Paper 105-2009

BioSurveillance and the Holy Grid – Does it Compute?

Russell Gann
Centers for Disease Control and Prevention
Atlanta, GA

Abstract

The Centers for Disease Control and Prevention currently use Base SAS® and SAS® BI (OLAP, Portal, WebReport Studio) for the BioSense Application. Despite all caveats, the search for technology silver bullets continues in every field. The concept of a grid of computing resources originated in the 1970's, and has re-emerged in the 90's and this century as computing power, networks and middleware capabilities advanced to support it. Grid computing can provide a powerful backbone for healthcare research issues. However, establishing a grid is not just a matter of clustering large numbers of disks or blade servers, nor is it a solution for every public health issue. Grid-like resource usage can enable systems not typically associated with massively parallelizable problems to achieve higher levels of availability and reliability. This presentation will review the basics of grid architecture (i.e. grid computing, grid storage, grid collaboration) and existing uses in public health (for example, caBIG™), potential future uses relative to biosurveillance systems such as BioSense, and the pros and cons of introducing grid models in this arena. This presentation will be based upon experience with BioSense, IDS and CRA from the CDC and both SAS® and BioSense's relevance to the move towards a federated model based on grid technology.

KEYWORDS: biosurveillance, computing grid, business intelligence, federated data

Executive Summary

Public health surveillance is intrinsically a massive collaboration and data sharing effort. Components vary from a simple system collecting data from a single source, to electronic systems that receive data from many sources in multiple formats, to complex surveys. The number and variety of systems will likely increase with advances in electronic data interchange and integration of data, which will also heighten the importance of patient privacy, data confidentiality and system security. This additional variety among the collaborating systems will also have a tendency to negate the ability for any single technology to satisfy all requirements.

Bio-surveillance is often thought of simply as syndromic surveillance; however the ability to detect events requires a broader set of information than that of syndromes. Bio-surveillance systems need to take advantage of integrated data from multiple sources including electronic health information not traditionally monitored by public health.

BioSense is an automated early event detection and situational awareness system to evaluate case and suspect case reporting along with statistical surveillance and data visualization of pre-diagnostic data to support the earliest possible detection of potential public health emergencies, whether natural or man-made.

The BioSense Influenza Data Store (IDS) is designed as an integral component of the Centers for Disease Control's (CDC) BioSense application. The IDS component provides web-browser-based-views of available influenza data. The IDS architecture is based on the Rational 4+1 View Model for software architecture, as is the entire BioSense application. This document describes the 4+1 View Model's implementation as it relates to the IDS component and the overall design of the component. The IDS is provided as a mechanism for public health users at the National, State and local levels, to monitor, assess and track the events related to Influenza and Influenza outbreak/pandemic preparedness within the general population.

Grid architectures have been hyped to be another silver bullet of technology. Due the complexity of requirements and disparity of data sources necessary to bio-surveillance, there will not be a single solution to all requirements, but grid architectures can be the answer to some of these requirements.

At present, BioSense utilizes base SAS® for it underlying data warehousing and data management tools. SAS® offers the flexibility to implement any number of methodologies, from a balanced scorecard to Baldrige, Six Sigma or any other customized framework. SAS®' capabilities provide a number of benefits:

- Better aligned resources with strategies. Understanding which measures drive your business, you can begin to align operations toward those drivers. Predictive score carding capabilities let you tie traditional scorecard features such as traffic lights and speedometers to powerful predictive analytics.
- Synchronization of financial and operational strategies. Collaboration on the development of scenarios and associated budgets. With one integrated solution for financial planning, budgeting and reporting, SAS® gives you the ability to create more frequent, accurate forecasts, ensure financial transparency and publish reports more quickly. Dynamic consolidation also ensures that the latest information is always available for decision makers and regulatory agencies.

- Quickly respond to business and market changes. At-a-glance dashboard visuals ensure that you have the most efficient access to your organization's current status. Automatic alerts inform you immediately of important events. You can also view cause-and-effect relationships between events, which help illustrate results in the proper context. A diagram editor makes it easy for you to visualize strategies and process flows. All these capabilities are offered through one integrated solution.

- Ensure a single view of enterprise information. To ensure a clean, consistent view of information within each department, SAS®' performance management offerings are all built on the SAS® Enterprise Intelligence Platform. This platform offers unmatched data integration, storage, business intelligence and analytic intelligence, so you can always be confident that your choices will deliver the expected results.

- Reduce costs without compromising profitability. SAS® provides activity-based management to let you get a complete operational view of cost and cost drivers. You can improve financial results by analyzing the profitability of products, customers and channels to help you make the right decisions for product development, marketing and distribution.

SAS® provides a number of focused business solutions to enable performance management in various departments within your organization:

- IT – Clarify the real costs and value of your information assets, and identify the best projects and strategies to maximize that value.

- Human resources – With a holistic view of your workforce, analyze and optimize human capital.

- Finance – Create a cohesive, holistic perspective of true costs, profitability, regulatory compliance, forecast scenarios and optimization strategies.

- Marketing – Understand customer behavior, risk and profitability, then drive more effective offers and marketing campaigns.

- Procurement – Analyze spend and supplier performance, so you can make strategic sourcing decisions and optimize your supplier base.

Introduction

The BioSense Application is designed as an object-oriented product which follows the Rational 4+1 Model View architectural approach. The following diagram is a high level graphical depiction of the 4+1 Model View:

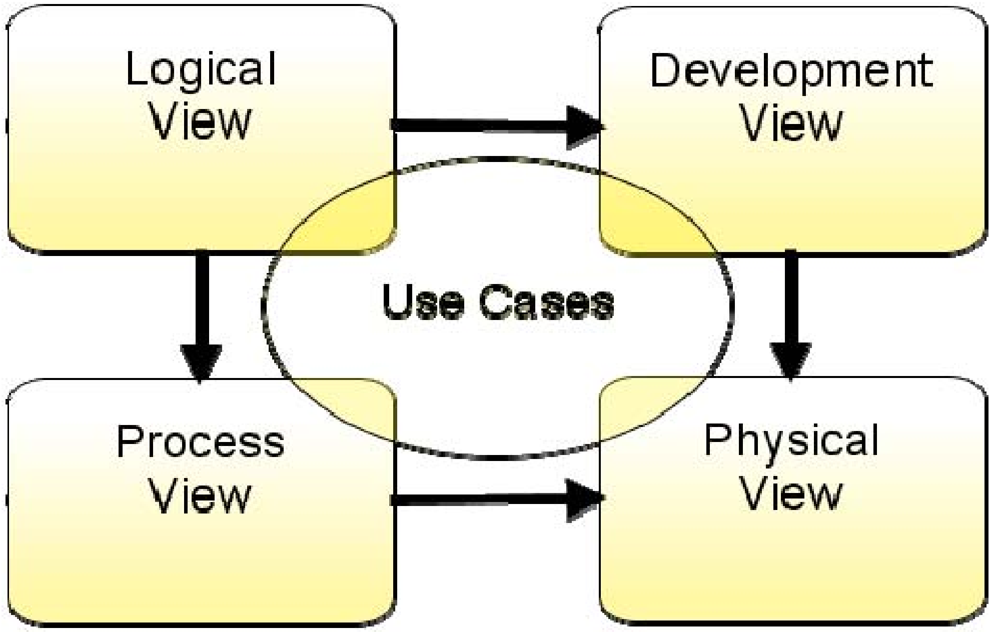


Figure 1

The remainder of this document will define the 4 architectural views and the over arching use cases with respect to the IDS component of the BioSense application. These views are utilized to address various stakeholders' concerns and are intended to be concurrent representations of the IDS.

Logical View

Logical view is intended to provide a better understanding the functionality of the IDS component. This view is primarily based on the functional requirements as defined by the CDC client. As previously stated, the application design is object-oriented-based; therefore, the logical view is represented as an object model.

The data for the IDS is based on Jurisdiction:

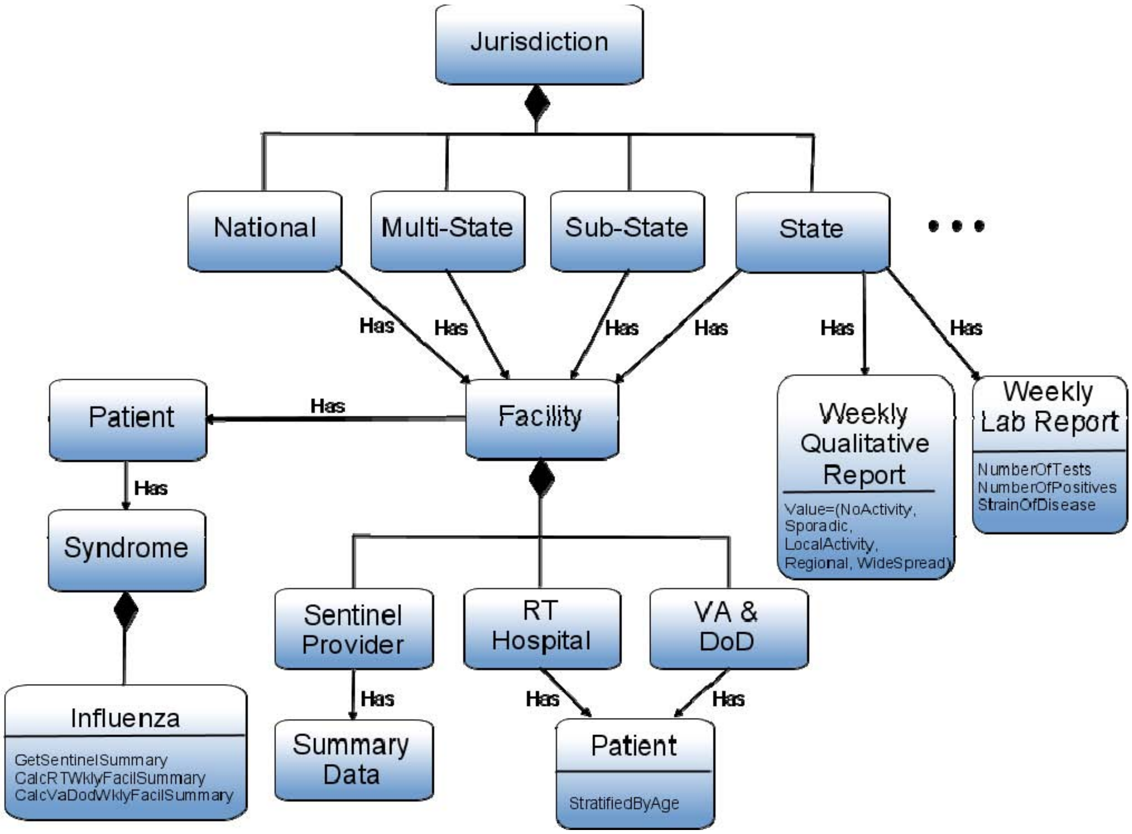


Figure 2

Jurisdictional decomposition exists across numerous geo-divisions (i.e. National, Multi-State, Sub-State, State, etc.). These divisions contain facilities which contribute to the diagnosis and/or treatment of patients with influence like illnesses.

States also produce two weekly data reports which are not specifically related to facilities:

> Qualitative Reports – Where the state reports one of the following flu related statuses: No Activity, Sporadic, Local Activity, Regional, or Wide Spread.

Facilities are presently Sentinel Providers, Real-Time Hospitals, or VA/DoD facilities. The Sentinel Providers supply only summary data, where the RT Hospitals, VA, and DoD provide detailed patient information from which summary data relative to influenza can be calculated.

