NATIONAL SYNDROMIC SURVEILLANCE PROGRAM (NSSP) IMPLEMENTATION PLAN FOR ESSENCE AND SAS

8/14/2015 DRAFT

Table of Contents

Implementation Plan Vision

The goal of National Syndromic Surveillance Program (NSSP) is to strengthen the practice of syndromic surveillance (SyS) at the local and state public health agencies and to support the use of syndromic surveillance data for regional and national situational awareness.

Background

The NSSP is supported by the Division of Health Informatics and Surveillance (DHIS) at the Centers for Disease Control and Prevention (CDC). Recently, DHIS conducted a pilot to test the introduction of two new tools (ESSENCE and SAS) into the BioSense Platform, a syndromic surveillance GovCloud environment. The pilot was in response to a functionality needs survey conducted by the BioSense Governance Group (BGG) to inform decisions about what software products CDC should make available in the cloud for the syndromic surveillance community. Based upon the results of the survey, DHIS planned and executed a pilot project of the potential use of ESSENCE and SAS tools in the Amazon Web Services (AWS) GovCloud environment in collaboration with eight jurisdictions, the US Department of Veterans Affairs (VA), and the Department of Defense (DoD) from December 2014 to May 2015. The objectives of this pilot were to test ESSENCE and SAS applications for security issues and obtain approval to operate each application in the cloud; test a limited set of data receipt and management processes (prior to binning and/or ingestion by analytic tools); and perform a test of basic ESSENCE functions. The overall goal of providing new tools on the platform is to improve data quality, efficiency, and usefulness of the data collected as part of the NSSP (Appendix A: Collaborative Pilot Report).

At the conclusion of the collaborative pilot, DHIS conducted an internal decision making process (PDMP - Planning and Decision Making Process) to review advantages and disadvantages of a combination of possible options of new analytic tools and existing tools. A total of five possible courses of action were considered during this process. At the end of this process, the NSSP team in DHIS recommended placing ESSENCE, SAS Tools, and RStudio Professional in the AWS GovCloud environment as the primary tools to support syndromic surveillance and data quality improvement activities. DHIS also recommended discontinuing the development and utilization of the existing BioSense 2.0 front end user web application (Appendix B: NSSP Planning and Decision Making Process). This recommendation was submitted for approval to the Division and Center leadership before being shared for input from the BGG and the wider NSSP community. The recommendations were shared with the BGG during their in-person meeting in June 2015. After detailed discussion, the BGG endorsed the recommendation. It was decided at this meeting that an implementation plan would be created by CDC and shared with the BGG for further input and guidance.

BioSense Platform terminology used:

> **MUB or "Locker"** = A legacy term that refers to a table or view known also as the Meaningful Use Base table or view (MUB), which is a step in the data journey that houses jurisdiction-specific data associated with civilian hospitals, urgent care centers, and/or ambulatory care providers that hold DUAs with the jurisdiction.
>
> **Pre-locker** = Tables, views, and rules applied to decrypt, extract, and post HL7 message segments into tables along the data journey before the Locker or MUB table/view.
>
> **BioSense back end** = Tables, rules, and transformations that occur from the Locker or MUB table/view forward, but before posting data to the user front end web application of BioSense. This includes things like binning, 24 hour rule, 100 mile rule, etc.

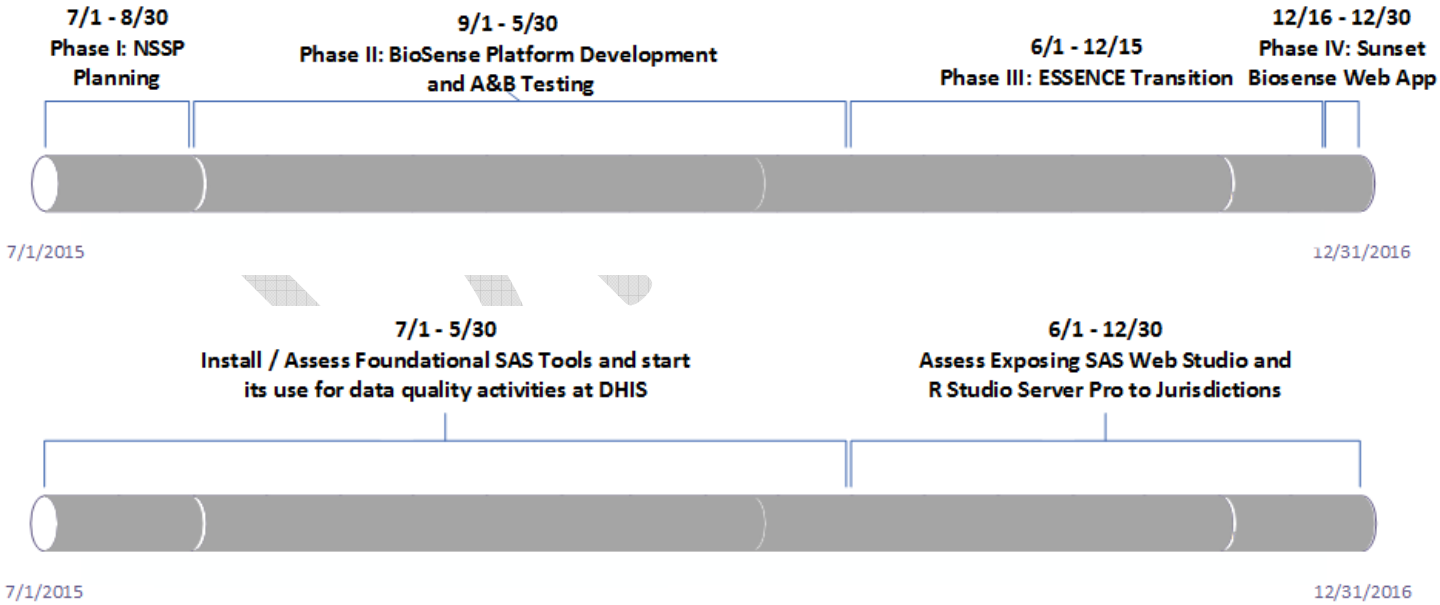
BioSense user front end web application = The analysis, visualization, and data access administration user interfaces that present data and visualizations through an HTML based web application. This is supported by a Mongo Database that receives counts and information from the BioSense back end.

Enhanced data flow = This will be an ongoing and iterative process that will begin during this period through 12/2016 and beyond by leveraging what was learned in the pilot and will be initially focused on getting appropriate data to ESSENCE. Future improvements will build upon this new foundation as we improve and move forward to create a new staging area and explore how to better support jurisdiction access to more raw data, analytic tools and queries, data quality feedback from jurisdictions, etc. It is important to note that we will be using new MIRTH channels in this enhanced data flow. These updated and enhanced processes will begin with minimal adjustments to key fields informed by our experience during the pilot, adding missing fields described in the PHIN messaging guide v2.0, as well as improvements to the facility information. These enhancements are intended to ensure that the ESSENCE application will receive the data expected to be present in a message. Future enhancements will require much more in depth discussion with the community to better describe requirements beyond 12/2016.

