due to the extensive space requirements of geo-data mappings.

CONCLUSION

SAS EBI deployments launched in the federal sector that ultimately succeed and gain widespread user adoption and devotion are those in which the deployment team has established the clear business value of SAS EBI, successfully navigated through the political and technical complexities of the SELC process, remained mindful of security considerations at all times, and enthusiastically engaged the user community with respect and understanding of their objectives. These successful deployments improve the efficiency and impact of the sponsoring agency as well as increase visibility of the deployment team within the agency and before those whom the agency serves.



Figure 9: SAS EBI Deployment Success

ACKNOWLEDGMENTS

I am especially grateful to Christine Vitron of SAS who provided those first introductory lessons on SAS administration which serve as the foundation for all of my SAS knowledge. Likewise, Maria Nicholson of Zencos and I worked side by side on a SAS project for a federal contractor for four years during which time she allowed me to absorb as much of her SAS wisdom as I was able. Robert Matthews of CSC has been a steady mentor on the federal project management life cycle since I began my tenure with CSC. Finally I want to thank my fellow SAS users in the community, including those on the SAS Institute R&D, product management, and sales teams who are always at the ready to answer my questions via email, phone, or Twitter.

CONTACT INFORMATION

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

Name: Jennifer Parks
Enterprise: CSC
Address: 8960 Milroy Court
City, State ZIP: Manassas, VA 20110
Work Phone: 703.470.6067
E-mail: jparks23@csc.com
Twitter: [jenniparks](https://twitter.com/jenniparks)
Linked In: <http://www.linkedin.com/in/jennimparks>

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.