The IDS user population can also be decomposed into permissions based groups:

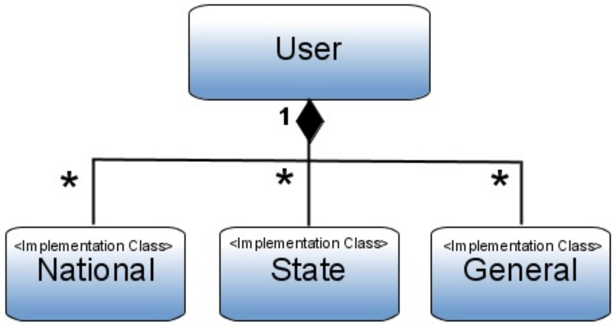


Figure 3

User roles have been created to manage the level at which a user may gain access. Data access privileges will be based on the following user roles:

- National level users have access to all data available from all states, jurisdictions and data sources without any geographical or Jurisdictional boundaries. CDC users are considered National users.

- State Level users have access to data available from within their State boundaries and jurisdiction.

- General users (i.e. all users which fall outside National or State privileges) only have access to summary data. These summaries are only provided for National/Regional levels; therefore, the General user is restricted from seeing detailed data.

The IDS application may be decomposed into screens of presentation or components:

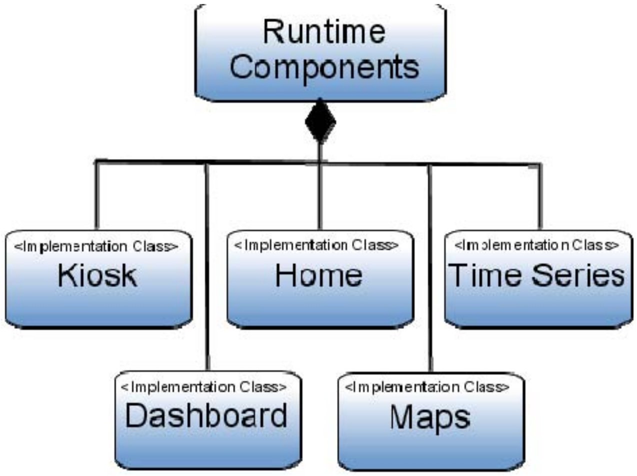


Figure 4

Runtime Components

There are five modules within the IDS component. All modules are portal-based. By providing IDS in a portal-based design the users will be provided the means to configure the IDS portion of the application to their preferred screen layout, within the limitations of each screen. The following defines each of the modules in this design:

Home

The users will be able to add links and bookmarks to a variety of information outside of the IDS and have them retained on the page for future reference. The user is also provided with the means to search external files, applications, news feeds, and websites for specific terms (e.g., avian influenza). Users are then able to choose which of the results they would like appear on their Home page.

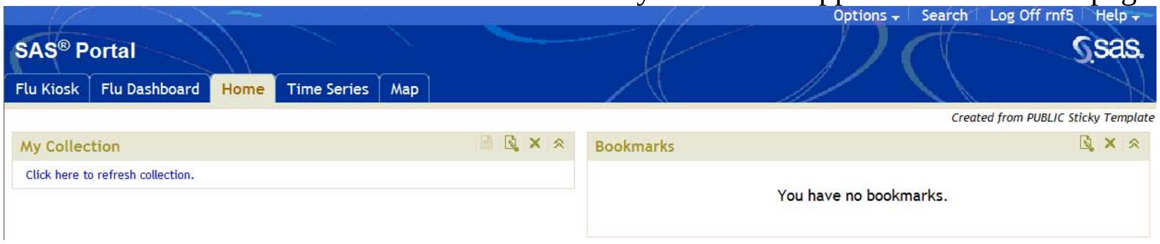


Figure 5

Kiosk

The Kiosk will have the following components and is configured by CDC personnel:

- **BioSense Flu Status**: Which is a brief influenza status report containing bullets or other text summary of current national influenza activity as described in the weekly influenza surveillance report.

- **Flu Links**: Links to influenza-related websites and news feeds.

- **Flu Status Dashboard**: See below for additional information on this component.

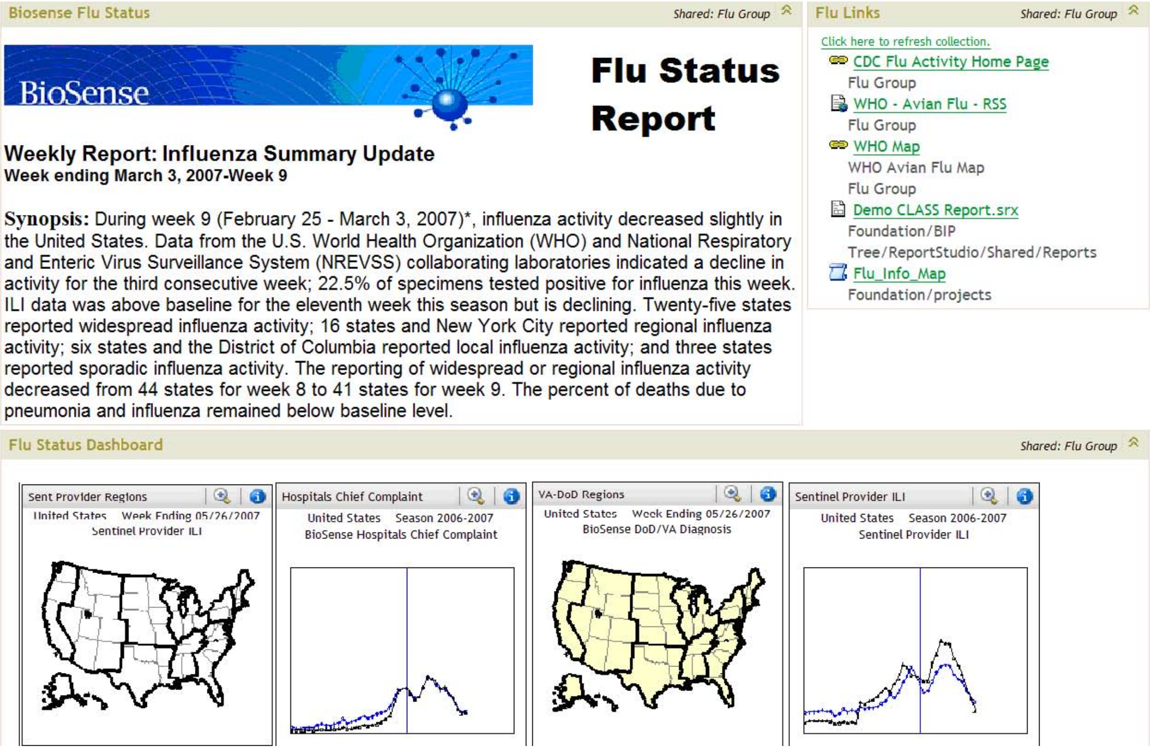


Figure 6

Time Series

Time Series plots representing event numbers and/or rates by week will be displayed with the associated data. Users have the ability to filter results by selecting the data type (e.g., Sentinel Providers), influenza season, and geographic area. The system will also have mouse over capabilities to display all metadata associated with the point on these plots.

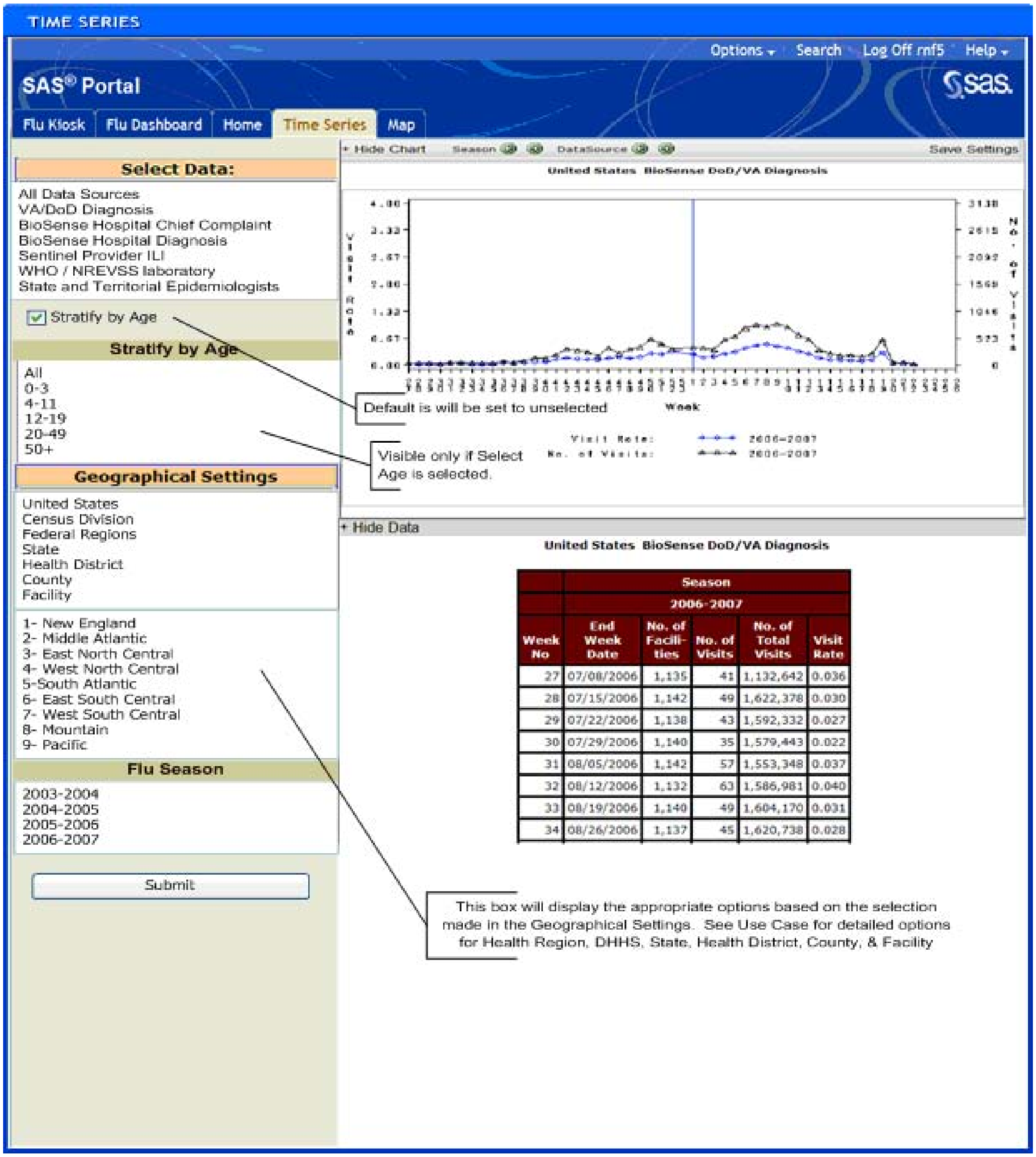


Figure 7

Dashboard

The Dashboard consists of a variable number of saved images from the Time Series and Map pages.

The content of these dashboard images is determined by the user. For example, users from a given geographic area can create a summary similar to the Flu Kiosk, but specific for their area. Thus, the user can simultaneously view all data applicable to their geographic area. The user can rename this screen and create and save additional customized dashboards. As with the Flu Kiosk, these dashboard images can be set to auto-update with the latest available data.