Implementation Mission

By December 31st, 2016, DHIS will place RStudio Professional, ESSENCE, and SAS tools in the BioSense platform, an AWS GovCloud environment, expose ESSENCE to all jurisdictions, transition all jurisdictions into the ESSENCE application, assess exposing SAS and RStudio to jurisdictions, and discontinue the use of the BioSense 2.0 front end user web application.

The implementation phases for SAS and ESSENCE are presented in following figures:



Detailed planning timeline and associated activities are included in attachment 1: NSSP Implementation Plan Timeline.

Lessons learned during the pilot as well as experience gained by DHIS and collaborating partners' use of BioSense, indicate a need for a major overhaul of the data work flow. DHIS is recommending that minor changes be made to the data work flow in order to get ESSENCE up and running while, at the same time, the team begins to gather requirements for an

enhanced data flow and data model that will be implemented in later phases. This will allow implementing ESSENCE in the GovCloud in the near-term to ensure continuation of syndromic surveillance operations. No changes will be made to the current BioSense 2.0 front end web application, which will be phased out, or its underlying data flow.

The near-term plan (July 2015-December 2016) includes making minor changes to the data flow feeding ESSENCE that are intended to capture all of the important elements spelled out in the PHIN Messaging Guide v2.0, working with stakeholders to improve facility tables and the basic process for updating the information in these tables, and assisting local jurisdictions with queries that work across new and legacy structures. It is important to note that each jurisdiction's Meaningful Use Base (MUB) table (sometimes called "locker"), legacy facility master, and stage one table archives (part of the Pre-locker data flow) will be maintained so that existing codes/processes set up at the local jurisdiction level can be more easily transitioned. In addition, basic or foundational SAS tools will be assessed and installed for use by DHIS staff for data quality improvement tasks. Additional activities will be conducted to assess the cost and feasibility of making SAS web studio and RStudio professional available in GovCloud environment to all jurisdictions.

Once the tasks from the near-term plan are complete, DHIS will begin to work with stakeholders to map out requirements and a future implementation plan for a more robust data model that can assist both local and central needs for improved data management, quality, and flexibility outside of the web applications like ESSENCE. Future enhancements may include a more relational data model, shared data sets for SAS or R analyses, and an ability for local administrators to feedback data quality changes into their data stream. In addition, as jurisdictions are transitioned to the new tools and enhanced underlying data flows, the current differences and limitations associated with the legacy implementation of PHIN-MS should be minimized.

Phase I – NSSP Planning (1 July – 30 August 2015)

DHIS staff will be engaged in identifying tasks needed to accomplish the mission and develop timelines for each task resulting in an implementation plan. The plan will initially have mid-level detail and be finalized with input from the BGG and syndromic surveillance community. The plan will be used to provide briefings and documentation to CDC leadership. In addition, DHIS will configure all needed servers and install applications into the AWS GovCloud environment. Two environments will be established, a staging environment as well as a production environment.

Phase II – BioSense Platform Development and Alpha/Beta Testing (1 September 2015 – 30 May 2016)

During Phase II, DHIS will work with collaborative pilot jurisdictions to develop and document enhancements to both the processing of data and the environment. Pilot jurisdictions include Alabama, Tennessee, Wisconsin, Virginia, Marion County, IN, Seattle-King County, WA, Michigan, Tri-County Health District, CO (Adams, CO; Arapahoe CO; Douglas, CO), the US Department of Veterans Affairs (VA), and the Department of Defense (DoD). These tasks will be conducted in a staging area without compromising the current BioSense application operations, underscoring the importance of not making changes in the existing data flow. This activity will include, but not be limited to:

(1) Creating a new more robust master facility table and a manual process (for the near-term) for updating facility information
(2) Adding required variables from the PHIN Messaging Guide v 2.0 - table 4.2 to the enhanced data flow and ESSENCE databases
(3) Separating transactional data flows from report generation and data quality assessment activities to be performed by DHIS staff by duplicating tables into a data mart. In Phase IV, this will likely become the basis for a new "locker" or local data store (we will be moving away from the term "locker")

(4) Developing an ESSENCE administration tool interface that will allow jurisdictional administrators to control data sharing
(5) Research Approaches for a Single Sign-on Solution
(6) Adjusting ESSENCE settings as required.
(7) Developing technical assistance documentation

For more detailed description of Phase II activities, please see appendix C: Detailed description of Phase II activities.

Phase III –ESSENCE Transition (1 June – 15 December 2016)

During this Phase, DHIS expects to transition nine Jurisdictions per month starting in June 2016. Multiple teams of ESSENCE transition subject matter experts will be established to assist groups of jurisdictions with the transition operations from BioSense to ESSENCE. As recommended by the BGG, Jurisdictions that use BioSense only as their primary syndromic surveillance system will be the first group to be invited for transition to ESSENCE. Among this group, jurisdictions that are ready and represent the largest group of ED visits available will be first in line. The second tier will include jurisdictions that use other systems/applications in addition to or without use of BioSense and will follow the same order criteria i.e. ready and largest group of ED visits (see appendix D: Jurisdiction Transition Order).

A more detailed description of Phase III activities can be found in Appendix E: Detailed Description of Phase III Activities. Of course, the activities and timelines associated with Phase III operations can change as details are fleshed out and as lessons are learned during Phase II operations. The goal is that by 16 December 2016, all jurisdictions will have been transitioned into the ESSENCE environment and will be functioning properly.

As jurisdictions are transitioned into the ESSENCE environment, DHIS will continue to gather stakeholder input and work towards the future plan. This includes mapping out requirements and an implementation plan for a more robust data model that can assist both local and central needs for improved data management, quality, and flexibility outside of the web applications (such as ESSENCE). Initially some of the enhancements may include (1) a more automated facility administration tool (2) jurisdictional access to RStudio Professional tools and (3) jurisdictional access to SAS Web Studio.

Phase IV - Sunset BioSense Web Application (16 – 30 December 2016)

The last phase of this implementation plan will be to sunset the BioSense 2.0 front end user web application. Before turning off all data flows to the BioSense 2.0 front end user web application, DHIS will ensure that (i) all jurisdictions are properly functioning in the new environment; (ii) historic data have been transitioned to the enhanced local data store in the new environment; (iii) there are strategies for looking across the legacy and new data structures; and (iv) that CDC leadership, the governance group, and the syndromic surveillance community as a whole have an opportunity to voice any final concerns. Once feedback has been assessed, the BioSense 2.0 front end user web application will be shut down on or about 31 December 2016.