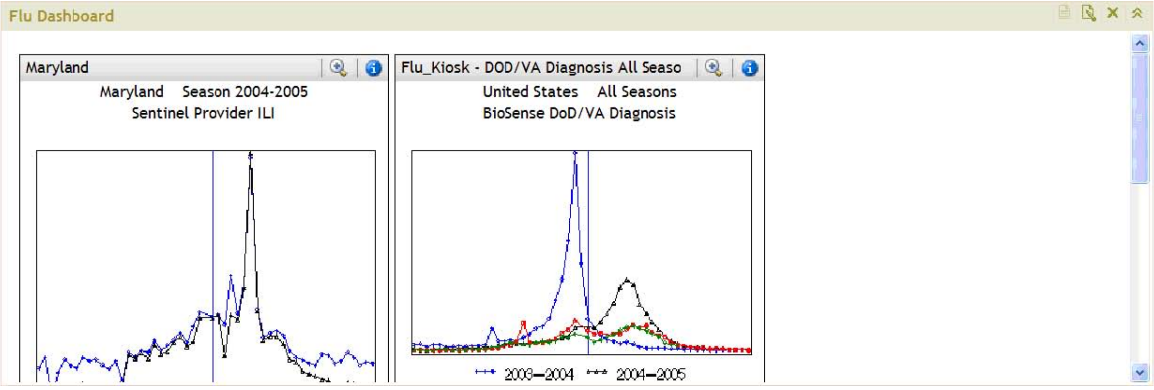


Figure 8

Maps

Maps allow the user to select data type, influenza season and week, and geographic area. Users can click on geographic regions to drill down to finer geographic detail. For outpatient ILI data, the number of facilities (or providers), ILI visit count, total visit count, and ILI rate can be obtained by mousing over the desired geographic region. Maps may also be filtered by geo-region, data source by year or week with a given year, flu season or week with in a flu season and maps session may be dynamic or static images

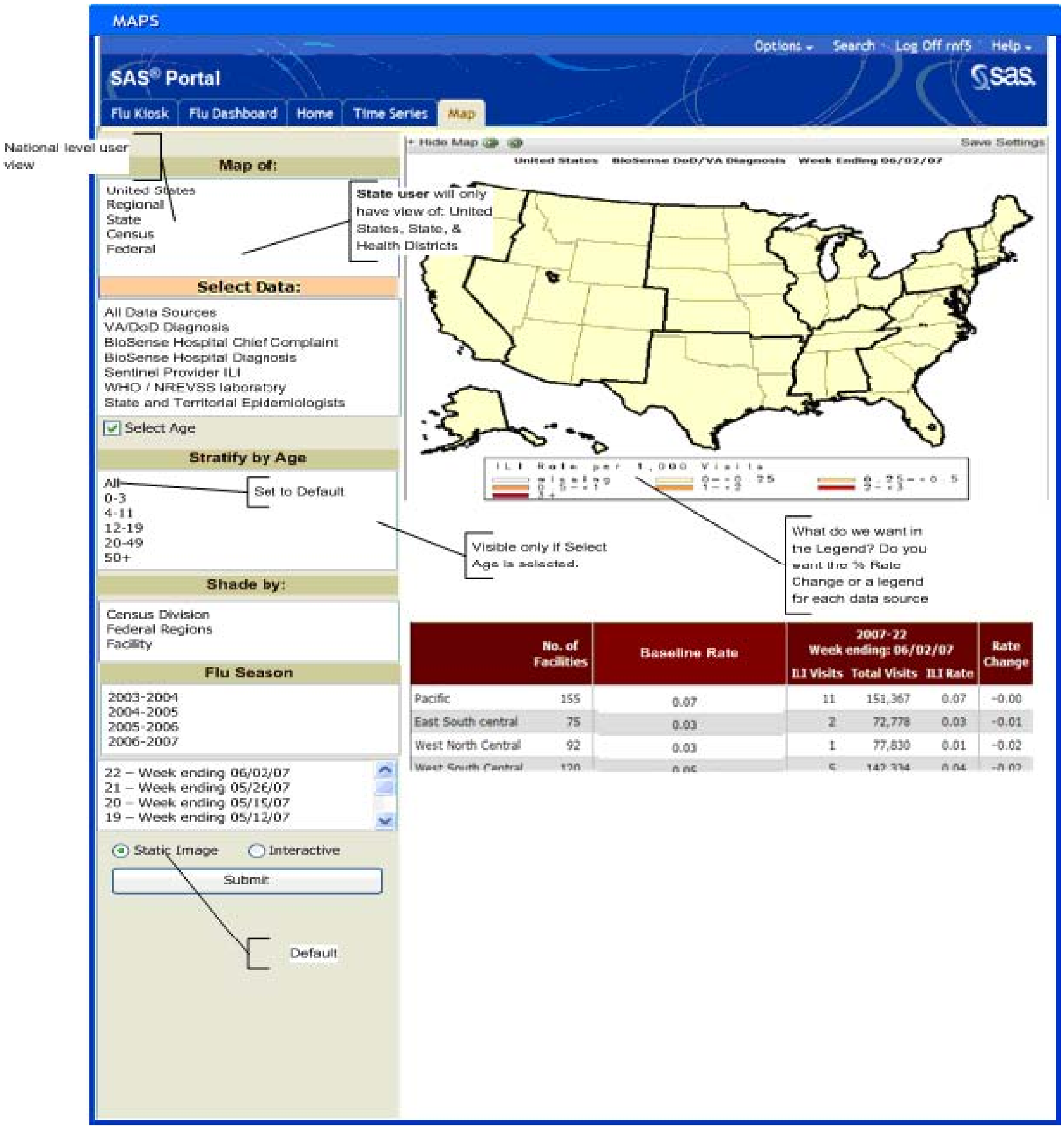


Figure 9

The IDS runtime components will be access via the SDN through the BioSense application:

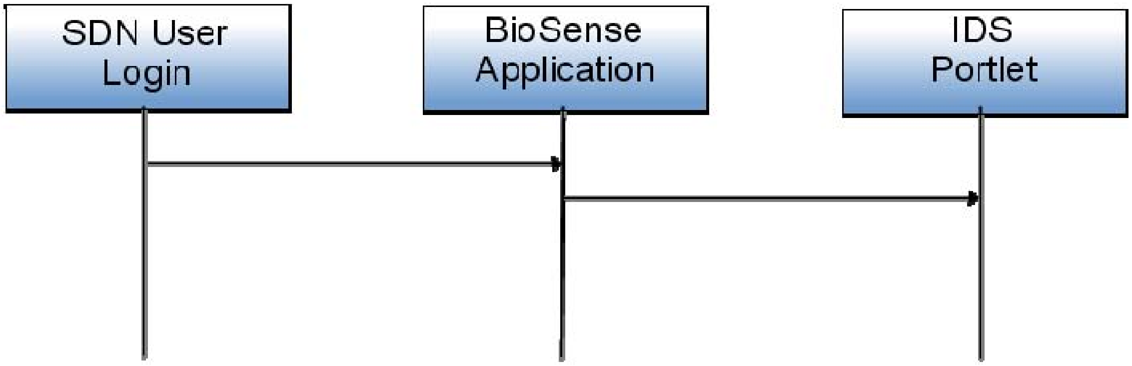


Figure 10

The following is a representation of the logical software stacks for IDS and SDN respectively:

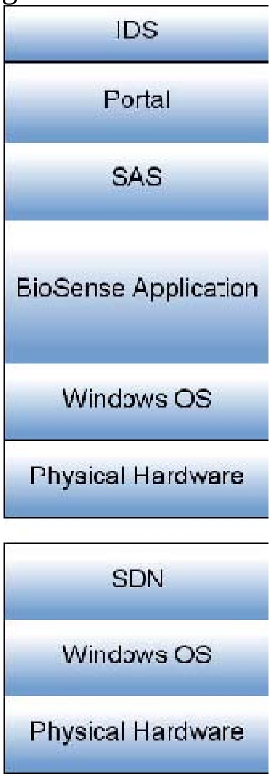


Figure 11

Development View

The IDS is packaged with the existing BioSense Application and as such follows the following java packaging structure:

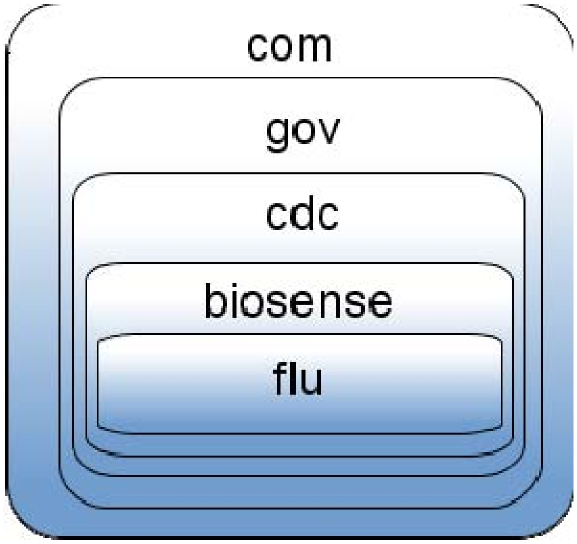


Figure 12

The IDS will be coded within a portal framework to allow the users to maintain their own configurations of the component.

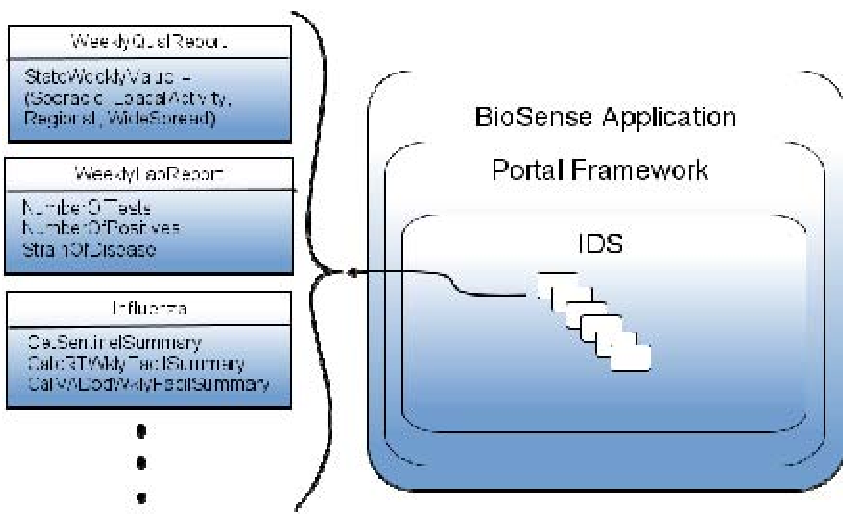


Figure 13

Portal Definition

The SAS® portal is used in the IDS component. The following is the definition of the portal as provide in the *SAS® Portal User's Guide* (http://www.regione.piemonte.it/osservatorio/turismo/ptur/help/en_US/ doc/portal/uguide/index.htm) :

> The SAS® Information Delivery Portal is your personalized gateway to up-to-the-minute information. The portal gives you single-click access to the data, applications, and documents that you need to make decisions and do your job. It differs from other portals in that it has the power and flexibility of SAS® behind it.
>
> The portal gives you easy access to SAS® reports, tables, databases, and stored processes. It lets you launch applications and view customized output. It gives you access to sites on the Web or your company intranet. It also lets you subscribe to publication or content channels that feed you information on a continual basis.
>
> When you become an authorized portal user, the portal desktop is yours to modify at will—set up your own portal windows; set up your own lists of data, subscriptions, applications, or links; and organize them in the way that makes the most sense to you. You can change the desktop at anytime; the changes affect only your view of the portal.