Appendix A

Collaborative Pilot Executive Summary

Background

Mandated in the Public Health Security and Bioterrorism Preparedness and Response Act of 2002, BioSense was launched in 2003 to establish an integrated national public health surveillance system for early detection and rapid assessment of potential bioterrorism-related illness. Congress funds BioSense, and, in turn, the Centers for Disease Control and Prevention (CDC) funds state health departments and local health programs and projects through cooperative agreements to support this exchange of syndromic data.

Syndromic surveillance uses near real-time pre-diagnostic data, primarily from emergency departments, and statistical tools to detect and characterize unusual activity for further public health investigation or response. CDC has been a driver in syndromic surveillance nationwide for more than a decade through BioSense.

In 2012, the BioSense Governance Group assessed user software preferences that pointed to ESSENCE and SAS as top contenders for inclusion into the BioSense Platform. In 2013, ASTHO surveyed users regarding desired functionalities for BioSense. And an internal CDC investment review was performed as well, in which defects and limited functionality and capability were cited as issues in the current BioSense Application. All of these efforts led to the need for including highly functional tools into the AWS GovCloud environment.

In December 2014, CDC began the ESSENCE and SAS Collaborative Pilot with eight jurisdictions (Alabama, Tennessee, Wisconsin, Virginia, Marion County, IN, Seattle-King County, WA, Michigan, Tri-County Health District, CO), the US Department of Veterans Affairs (VA), and the Department of Defense (DoD). The scope of the pilot was threefold: (1) evaluate security scans, (2) identify ways to improve pre-locker data processing, and (3) evaluate that the ESSENCE installation works. This Executive Summary describes the second and third parts: Data Processing and ESSENCE functionality.

NSSP Pilot – Data Activity Summary

Overview

The NSSP Pilot provided a ripe opportunity to study the data landscape, end to end data flow, and the processing rules applied as incoming data journey from receipt in the AWS GovCloud to use within the BioSense 2.0 front-end user web based application. To date, there had been a lack of transparency and sharing of information on processing applied to jurisdictions' data feeds. Reversing that model, and openly sharing and disclosing information about the data, was a foundational goal for this pilot. With that, the pilot also provided an opportunity to document and share findings, solicit valuable input from pilot participants, and collaboratively form suggestions around processing changes to share with the larger community for their consideration.

It is important to point out that there was a limited time allotted for the pilot period, and limited documentation describing the data processing supported by the current system. Therefore findings described in this summary should be considered preliminary and subject to change as additional review of the system is considered and supported.

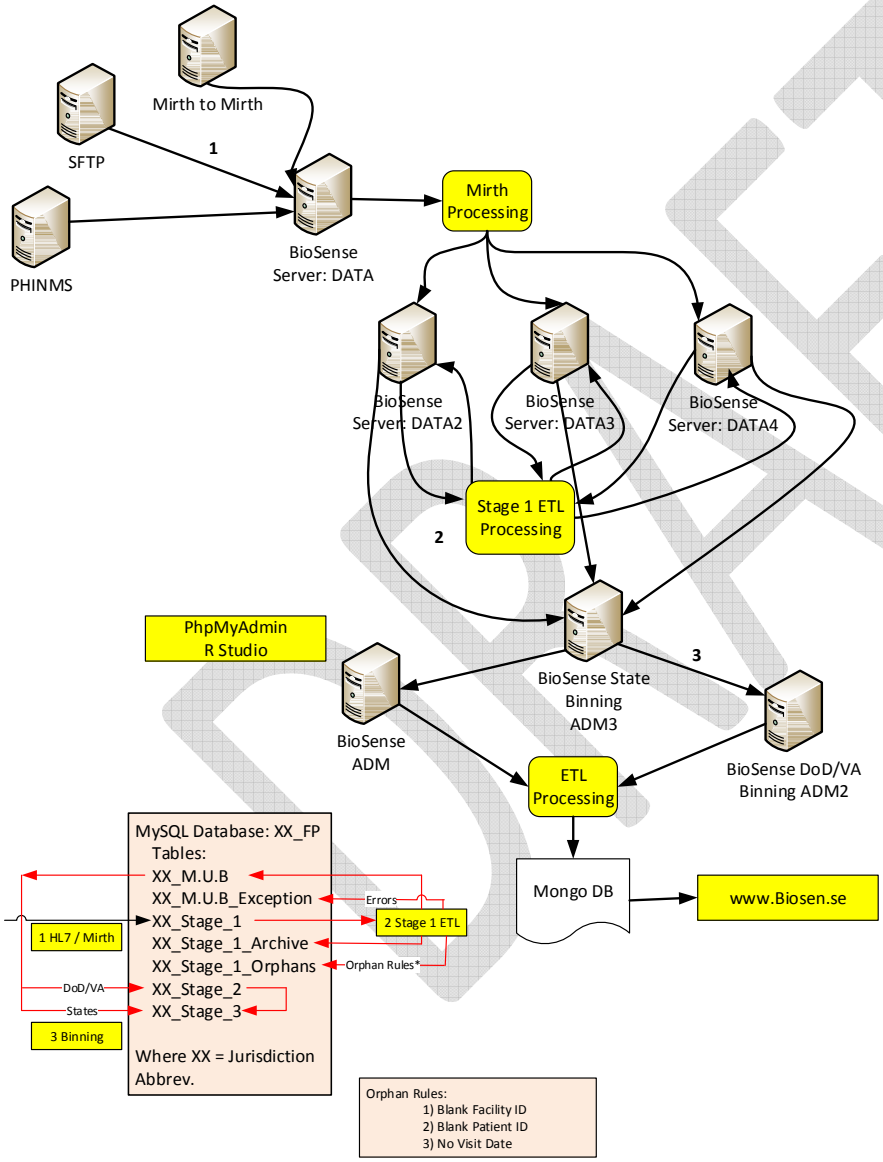
The data review activities applied are as follows:

Data Landscape and Current Data Flow
Document the Processing Rules Applied
Data Element Profiling
Facility Master clean-up
Collapsing of records into a “patient visit event”

Data Landscape and Current Data Flow

Four core processing Steps were identified as illustrated in the data flow diagram in Figure 1.

Figure 1



Step 1: Messages Deposited and Mirth Launched
Step 2a: Data parsed out to Stage 1 Tables, Orphans, and Exceptions
Step 2b: Meaningful Use Base (MUB) Table (the “Locker”) updated from Stage 1 Tables
Step 3: Binning
Step 4: Mongo Data updated for use with the Application

Recommendations: Data Landscape and Current Data Flow

- Provide jurisdictions access to all tables in the "fat pipe" in the current BioSense system. This will require development of additional Views in some cases given not all jurisdictions had Views/access to all tables.
- Work with jurisdictions on reconciling MUB Exceptions (facility update; re-processing of data triaged to exceptions)
- Develop a quality assurance check on records added which have visit dates over 2 weeks ago so as to identify when a "forced" update of the Mongo application data is required.
- Re-architect the "landing pad" to a structure that is more sustainable and flexible. This includes separating out PhinMS data into jurisdiction-specific databases. In addition, this includes de-coupling facility characteristics from analytic data where any change in facility information would be able to be applied across retrospective data.
- Develop data marts for use in ad-hoc analyses to optimize runs and reduce "conflicts" with transactional processing

Document the Processing Rules Applied

We performed a crosswalk of how Stage 1 Meaningful Use (MU) variables align with the variables in the Meaningful Use Base (MUB) table, studying documentation and underlying processes. A draft "crosswalk" document was developed and includes information on the HL-7 segments and data fields that are associated with Stage 1 columns, as well as how Stage 1 data is mapped to MUB. The crosswalk includes information on rules applied during the Stage 1 to MUB process as well as information on calculated fields (and algorithms used in the calculations).

The crosswalk exercise also identified elements of interest that were in Stage 1 but not in MUB, as well as elements that were in the Messaging Guide but not processed into Stage 1.

See below for categories of data elements included in the crosswalk documentation:

MUBTRANS Key	
1	Direct Transfer From Stage_1_Archive to MUB without modification
2	Stage_1 Column To Mub Column Transfer but with a Type Change ;i.e., MSH_7_1_Message_Date_Time is character in Stage_1 but Date_Time in the MUB.
3	Mub Column is sourced from multiple columns in Stage_1
4	Mub Column Only with no corresponding value in Stage_1
5	Element of Interest in the Phin Messaging Guide 2.0 but not in the MUB
6	Exists in Stage_1 but is not carried over to the MUB
7	May have been included in past versions of the MUB but does not exist now.

Recommendations: Document the Processing Rules

- Verify content of draft crosswalk document
- Reformat crosswalk for end user consumption including a "bottom line" section for quick reference
- Clear crosswalk and share with community
- Ensure all data elements are processed and stored where appropriate

Data Element Profiling

Data profiling was performed in order to confirm our current understanding of the processing rules being applied and provide information on content of Orphans and Exceptions Tables. The profile reports provided pilot participants with insight on how the Stage 1 and MUB data relate and the mix of HL-7 data elements that were used to populate Stage 1 and MUB. The reports also provided an opportunity to spark discussion around potential rule changes to suggest to the larger community. In addition, the reports included information on the records that were triaged to the Orphans holding tank or the Exceptions holding tank.

Recommendations: Data Profiling Results

It is important to acknowledge that the profile reports were based on a static cut of data from around November 2014. The contents of the jurisdictions' reports often revealed a mix of elements being populated or not populated. The group acknowledged that a number of processing updates were made by the legacy contractor around August 2014 after which incoming HL7 data were landed and stored in the Stage 1 table. However, retrospective data were not processed. Additional fields that were processed starting August 2014 include EVN7 related fields and PV1-19 visit ID. The group acknowledged and recommended that profiling of more recent data would more accurately reflect the content of current incoming data which is being landed and stored in Stage 1.

Understanding that we may not have seen the full more-recent picture of data in the data profile reports, the group did provide suggestions for processing changes. Please see below for highlights of recommendations on rule changes per each priority data element:

Unique Visit ID

Use the Patient Visit ID (PV1-19.1) as described in the messaging guide. Pilot participants acknowledged that additional data may be needed to form a unique key. Participants also acknowledged that further study is needed to assess where, why, and how a patient's initial-visit person-event ID code might change, perhaps multiple times, as the patient is seen in different departments or at different facilities within the hospital system. Some participating pilot jurisdictions were confident that the ID code would stay the same at the hospitals within their jurisdictions. Other jurisdictions were uncertain whether in such situations a patient's ID code at their hospitals would remain unchanged.

The pilot group also acknowledged the possible need to support multiple ID code methods. This is because no universal all-situation patient ID coding method might exist, or even be practical, across multiple hospital systems. While such lack of patient-identification universality is a possibility, addressing this and similar issues must be balanced against the hazard of offering more choice options than system resources can realistically support. Again, further investigation is needed to explore models that would work across multiple hospital systems. Also needing examination are a small number of models that would accommodate the various flavors of incoming data.

Facility ID

Use EVN7.2 to seed the Facility ID.

Medical Record ID

Use PID-3.1 if the accompanying identifier type in PID-3.5 indicates the identifier is a medical record (MR). Do not try to change the incoming data to insure MR is always sent as the first PID-3 field. This is unreliable.

Chief Complaint

Exclude Triage Notes when building the Chief Complaint concatenated field.

Chief Complaint/Diagnosis/Procedure Storage approach

Some pilot-test participants suggested that the table structure to store chief complaint and other health-indicator related data (Diagnosis) be adjusted to make discernable the HL7 data source segment. Suggestions included a pivoted version of the concatenated data with indicators as to type of data and HL7 source. Such a version would complement the concatenated style and would not replace it. Note: adopting a pivoted-type style would also resolve problems of concatenated data becoming truncated due to field length limitations.

Age/Age Units

Some study participants urged that a patient's age be universally calculated on the basis of birthdate. Their justification was that a patient's age as identified in the message sometimes was the current real-time age instead of the patient's age at the time of their visit. However, other participants noted that birthdate may not be sent and that this approach may not work for others. The study participants recommended further study of age related data including review of the differences among age calculated based on birthdate and visit data, calculated age sent in incoming OBX segments, and reported age sent in incoming OBX segments. If we find that calculated age is more accurate than reported age then we will go with that one.

Death Indicator

Study participants noted that while PID-30.1 is required, we need to remember that it is not required to note "No." Hence the reason for the large volume of missing death indicators. The participants suggested other data be used to determine mortality status and not depend on PID-30 alone. Recommendations included the use of date-of-death as an indicator of death, the Chief Complaint or Diagnosis Text of "DOA" (Dead on arrival); and/or the disposition. The group recommended we determine what is available in data elements that may assist in determining death status.

Facility Master clean-up

The DHIS staff reviewed facility-related data in support of the pilot's data-profiling activity. This effort led to a clean-up of facility master tables (from pilot jurisdictions) in preparation for use of data in Essence.

Recommendations: Facility Master clean-up

Given the data used during the pilot reflected a static cut of data up to November 2014, more recent data from the pilot jurisdictions should be used to determine if there are any other differences/issues so as to close the loop with pilot sites. In addition, a similar facility clean-up process should be launched across all jurisdictions.

The clean-up process is to include enhancing facility master tables to capture "metadata" about facilities including NSSP-sending entity, parent facility, facility type, open/close/changed status, date added, date ended, electronic medical record system associated with the facility feed. The onboarding teams should work closely with other team members to ensure official registration of facilities in master tables, with all required metadata completed during the registration.