Dashboard Design Model

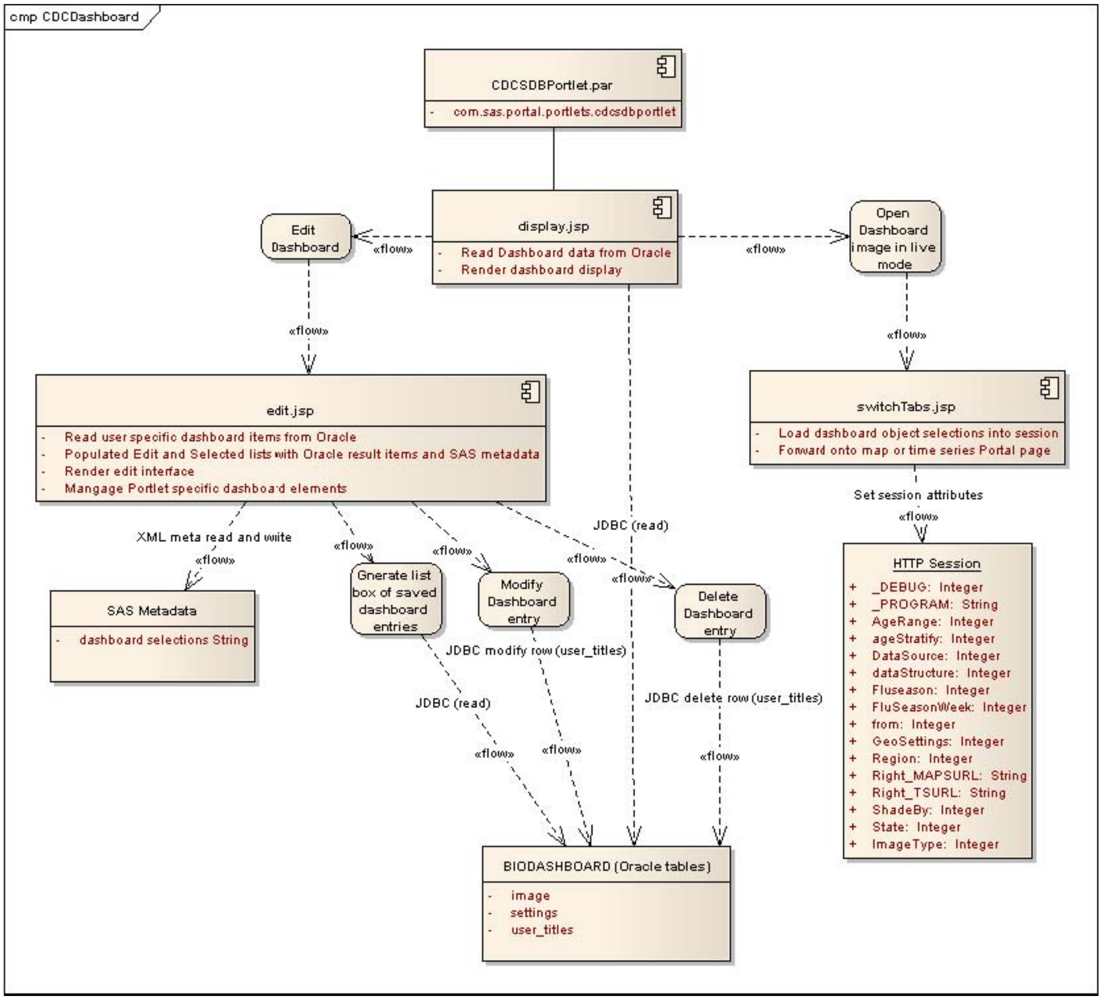


Figure 14

Time Series Design Model

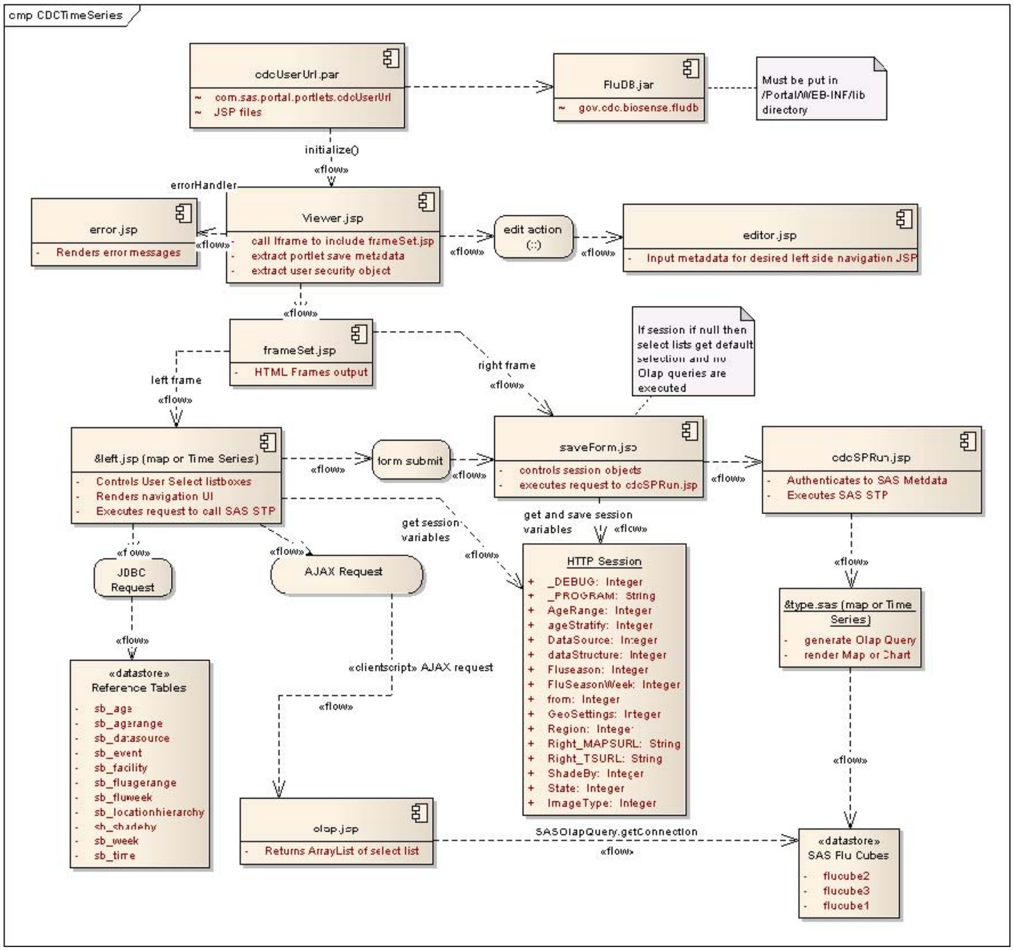


Figure 15

Portal Design Model

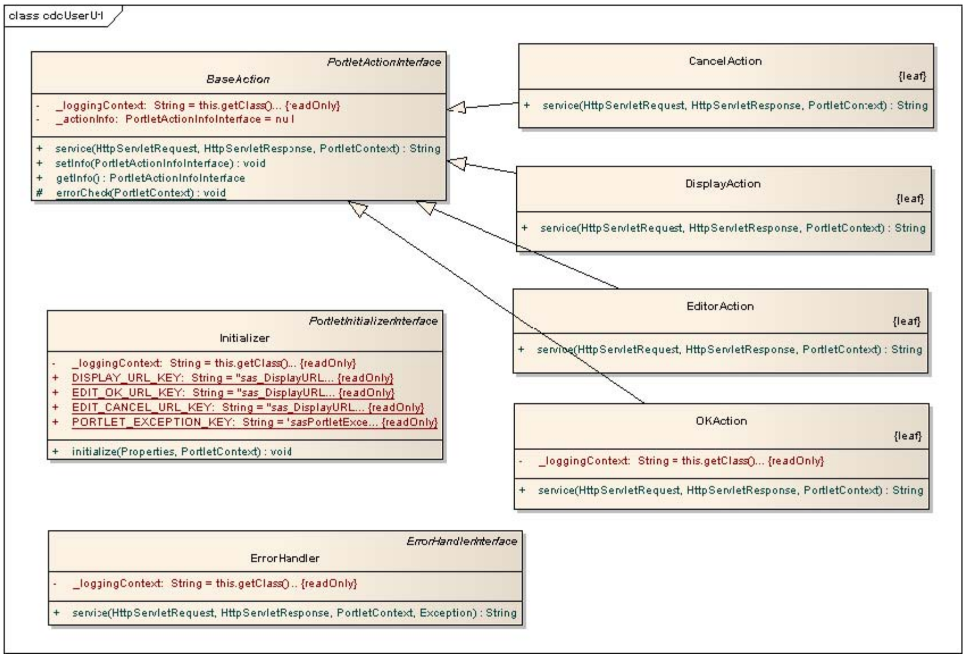


Figure 16

Database Class Model

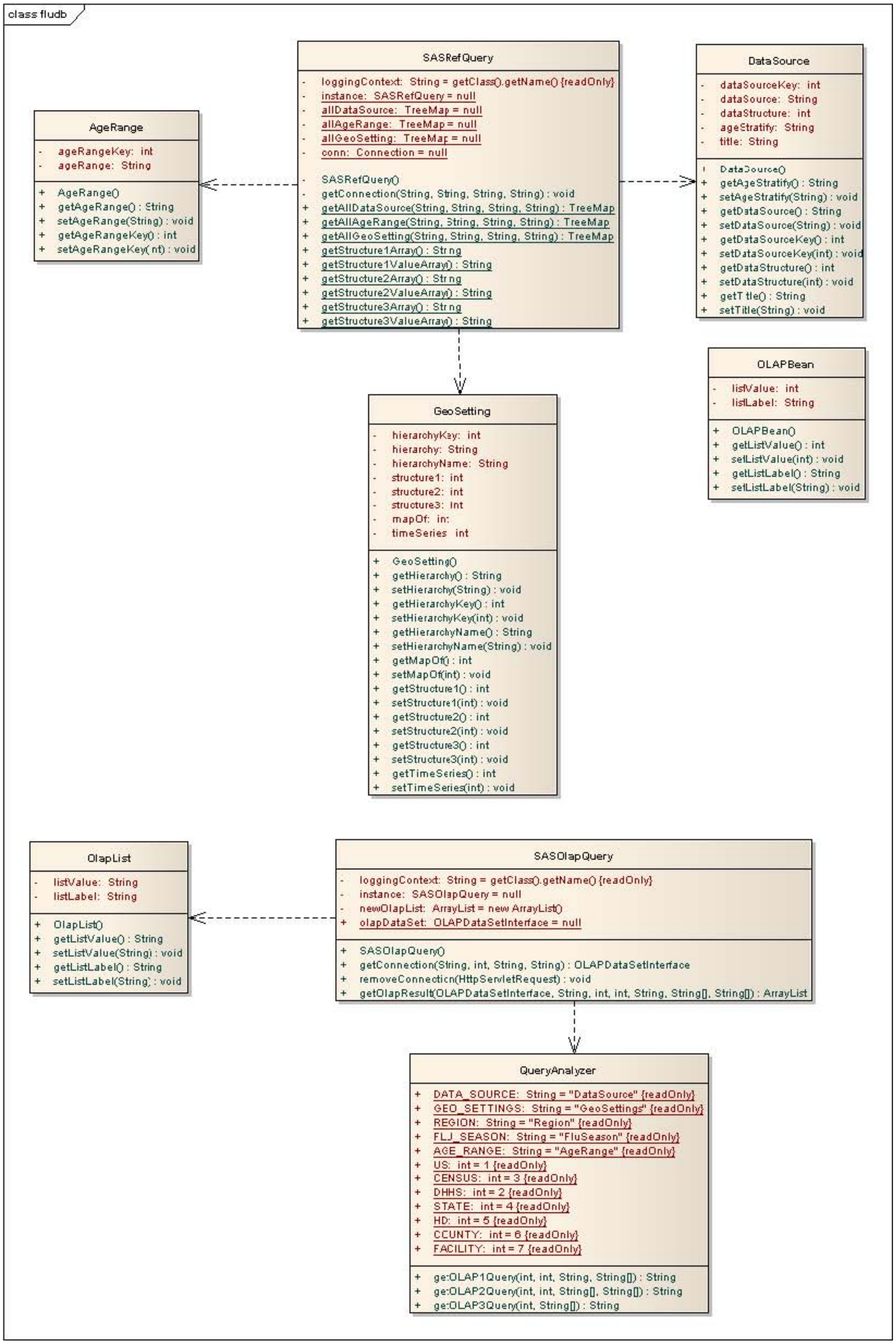


Figure 17

Oracle Data Model – for Dashboard

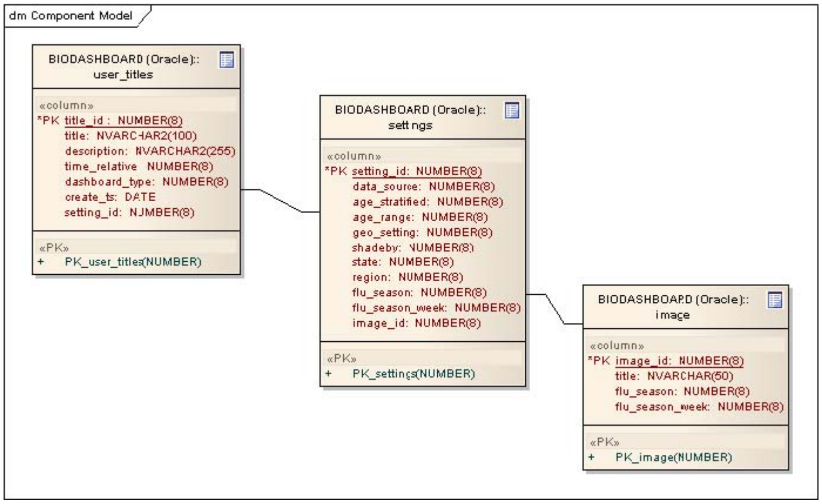


Figure 18

Image Save Model – for Dashboard

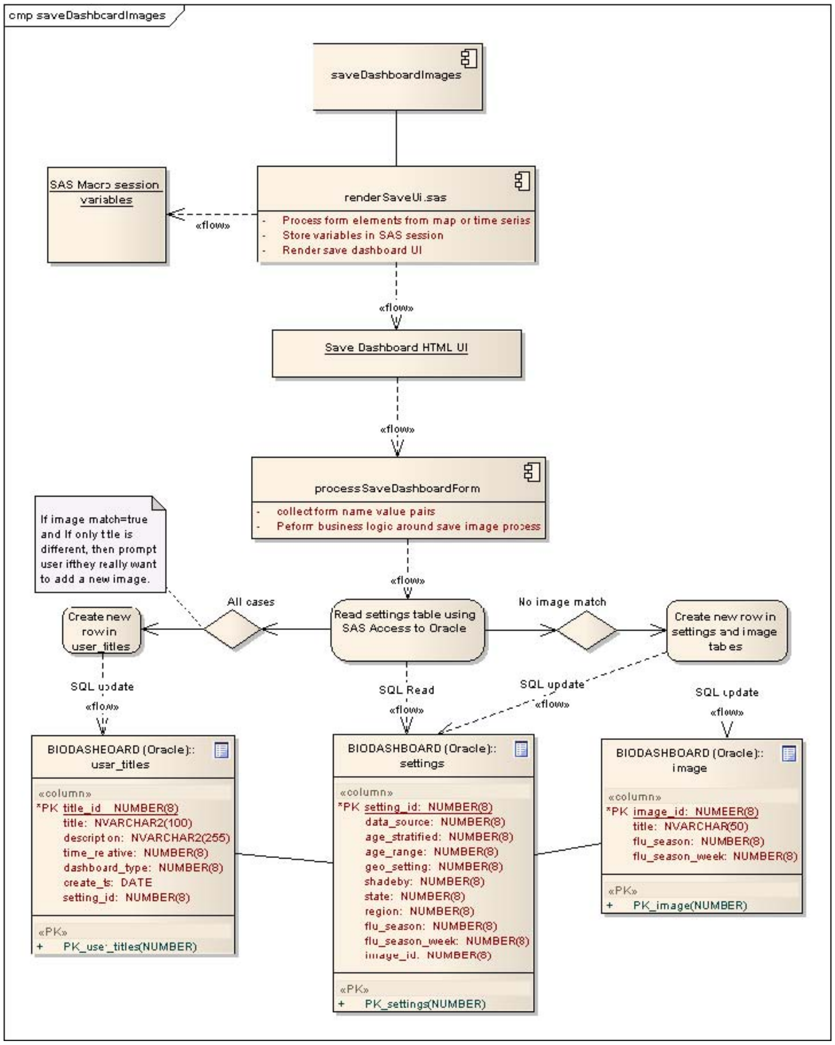


Figure 19

Process View

Data Flow

There are 3 data sourcing types viewable in the IDS component: Type I, Type II, Type III
Below are the respective data flows of these three sourcing types:

Data Source Type I

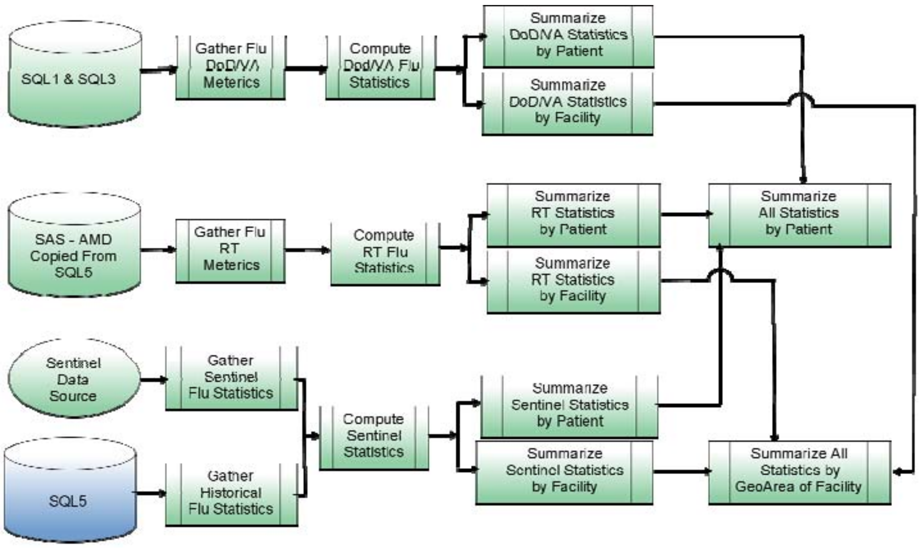


Figure 20

Data Source Type II

Figure 21

Data Source Type III

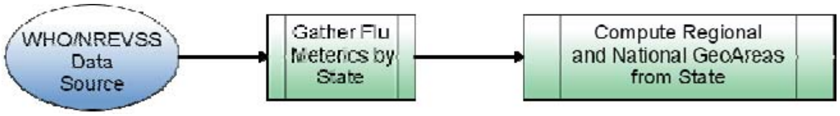


Figure 22

Time Series – Menu Flow

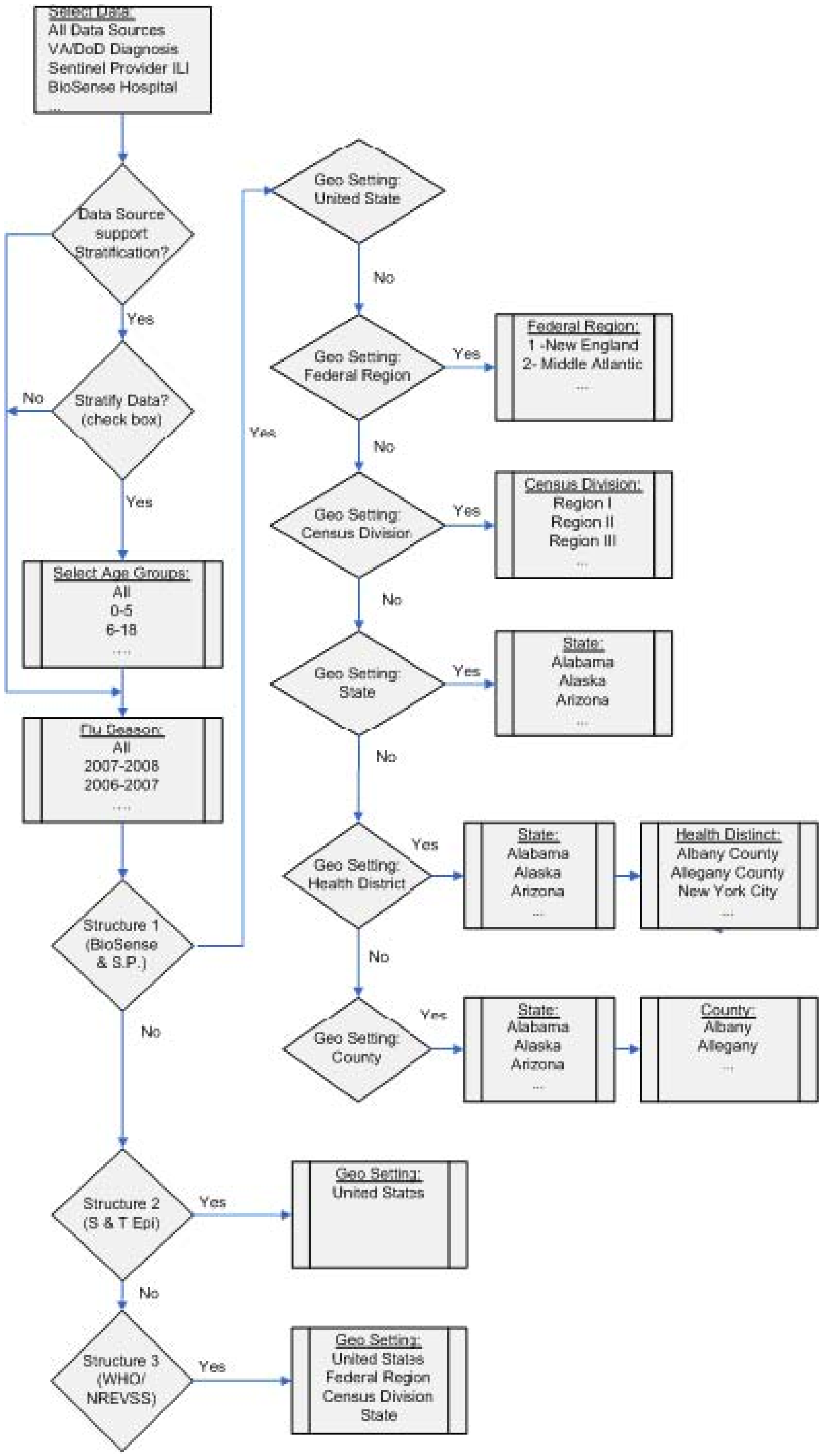


Figure 23

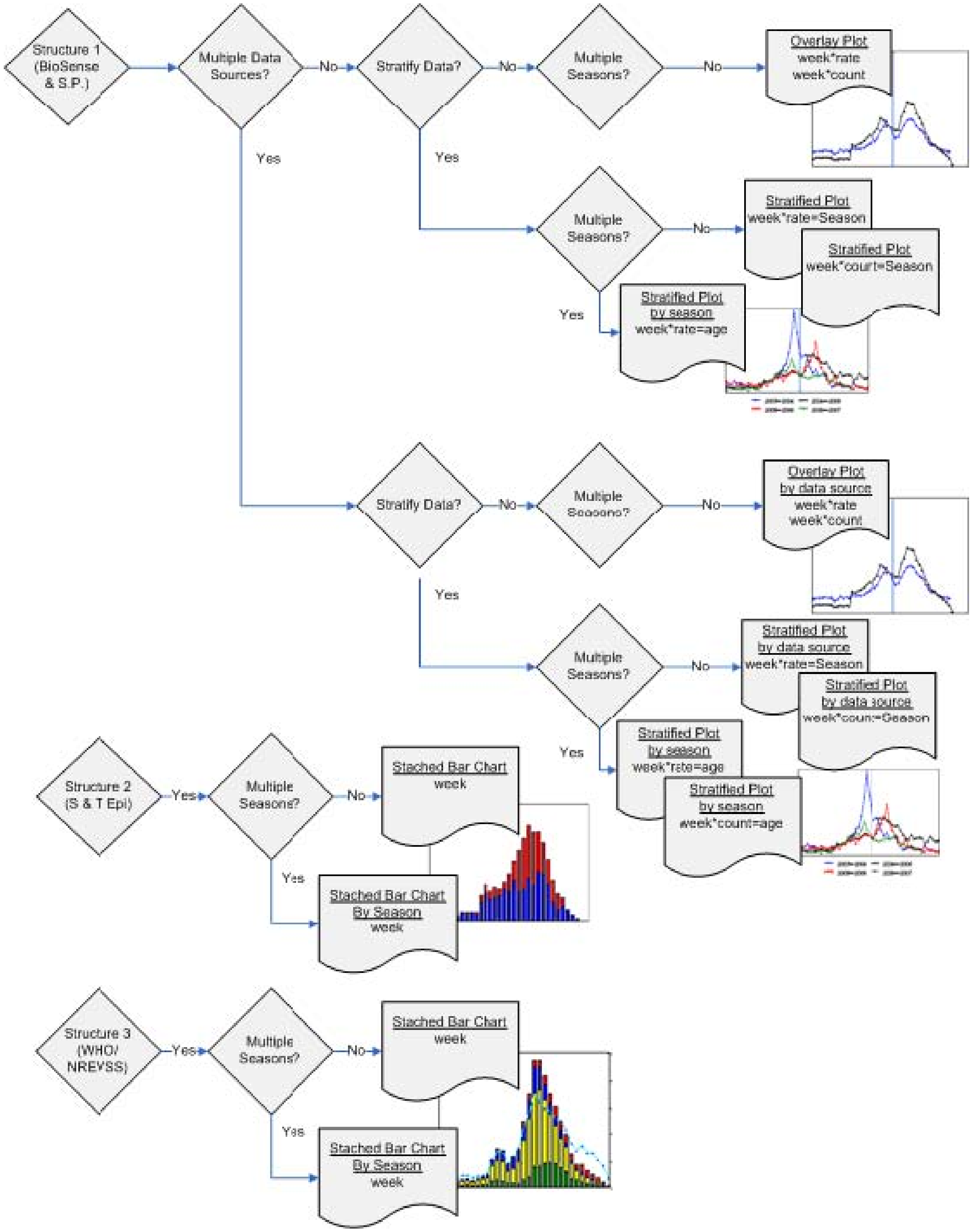

Structure 1 (BioSense & S.P.)
Multiple Data Sources?
No
Stratify Data?
No
Multiple Seasons?
No
Overlay Plot
week*rate
week*count
Yes
Yes
Multiple Seasons?
No
Stratified Plot
week*rate=Season
Stratified Plot
week*count=Season
Yes
Stratified Plot
by season
week*rate=age
Stratify Data?
No
Multiple Seasons?
No
Overlay Plot
by data source
week*rate
week*count
Yes
Multiple Seasons?
No
Stratified Plot
by data source
week*rate=Season
Stratified Plot
by data source
week*count=Season
Yes
Stratified Plot
by season
week*rate=age
Stratified Plot
by season
week*count=age
Structure 2 (S & T Epi)
Yes
Multiple Seasons?
No
Stached Bar Chart
week
Yes
Stached Bar Chart
By Season
week
Structure 3 (WHO/ NREVSS)
Yes
Multiple Seasons?
No
Stached Bar Chart
week
Yes
Stached Bar Chart
By Season
week

Figure 24

Map - Menu Flow

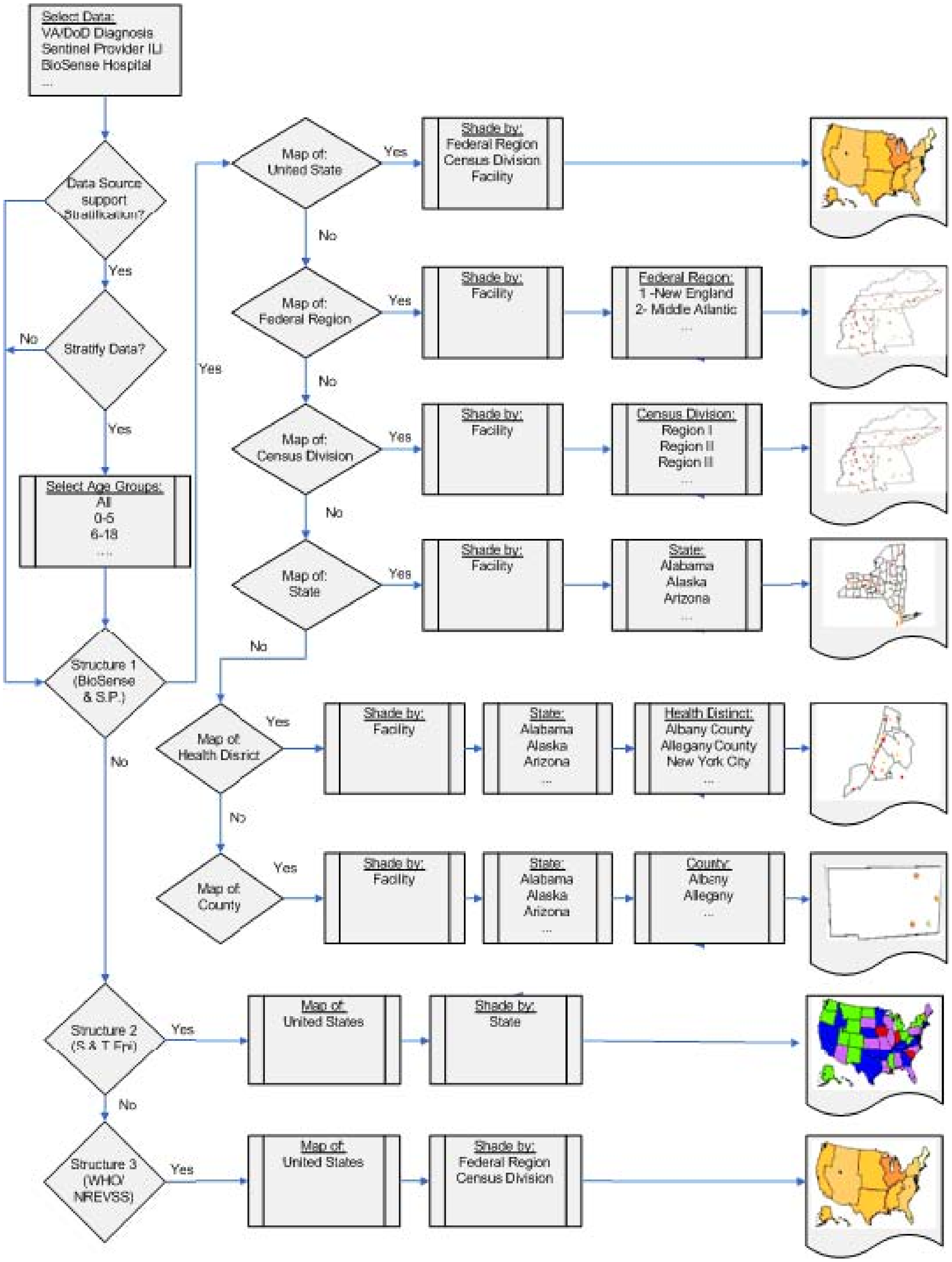


Figure 25

Business Use Cases

The use case for the IDS can be found in:

UC View IDS Home Page
UC View IDS Time Series
UC View IDS Maps
UC View IDS Dashboard
UC View IDS Kiosk
UC Edit Custom Display