Collapsing of records into a "patient visit event"

ESSENCE uses the last record and overwrites information. Many jurisdictions see the chief complaint degrade overtime – starting with rich text and resolving to some drop down value. Further, although the A^08 messages are supposed to include the same data previously sent if unchanged, some of the pilot sites' experiences suggest that A^08 updates may include NULL data in segments that had non-NULL data sent through in previously issued messages. The ESSENCE method may not be the best solution.

The group discussed approaches to take, if any, to collapse information found across records associated with the same patient visit event. The group acknowledged that this topic is tethered to the Unique Visit ID. (Unique Visit ID is further discussed in the Data Profiling section.) It was important to make sure we were talking about the same thing when using the term collapsing. The collapsing of records may tie to the following scenarios:

- **Transactional updates**: updates are sent for the same patient visit event. For example, Aˆ08 updates are sent for the same event previously sent in Aˆ01 messages, where information supporting "a unique visit ID" may or may not be the same across messages.

- **"Roll-up" of data:** Combining data across different messages (with different "unique visit ID") that are associated with the same patient visit event. For example, the same person-visit scenario could possibly have

different visit IDs if the person went to different facilities within the hospital system, or went to different departments within the same facility. In both examples, facility ID changes could influence the Unique Visit ID "key" if facility ID is used as part of the key. The Unique Visit ID would be different in these examples and thus, the patient's visit event data would not be able to be "connected" via Unique Visit ID. In addition, a patient-visit event may continue over time such that incoming data may have different dates of events. This could potentially influence the Unique Visit ID "key" if decisions around Unique Visit ID algorithms included use of date data.

- *Note: In both scenarios, decisions around reconciling differences found in foundational data (e.g., demographic data) would have to be made as far as the "final" data used.*

Recommendations: Collapsing of records

In general, participants acknowledged this is a complicated topic that requires additional data review across jurisdictions given the differences in experiences cited by the pilot participants. Findings will help the community make more informed decisions around the area of record collapsing.

Data review recommendations are:

Transactional-related

- Check if duplicate Chief Complaints are sent over time.
- Compare "first and last" records and review differences in data first sent vs. data last sent (for the same Unique Visit ID).
- Report out by jurisdiction, facility, and "time."

Roll-up- related

- Assess frequency of "different Unique Visit IDs" that have "close enough connection" to suggest that the different events are really associated with the same patient event. Include chief complaint/diagnosis data in the review to better assess if the separate "Events" are really tied to the same situations.
- Use results to determine how often this happens and whether it is necessary to put effort into this more challenging connecting the dots and then collapsing the data.

Issues related to different treatment settings and the interactions between Treating Facility, Facility Type, Visit Type, and Patient Class. This is critical when trying to determine admission status or how to properly segregate visit types for analysis. This is an issue that we touched on during the pilot, but for which we really didn't come close to developing recommendations.

ESSENCE Evaluation Summary

A collaborative team was assembled to evaluate ESSENCE. Team members reviewed program documents and then met to clarify the goals of the ESSENCE evaluation, including how ESSENCE was being used, and by whom. Once this preliminary work was done, they identified five components of ESSENCE to evaluate: (1) functionality, (2) data sharing and permissions, (3) usability, (4) timeliness, and (5) user satisfaction. Next, they designed an online survey to evaluate each component. Ultimately, they needed to know whether ESSENCE, a surveillance system traditionally used for state-level data analysis, could handle the demands of gathering data from sources nationwide and perform as intended on the BioSense platform. They also wanted to know if the system would meet users' needs, be easy to use, and address complaints and improve the current system.

Eight jurisdictions volunteered to evaluate ESSENCE, individual participants self-identified as novice, intermediate, or advanced users. Users were given access to a test environment in which a static cloned database was used to feed ESSENCE data flows.[1] Permissions were set on the basis of jurisdiction in which users wanted to share data. Users were assigned real-world tasks that would prompt them to test system features. As the users performed each task, they were required to take a brief online survey that recorded their answers. Interviews were conducted after completion of the survey to gain insight into each user's overall experience.

Strengths

- **ESSENCE will meet users' needs (7 of 8 users)**. Users reported that ESSENCE would meet their needs for analyzing syndromic surveillance data. Further, they noted that ESSENCE would enhance the BioSense platform, enable customization and display elements of data quality, and increase analytic capabilities and data visualization. Supporting quotes: "*It is highly versatile, especially this newer version. I can do so many things that are relevant to the workflows I might want to adopt when doing syndromic surveillance investigations."*

- **Users had a good experience with the system (6 of 8 users)**. Responses were generally positive about ESSENCE, and users had a good experience. Supporting quotes: "*Very positive. Can't wait to use. I wish I had more time to work with the system." "A lot of functionality and customizability." "Based on my experience, it was incredibly quick for doing anything I wanted. This is very important to me."*

- **ESSENCE is a flexible system with features that enable users to streamline data workflows, customize data views, and share analyses (6 of 8 users).** Users especially liked ESSENCE's built-in data views and their ability to share analyses and customize the system and dashboards. Supporting quotes: "*I like that I can do most of the standard descriptive epidemiology using the ESSENCE interface, that it has functionality that tells me where there are observed differences from a baseline count so that I have a starting point for evaluating whether there's any public health significance." "I like the options for streamlining my workflow (myESSENCE, myAlerts, and Report Manager) so that I don't have to click around the user interface a lot to get what I need. The 'dashboard' functionality in myESSENCE is really helpful, and I like that I can share views of data within myESSENCE, myAlerts, and queries. Having built-in data quality views is also great."*

- **ESSENCE will address the enhancements requested by the user community (2 of 8 users).** Users reported that this tool would address the complaints and enhancement requests made previously and enhances the BioSense platform. Supporting quotes: "*I think having this in the NSSP cloud is a huge improvement on current state and will allow for efficient analysis of national data." "It is a good tool that will address many of the complaints and system enhancement requests that users have made and that have been noted in the functional requirements gathering process."*

[1] Test parameters were preset, depending on the jurisdiction in which the user wanted to share data.

Weaknesses

- **Users cannot control data sharing (5 of 8 users).** Some users reported that ESSENCE does not let them control the level of data they want to share. Supporting quotes: "*I can't figure out how to do this." "Instructions are not clear on how to select variables/fields to share*."

- **Data sharing and data source identification were confusing and not intuitive (4 of 8 users).** Users were confused about what they should have seen in their data views. One user noted not being able to see aggregate-level data for a jurisdiction within which data were being sharing. Users also were unable to understand what the labels for different data sources meant (e.g., ER by Patient Location and ER by Hospital Location). Supporting quotes: "*Michigan resident visits to CO, IN, TN, WI facilities downloaded at line level. Unclear if this is desired." "To view the hospital data, it required me to change the data source to ER Data by Patient Location. It took a while for me to figure this out. It would be ideal if this was the default data source for hospital accounts*."

- **Potentially steep learning curve for new and intermediate users (4 of 8 users).** Although ESSENCE received a good usability score (70.5), novices had difficulty understanding how to do many required tasks (i.e., creating and editing queries), whereas experienced users found the tool intuitive. Some novices reported that certain aspects of the tool are not intuitive or user-friendly. Intermediate users noted difficulty with certain tasks, such as creating a bookmark, but accomplished the task after consulting the training materials. Novices, however, found the training materials outdated and not helpful. Supporting quotes: "*Not user-friendly, and resources needed to support users are lacking." "There is a bit of a learning curve to understand where everything is*." "*Were it not for the training materials (and mimicking them), I don't think I would have been able to figure out how to do most of these things." "Steep learning curve. 101 and 201 probably required for new users."*

- **User interface does not explicitly tell the user what to do next, and some features are not easily navigable (3 of 8 users).** Users reported that the system lacked feedback. For example, they were unsure if a feature was working correctly because nothing, such as an hourglass, hinted that processing was ongoing. Users also were unsure if a query was actually being shared. They noted the tool's inability to quickly filter certain variables such as county or facility without scrolling through an entire list. Users suggested graying out data sources that are not selectable or having a text box to filter. Supporting quotes: "*Overview by region query took a long time to load. It would be nice to have something to let the user know they did something wrong*." "*For many tasks, a time-consuming trial and error process was involved. Even when a task was completed, I was often not sure whether the methods I used were ideal or the most efficient way to complete the task."* "*Some features seemed to be disabled, but it was hard to tell since there isn't a lot of feedback when the user makes an error versus when the system can't support the request. For instance, when you go to the Overview portal and try to get an overview by facility, ESSENCE returns a blank page*."

- **Some features are not functional.** Users reported the myAlerts and spatial alerts features did not work. Supporting quotes: "*No spatial alerts generated." "Couldn't view spatial alerts lists." "I couldn't do it*."

- **Users experienced long wait times for features intrinsic to data processing (2 of 8 users).** Users noted lag time or slow processing for some features (e.g., selecting facility as an overview parameter, downloading graphs and tables). Supporting quotes: "*Download all selected graph data tables by MS Excel action took a long time*."

Recommendations[2]

1. Provide comprehensive training (i.e., each level of implementation, all functionality) so that users understand how to use the system effectively, familiarize themselves with the data sources, and know how these data will be displayed (screen views). Possibly create a buddy system pairing new users and novices with intermediate and advance users.

2. Update the system operation documentation at the comprehension level of a new user.

3. Enable users to control administrative access (i.e., set their own permissions) for viewing and sharing data.

4. Test and validate that features are available and function properly.

5. Monitor and evaluate ESSENCE's performance when many users access the system simultaneously and download large amounts of data.

6. Modify select aspects of the user interface to improve the user experience (e.g., gray out non-selectable options, notify user when queries are shared, notify user when a task is processing, add a text box to filter variables).

[2] CDC will retain BioSense until ESSENCE is launched. These recommendations cannot be implemented until data are cleaned, a local administration tool is developed, agreements for data sharing are developed, and technical assistance for transitioning jurisdiction is provided. Once a decision is reached, DHIS staff will work with the BioSense Governance Group and NSSP CoP members on plans, timelines, and target dates for implementation.

Appendix B

NSSP Planning and Decision Making Process

At the completion of the Collaborative Pilot, DHIS staff were given the following planning task "After completing the ESSENCE/SAS Pilot on May 15, 2015, DHIS NSSP will make a recommendation no later than May 30, 2015 regarding which tools in the AWS GovCloud will be utilized to do data quality, ad-hoc analysis, and traditional syndromic surveillance". In order for the staff to provide a well thought out recommendation to CDC leadership, they used a structured Planning and Decision Making Process (PDMP) with representatives across all branches within DHIS.

During the eight hour PDMP session, the staff discussed and documented all facts and assumptions related to the planning task. Once complete, the staff developed five possible courses of action (COA's) that could meet the requirements of the planning task. The COA's developed were:

COA 1: Enhance BioSense and include RStudio Professional, PHPMyAdmin and MySQL Enterprise.

COA 2: Enhance BioSense and include RStudio Professional, SAS Products, PHPMyAdmin and MySQL Enterprise.

COA 3: Enhance BioSense and include RStudio Professional, SAS products, ESSENCE, PHPMyAdmin and MySQL Enterprise.

COA 4: Discontinue BioSense and include RStudio Professional and place ESSENCE in production.

COA 5: Discontinue BioSense and include RStudio Professional and place ESSENCE and SAS products into Production.

The staff then documented the advantages and disadvantages associated with each COA and developed criteria used for the COA analysis matrix. The subjective and objective criteria established were (1) Cost (2) Community of Practice (3) Leadership expectations (4) Complexity (5) Staffing (6) Analytic Capacity. At the completion of the process, the staff recommended COA 5 and received approval from the Division and Center Leadership. The recommendations were then shared with the Governance Group and received their endorsement.

Appendix C

Detailed description of Phase II activities

During phase II, great care will be taken to integrate ESSENCE into the BioSense platform with a primary goal of not compromising current BioSense data flows and operations. Phase II will focus on the ESSENCE transition.

Key Implementation Tasks

The initial deployment of ESSENCE into the BioSense platform requires the successful completion of a number of key items including development of key infrastructure, data flows, supporting interfaces, and the analysis and cleaning of existing BioSense facility data with input from the jurisdictions.

- Design and develop the new master facility table, site profile, and supporting tables, data flows, and processes to enable data loading into ESSENCE.
- Adding required variables from the PHIN Messaging Guide v 2.0 - table 4.2 to the enhanced data flow and ESSENCE databases
- Separate data analysis and reporting data sources from transactional (production) data sources via the stand-up of a new Microsoft SQL database server which will house duplicate copies of the data marts currently used (e.g., the Locker). In phase III DHIS staff will work to provide users access to upgraded analysis tools (e.g., R Professional; SAS Studio).
- Design and develop an ESSENCE Jurisdiction Administration application that will enable the jurisdiction administrators to control access and sharing of their data.
- Research Approaches for a new Single Sign-on portal with two factor authentication to enable users to securely access ESSENCE and other BioSense platform applications.
- Adjusting ESSENCE settings as required.
- Developing technical assistance documentation

Assumptions

- Development resources will be directed towards implementing ESSENCE. Prioritized legacy BioSense issues will be reviewed on a case by case basis.
- The jurisdictions will be available to assist with the analysis and adjudication of their facility data and provide guidance and direction regarding data to be loaded into ESSENCE.

Facility Clean-up Process

Integrating ESSENCE into the BioSense platform will require a review and cleaning of the existing BioSense facility data. This review is driven by structural differences at the database and application level between BioSense and ESSENCE.