Technical Use Cases

The following are technical use case for the Influenza Data Summary (IDS) tool (A.K.A. Flu tool). Here the architectural and technical detail will be defined as use cases. To better understand the use case the following diagrams are provided on the SiteMinder Agent (i.e. message level interceptor used to manage internet URL processing).

Logical SiteMinder Flow

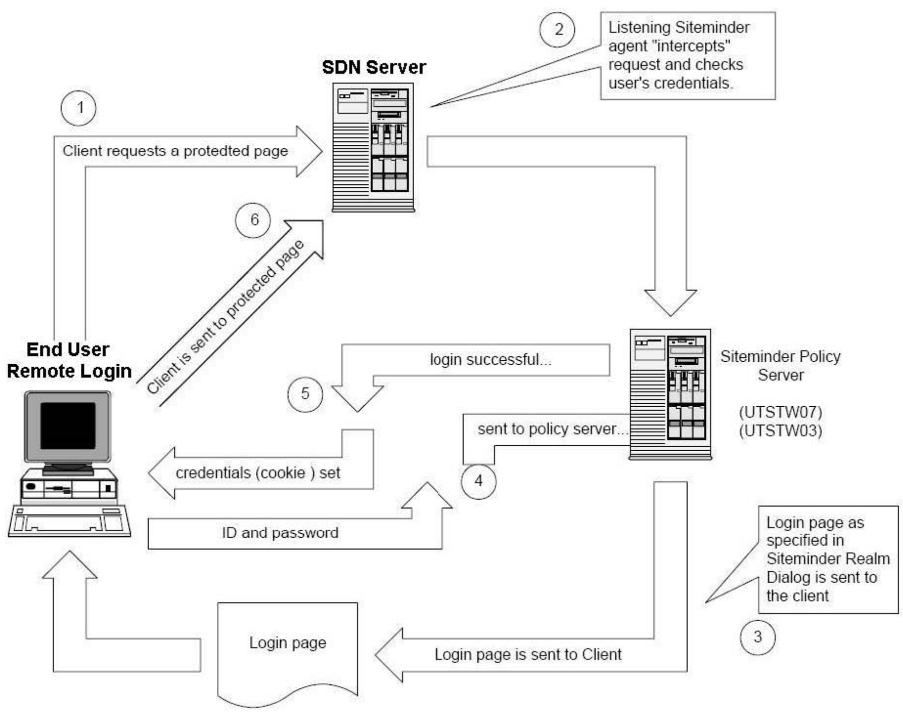


Figure 26

BioSense Activity with other SDN Activities and URL Proxies

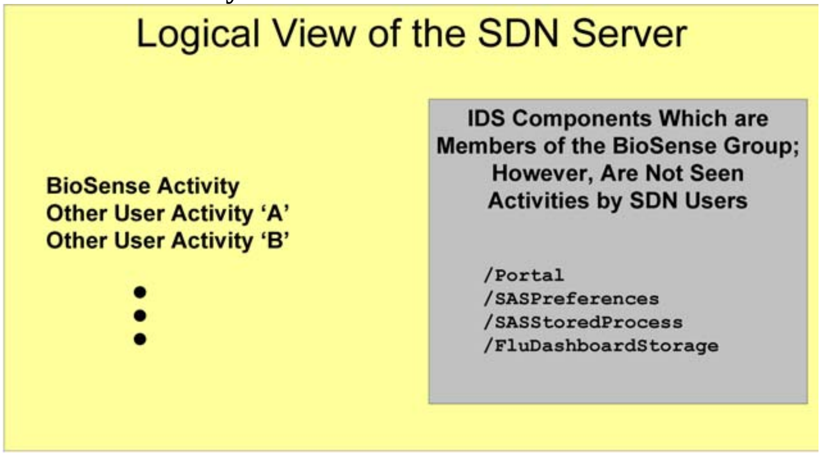


Figure 27

IDS Access

Step 1: An end user of BioSense must be a certified SDN user. Contact the CDC BioSense Application Administrator to gain this certificate.
Step 2: The certified user logs in to SDN via a web browser to enter their user id and password.
Step 3: On successful login the user will be delivered to the SDN Activities: Where one such activity is BioSense.
Step 4: Select the BioSense Activity. *To access any of the features of BioSense such as IDS (Flu) one must do so from within the BioSense application.*
Step 5. Once the user has accepted the User Rights, a requirement of using the BioSense application, the BioSense home page will be presented. One of the possible selections on the home page is the IDS component. By selecting this button the user will be forwarded to the IDS application.