The ESSENCE ingestion process requires use of a facility master table containing unique identifiers for each facility. While a number of sites have facility masters that meet this requirement, the pilot experience revealed that some sites' facility masters contain duplicate facility IDs referencing the same facility. Related, we discovered that the processed data also contained duplicate facility identifiers (FacilityID_UID) for the same facility. In addition, we found that the characteristics of the same facility differed in some cases as the facility data were compared across various data sources and views (e.g., slightly different names; address missing vs. non-missing,; spelling differences in city; zip code differences including blank vs. 5 digit vs. 5 digit plus 4 digit zip extension).

The differences are likely due to facility master changes over time, where updates to identifiers and characteristics were inherited in data that were prospectively processed, but not reflected in retrospective data that had been previously processed. In addition, the current BioSense receives and processes incoming data into one of two major MySQL database structures, jurisdictional sFTP databases (aka. "Fat Pipe") and PHIN-MS. These two source databases have their own facility tables and are not integrated. This non-integration between facility table entries introduces potential duplicates and differences in facility characteristics for those sites that either at one time transmitted data through PHIN-MS and then moved to transmitting through sFTP, or, for those sites that continue to transmit data through both PHIN-MS and sFTP. Essentially the dual facility masters fell out of synch over time resulting in facility-related discrepancies in the data.

BioSense was built in such a manner that this disconnect did not preclude downstream processing of data (although the data processed will of course reflect the mix of facility related data). In contrast ESSENCE was designed assuming each facility exists with only one record and each facility must be unique include facility name, address, zip code etc. This difference advances the need to support the long standing requirement to standardize and improve the facility data.

To support the ESSENCE requirement for a single unique database entry per facility, the NSSP team will work with sites to review existing facility masters and adjudicate duplicates and/or discrepancies found in the current master files such that we ultimately have one identifier selected as the winning identifier to use moving forward. In an effort to assist with this decision making, the NSSP team will provide information on volume (count of records associated with a specific ID) and currency (timeframe of maximum message date associated with a specific ID). Based on volume and currency, the team will generate a first cut proposal for jurisdiction consideration as to which of the multiple identifiers to use as the official winning identifier from this point forward as part of the ESSENCE process.

The winning ID will be selected from the multiple FacilityID_UUID values based on the following rules:

- The facility ID with the larger volume would win over another facility ID with lower volume
- A facility ID with more current message date values would win over another facility ID with earlier maximum message dates, regardless of volume

Note: The approach taken is a best guess for jurisdiction consideration as a starting point to resolve duplicates and inconsistencies found in facility-related data.

Depending on the site, the existing BioSense data may contain a mix of identifiers due to the above noted duplication. There may be some cases where current incoming feeds also have a mix of identifiers. In an effort to standardize the data in ESSENCE, the NSSP team will provide a mapping of duplicate facility IDs to the winning facility ID . The mapping information will be used in the ESSENCE ingestion process to map and adjust an incoming identifier, perhaps now considered deprecated, to the current winning facility ID.

The important facility master cleaning exercise presents an opportunity to introduce a new master facility table to house the adjudicated and clean facility master data. The new master facility table will contain all of the BioSense facility table columns plus additional columns to capture desired data elements to support processing and analysis. Where appropriate, the columns will support controlled entry such that standard codes and labels are used to minimize the variation of related information found in current master files.

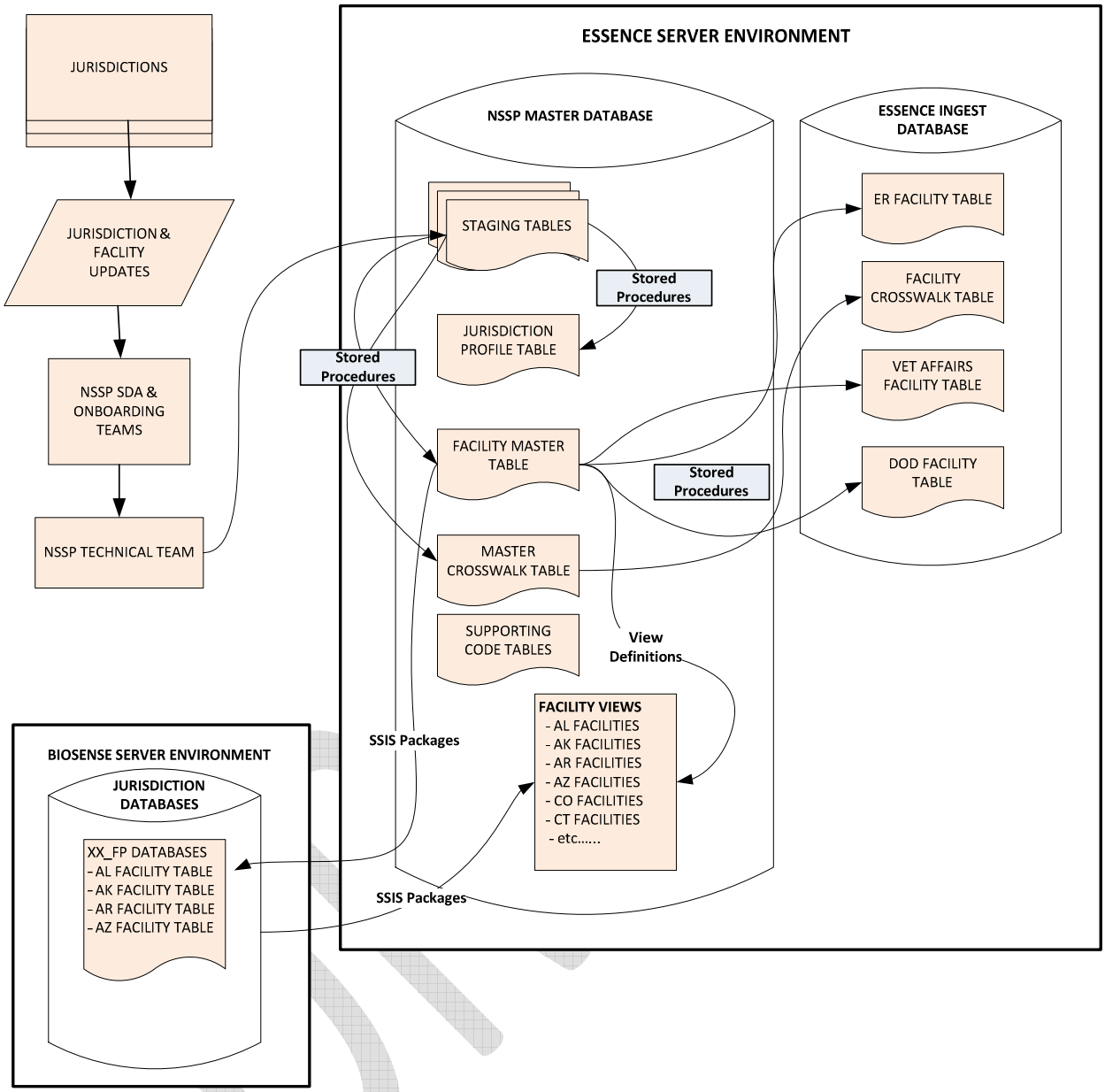
Facility Data and Flow Design

The central feature of the ESSENCE transition is a new "Facility Master" data table that will serve as the solution's primary reference source for facility data. Information from this table will then be deployed to the various parts of the BioSense platform where it is needed to support operations. The new master facility table will exist not only to support ESSENCE data ingestion but will also function as a central database to be used to keep the existing jurisdictional facility tables synchronized and ensure data quality.