BioSense Implementation of SiteMinder Agents

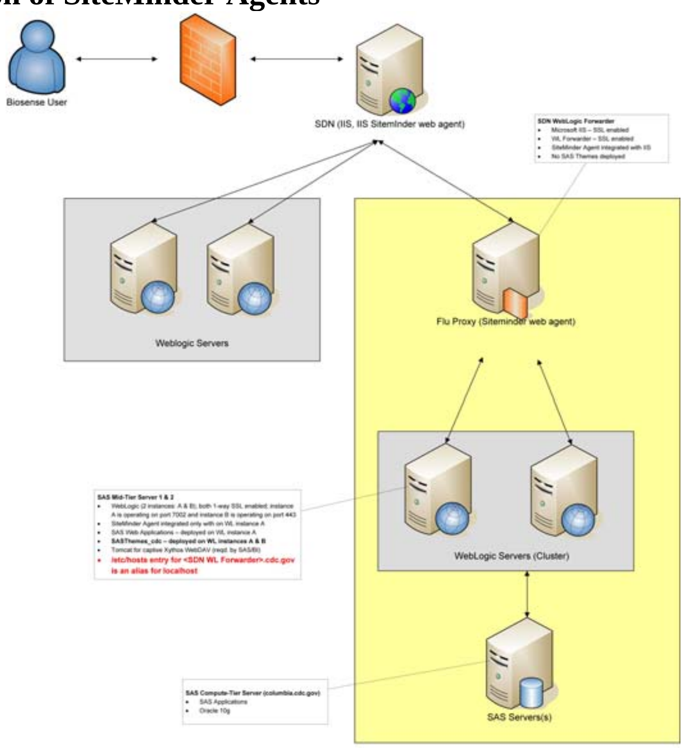


Figure 28

Forwarding the User from BioSense to the IDS

After a successful login is achieve and the user has navigated to the BioSense Home page. Gaining access to the IDS is made by selecting the IDS button. Unseen by the application user is the agent to agent communication which is defined here:

Step 1: Select - From within the BioSense application via the user's web browser the user selects IDS. On this select a message is transmitted to the SDN server. The SDN SiteMinder agent intercepts the message inspects the message header. This agent then route the message to the appropriate sub-system as defined by the agents configuration. In this case the BioSense message will be routed to the IDS (A.K.A. Flu) System (See Figure 3 in the yellow highlight).

Step 2: Initial Routing - This routing take place with the aide of the proxies defined in Figure 2 in the grey highlight. All defined IDS proxies are members of the BioSense group; therefore, no additional login is required.

Step 3: SiteMinder IDS Agent - The SiteMinder agent on the IDS system further inspects the message and establishes a link to the appropriate URL.

Physical View

The following diagram is intended to define the multiple BioSense system environments. The yellow highlighted area represents a proposed single instance of SDN which would support DEV, TEST and STAGING. The hardware implementation will be the responsibility of ITSO for non-Solaris equipment in DEV & TEST environments and MTDC for STAGING, PROD, and Solaris equipment in DEV and/or TEST. For specific hardware information pleases contact the ITSO or MTDC teams.

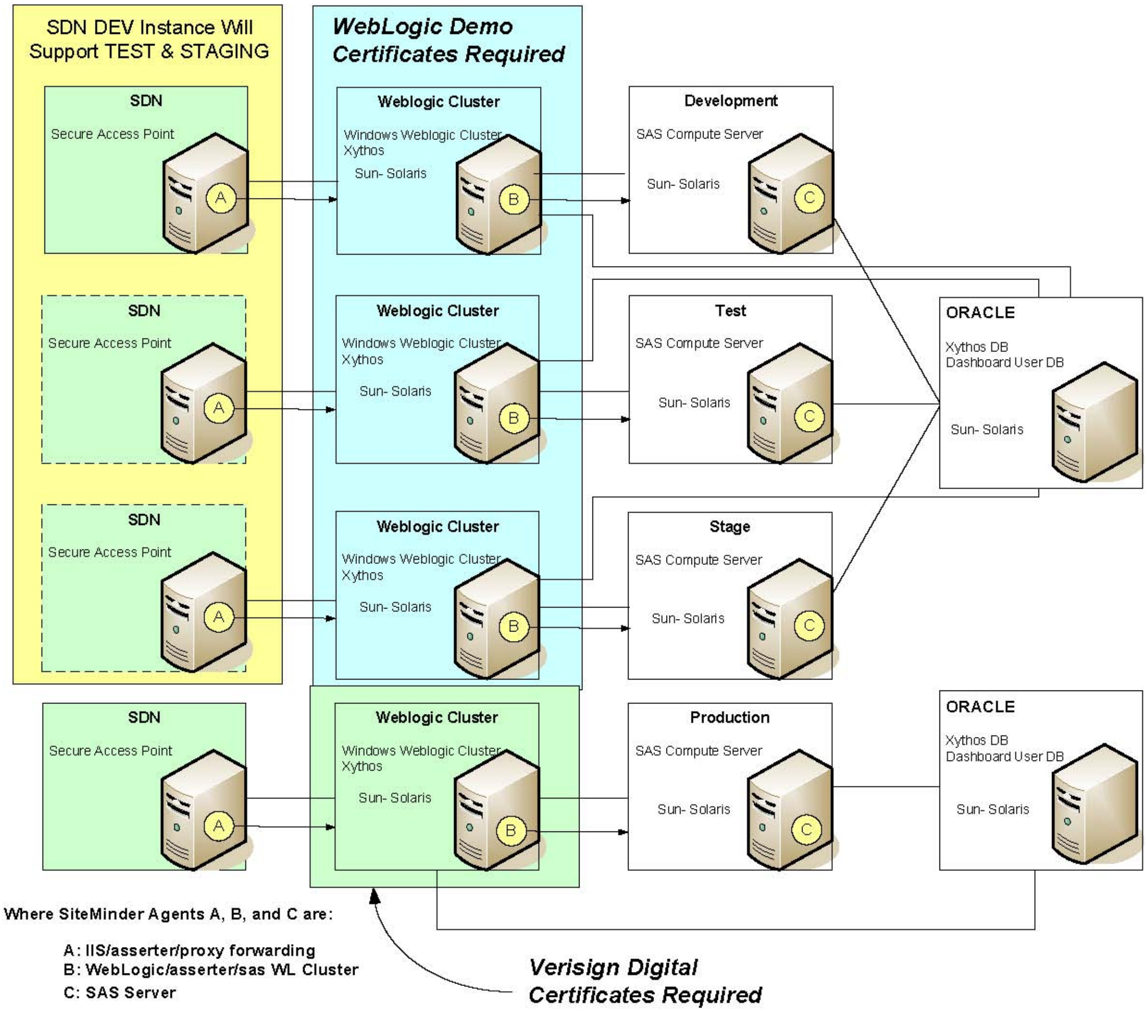


Figure 29

If the physical hardware is provided as defined above there will be 2 SDN server instances:

- Server 1 in support of DEV, TEST, & STAGING
- Server 2 in support of PROD

BioSense Activities will need to be provided for all 4 environments where the application development team has full access to these activities in DEV and TEST and only limited, as per SLA agreement, in STAGING and PROD.

Priority and Expected Dates for Specific Environments

Priority 1 DEV Available for Development Use: August 16, 2007
Priority 2 TEST Available for Development Use: August 21, 2007
Priority 3 STAGING Available for Deployment: August 21, 2007
Priority 4 PROD Available for Deployment: August 21, 2007

System Configuration

It would be preferred that all for environments be identical; however, it is necessary for the DEV and TEST to be identical and for STAGING and PROD to be identical.

Software Deployed

Application Tier (on Solaris 10 Update 3)

- SAS® Business Intelligence Platform v9.1.3 SP4
- SAS® Foundation Services v9.1.3 SP4
- SAS® Management Console v9.1.3 SP4
- SAS® WebReportStudio v9.1.3 SP4
- SAS® WebReportViewer v9.1.3 SP4
- SAS® WebOlapViewerforJava v9.1.3 SP4
- SAS® Query & Reporting Services v9.1.3 SP4
- Xythos WebDAV v2.2
- Tomcat v4.1.31 (captive installation for Xythos by vendor)
- WebLogic Portal® 8.1.4
- Sun JDK 1.4.2_05
- WebLogic Identity Asserter (SDN Agent), for specific release information contact SDN personnel

Compute Tier (on Solaris 10 Update 3)

- SAS® MetaData Server v9.1.3 SP4
- SAS® v9.1.3 SP4
- SAS® Connect Server v9.1.3 SP4
- SAS® Object Spawner v9.1.3 SP4
- SAS® Share Server v9.1.3 SP4
- SAS® OLAP Server v9.1.3 SP4
- SAS® Foundation Services v9.1.3 SP4
- SAS® Remote Services v9.1.3 SP4

Database Tier (on Solaris 10 Update 3)

- Oracle 10g v10.2.0.3.0

CONCLUSION

Shall we grid?

Grid architectures have been hyped as yet another silver bullet of technology. National public health surveillance is intrinsically a massive collaboration and data sharing effort. Simply adding a new technology or group of complementary technologies such as those that support grid computing is never the answer for a complex problem, such as what we face with integrating public health information. To take BioSense as an example, receiving real-time data from the divergent HRIS and clinical systems of hospitals, local public health, the VA, and other data sources means that we must pull data, sanitize it and relate it to the needs of CDC's bio-surveillance requirements in a number of different ways.

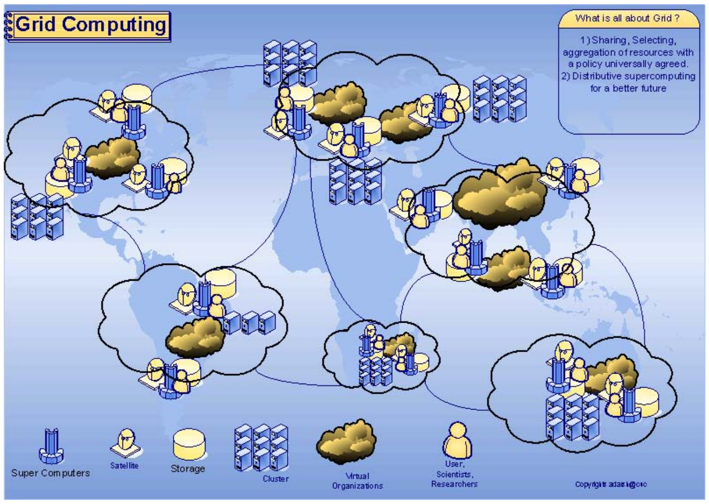
Furthermore, establishing a grid is not just a matter of clustering large numbers of disks or blade servers. Although throwing more horsepower at a problem sometimes solves issues, to make a grid infrastructure function appropriately requires

- High-speed data movement
- Ability to cache very large amounts of data
- Robust scheduling algorithms and security mechanisms
- Appropriate developer toolsets

These requirements have to be considered soberly when embarking on a grid-enablement program.

There are also more types of grid computing that just computing grids.

- Data grid is the controlled sharing and management of large amounts of distributed data.
- Computing grid (or the use of a computational grid) is applying the resources of many computers in a network to a single problem at the same time
- Collaboration grid is a combination of the sharing of large amounts of distributed data while utilizing the computational grid tools to process distributed data co-located with the tool sets.



Data grid storage approaches are being utilized to address the public health need to save patient imaging data. The requirements for housing this data increases in size as technology improvements increase image resolution. In other words, as equipment provides denser images that require more space, storage capacity needs to grow with the demand. Growth of digital medical images will increase from 308,000 terabytes in 2003 to 1,250,000 terabytes by the year 2006 (Yankee, 2003).

As an example, Banner Health, one of the largest not-for-profit healthcare systems in the United States, has 20 hospitals and other facilities in eight states. Its HP solution accumulates storage of 800,000 annual radiology/cardiology procedures. Its distributed, multi-tier storage uses grid technology for redundancy and fault tolerance.

The Iowa Health System (IHS) utilizes IBM® Grid Medical Archive Solution (GMAS) to allow cardiology, radiology and other digital images to be shared across multiple sites and securely stored for years to help improve patient care, advance research and reduce administrative costs. The 13 terabytes of Picture Archiving and Communication Systems (PACS) is expected to double in size over the next year alone.

Grid computing requires the use of software tools that can divide and farm out pieces of a program to as many as several thousand computers. Grid computing can be thought of as distributed and large-scale cluster computing and as a form of network-distributed parallel processing. Grid approaches to resource usage can enable systems, not typically associated with massively parallelizable problems, to achieve higher levels of availability and reliability.

SAS/Connect® software establishes connections between networked computers with various operating systems and enables organizations to combine the resources of these machines for maximum benefit. It efficiently distributes computing workloads across different CPUs and provides scalability through parallel processes. The National Institute of Environmental Health Sciences (NIEHS) used SAS® to enable a compute grid for a toxic genomics study.

Malaria kills more than one million people each year, most of them young children living in Africa. Now physicists in the UK have shared their computers with biologists from countries including France and Korea in an effort to combat the disease. Using an international computing Grid spanning 27 countries, scientists on the WISDOM (World-wide In Silico Docking On Malaria) project analyzed an average of 80,000 possible drug compounds against malaria every hour. Up to 5000 computers were used simultaneously, generating a total of 2000 GB (2,000,000,000,000 bytes) of useful data.

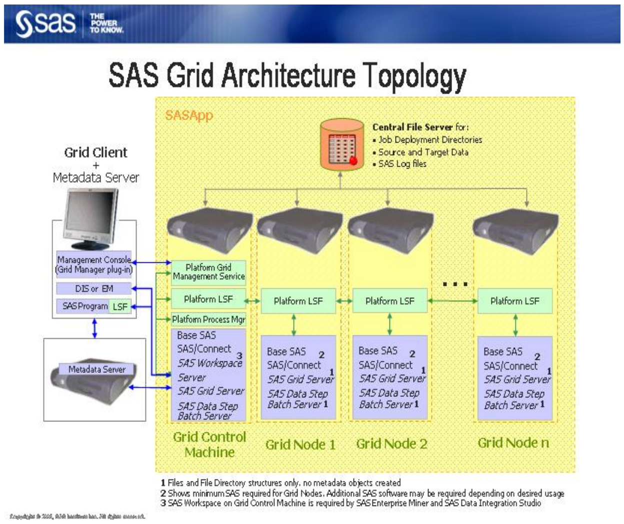
TeraGrid™ is an open scientific discovery infrastructure combining resources at nine partner sites to create an integrated, persistent computational resource. TeraGrid™ integrates computers, data resources and tools, and experimental facilities around the country. Currently, TeraGrid™ resources include more than 250 teraflops of computing capability and more than 30 petabytes of online and archival data storage, with access and retrieval over networks. Researchers can also access more than 100 discipline-specific databases. With this combination of resources, the TeraGrid™ is the world's largest, most comprehensive distributed cyberinfrastructure for open scientific research.

Again, the two important take-aways about grid computing are: you must have an appropriate problem that is parallelizable, and the software that coordinates the grid infrastructure is key to success, and can be very pricey.

Collaboration Grids are the combination of the computational grid tools to process distributed data and unite geo-graphically dispersed user communities.

One example of collaboration grid, which you will hear more about from later speakers, is caBIG™. The National Cancer Institute (NCI) launched the caBIG™ (cancer Biomedical Informatics Grid™) initiative to accelerate research discoveries and improve patient outcomes by linking researchers, physicians, and patients throughout the cancer community. caBIG™ allows sharing of both datasets and tools, and enables independent researchers around the world to coordinate efforts and outcomes on cancer research with each other.

caTRIP (Duke Comprehensive Cancer Center)



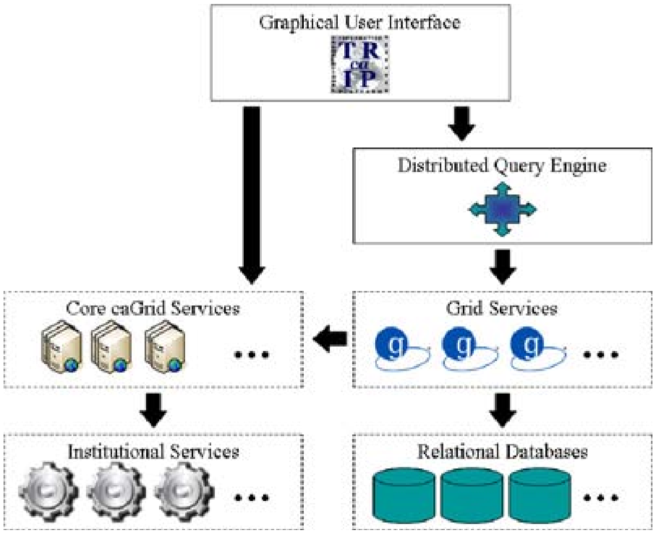
An example of tactical use of collaboration grids occurs in coordinating responses for large public events. For example, in 2003, an exercise event to prepare local San Diego emergency medicine providers was executed. One of the intended objectives of Shadow Bowl was to allow participating organizations to demonstrate rapid response, on-demand services and expertise for a mass casualty event. Shadow Bowl participants established a national medical communications grid to host the data, provide servers, database, and communication links to network sites. This communications grid included physiological, biological, and environmental sensors integrated into a distributed network with real-time displays to enhance critical communications for medical surge capacity. This collaboration between first responders and field units, the City of San Diego, and other agencies melded together real-time surveillance routines.

.

Grid Architecture and Bio-surveillance

Grid architecture could be utilized in bio-surveillance activities, in all three categories, data, collaboration and computing. As BioSense collects data from disparate sources like hospitals, labs, public health agencies, pharmaceutical companies and all, data retention and storage management policy becomes more critical. With a sense of mission, being whether its situational awareness or public health studies, BioSense is basically the one-stop shopping area for data related to national public health as well as individual emergency preparedness issues. Currently, storage usage is about 1 terabyte, but with growth of new data sources, this could double or triple within the coming year.

BioSense is not on the order of magnitude of a PACS or anything with the digital imaging repository world. But still, with demand growing for this type of information, its not trivial.



There are data mining requirements for this type of data as well as the need for real-time data access. The types of tools necessary to effectively use these applications are numerous, but there needs to be a balance between the needs for both support detection and situational awareness.

For bio-surveillance as well as for use in BioSense, it would be possible, during a catastrophic public health event that computers of all makes and processing power could be re-tasked to support additional user sessions and analytics. Companies playing in the Javaspaces environment include Gigaspaces. This type of technology enables the system to scale out horizontally across many dissimilar platforms taking advantage of all available CPUs and memory in the event of a “surge” of users.

Prototyping is underway on this technology. A proof-of-concept (POC) was undertaken by the BioSense Application Development Team (SAIC). In the POC the GigaSpaces product was thoroughly tested within the future architectural frameworks of the BioSense Application. The results were quite promising. GigaSpaces provided significant application time improvements over previously attempted approaches. If the product is found to be economically feasible it is seen as a valuable tool for improving the applications speed, fail-over robustness, and scalability.
The Division of Emergency Preparedness and Response is actively seeking better ways to securely share data sets and analytic tools between BioSense and our public health partners. Within a Service-Oriented Architecture, services are relatively large, intrinsically unassociated units of functionality, which have no calls to each other embedded in them. As Dr. Lenert noted in the opening plenary address, our goal is to allow fairly large chunks of functionality to be strung together to form ad-hoc applications which are built almost entirely from existing software services. The larger the chunks, the fewer the interface points required to implement any given set of functionality.

One of the problems of selling this vision to IT, IMHO, is the lack of successful implementations that weren’t basically riding on the back of some big enterprise-class product like SAP, or heavily invested in proprietary middleware and SOA management code (IBM, etc.). There is hope to leverage from experiences such as caBIG™ and the like for the next generation of BioSense development.

Grid Models for BioSense – Pros & Cons

A collaborative model for BioSense would have more flexible architecture for collaborating with other public health entities. Organizations like caBIG™ should be examples for us to follow to collaborate within public health environs. Because of examples like the Operation Shadow Bowl we can see that both researchers as well as 1st responders can benefit from grid models. Because of caBIG™ and SETI we can see a collaborative setting where research and science can share information in an environment already proved in the public sector.

Security experts recommend that IT administrators clearly identify and understand the security risks associated with large-scale grid computing deployments. Greater sharing of information and resources across traditional trust boundaries will result in increased risks that must be addressed as a matter of urgency. To ensure the confidentiality, integrity and availability of crucial data on the grid systems, administrators must implement proper classification to handle confidential data.

The return on investment of a silver bullet solution can be costly, both to purchase as well as to maintain. It appears that an expectation that some new technology or practice will easily cure a major prevailing problem. But, as in tradition, the silver bullet is only effective against a certain kind of monster.

Grid Computing – BioSense?

Within the scope of possibilities for BioSense, we have to consider a number of issues. These possibilities are of using grid computing, data grids or a collaborative combination is what we're here to work out. There are political, social, security, financial and public health implications to all potential solutions to this very high-visibility application.

CDC has taken the next step in its path towards a more productive, secure and shareable vision of BioSense by securing new leadership and with that leadership, a new direction. To have a BioSense that is receptive to SOA, includes collaborative grid architecture that scales to all levels of public health, and is federated [forgot to include federated in here somewhere] is the new track for the BioSense train. In the previous and coming months there have been and will be many discussions concerning BioSense that will lead to its new mission statement.

So, grid computing is not just the phrase of the day, but an actual set of architectural approaches that has been proven in a number of examples to benefit public health. These examples will lead the way to a better, more collaborative system of tools and datasets that will become CDC's bio-surveillance of the future. Keep in mind that close scrutiny needs to be made for a careful security-conscious implementation of grid computing techniques that will improve the extensibility and collaborative nature of BioSense, as a utility to CDC and its partners, as well as to the public at large.

Typically, advice is least heeded when most needed, but in conclusion, I would like to offer the following statements from some very learned authors:

- Software entities are more complex for their size than perhaps any other human construct, because no two parts are alike (at least above the statement level). … In this respect software systems differ profoundly from computers, buildings, or automobiles, where repeated elements abound (Brooks, 1995).

- There is no silver bullet and frankly you probably don't need one. It is far more important to be able to find the right kind of gun, be able to load the gun … and perhaps most importantly, be able to figure out where the werewolf is (Oliphant, 2006).

References

Brooks, F.(1995). *The Mythical Man-Month: Essays on Software Engineering, 20th Anniversary Edition*. Reading, MA: Addison-Wesley, 1995.

Oliphant, M. (2006) *Usability Works*.

Yankee (2003). Yankee Group Study Shows the Stakes Are High in 2003 for Storage System Customers and Vendors. Business Wire. June 16, 2003. FindArticles.com. 22 Jul. 2008. http://findarticles.com/p/articles/mi_m0EIN/is_2003_June_16/ai_103374226 .