Figure 1 is a summary of the existing and proposed future data flow and processes. The NSSP Data Analysts and Onboarding teams will be engaging the jurisdictions to review and organize the jurisdiction's facility data before loading the clean facility data into the new master facility table. In addition to the new master facility table there will be additional tables created to support processing and improve data quality including:

- Jurisdiction Profile table to capture details about each jurisdiction and their data profiles. Key use of this table will be to assign a Site Id which will be used as a key field in the database and used to support new ESSENCE functionality for data access, sharing and ownership.
- Facility Crosswalk table to be leveraged when ingesting pre-existing BioSense data that may contain duplicate facility identifiers for the same facility. The Facility Crosswalk will be use to map incoming identifiers to a single master facility Id.
- Facility Staging tables to store raw facility data that will be uploaded to the BioSense Platform before being validated and written to the master facility, facility crosswalk and ESSENCE facility tables.

Figure 1 - Facility Data Processes and Tables



Data Flows and Processing PHIN Message Guide Variables

Integrating ESSENCE into the BioSense Platform will require the development of enhanced data flows (Figure 2). These enhanced data flows will leverage existing BioSense data feeds from the jurisdictions without interrupting current BioSense data processing. The first iteration of the enhanced data flows will include all existing BioSense variables and will also include the additional variables, currently uncaptured in BioSense, as documented in the PHIN Syndromic Surveillance Messaging Guide v2.0 table 4.2. Data will be extracted and stored leveraging the existing BioSense servers and then fed into ESSENCE. In addition, consideration will be given to a new HL7 archive that could be leveraged in phase IV and beyond in order to ensure that quality data flows into ESSENCE while also storing the raw data as received (Figure 3). This enhancement will be a more relational data model to collect and store the raw HL7 message in an archive that preserves the integrity of the message as it was transmitted. Designing this relational data model will require intensive

analysis and input from the community before development can begin and when implemented will replace many of the existing BioSense servers. Copies of the raw HL7 messages from the archive will be made available to the jurisdictions for analysis in the enhanced data store.

Figure 2 - BioSense/ESSENCE Server Environment

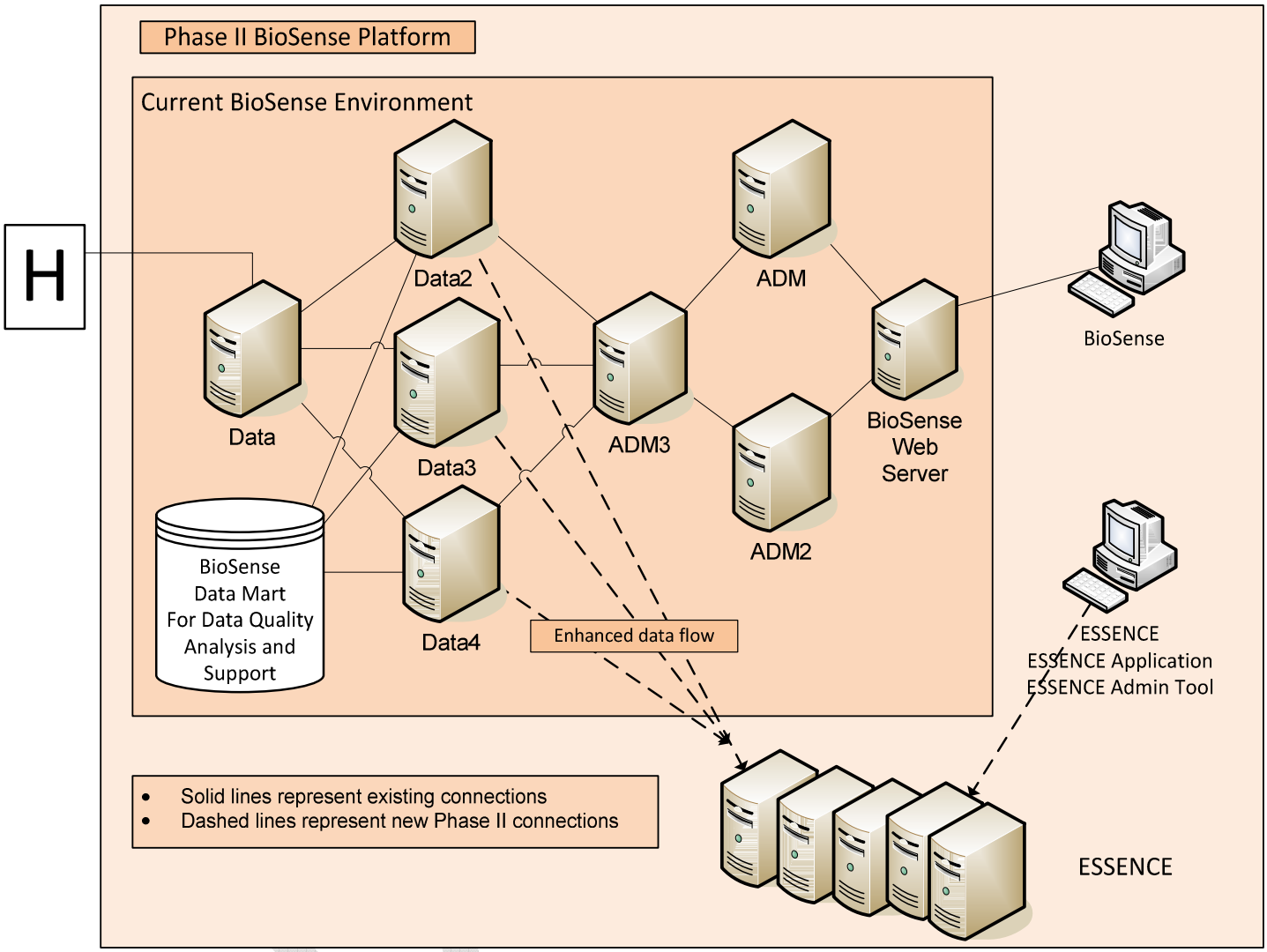
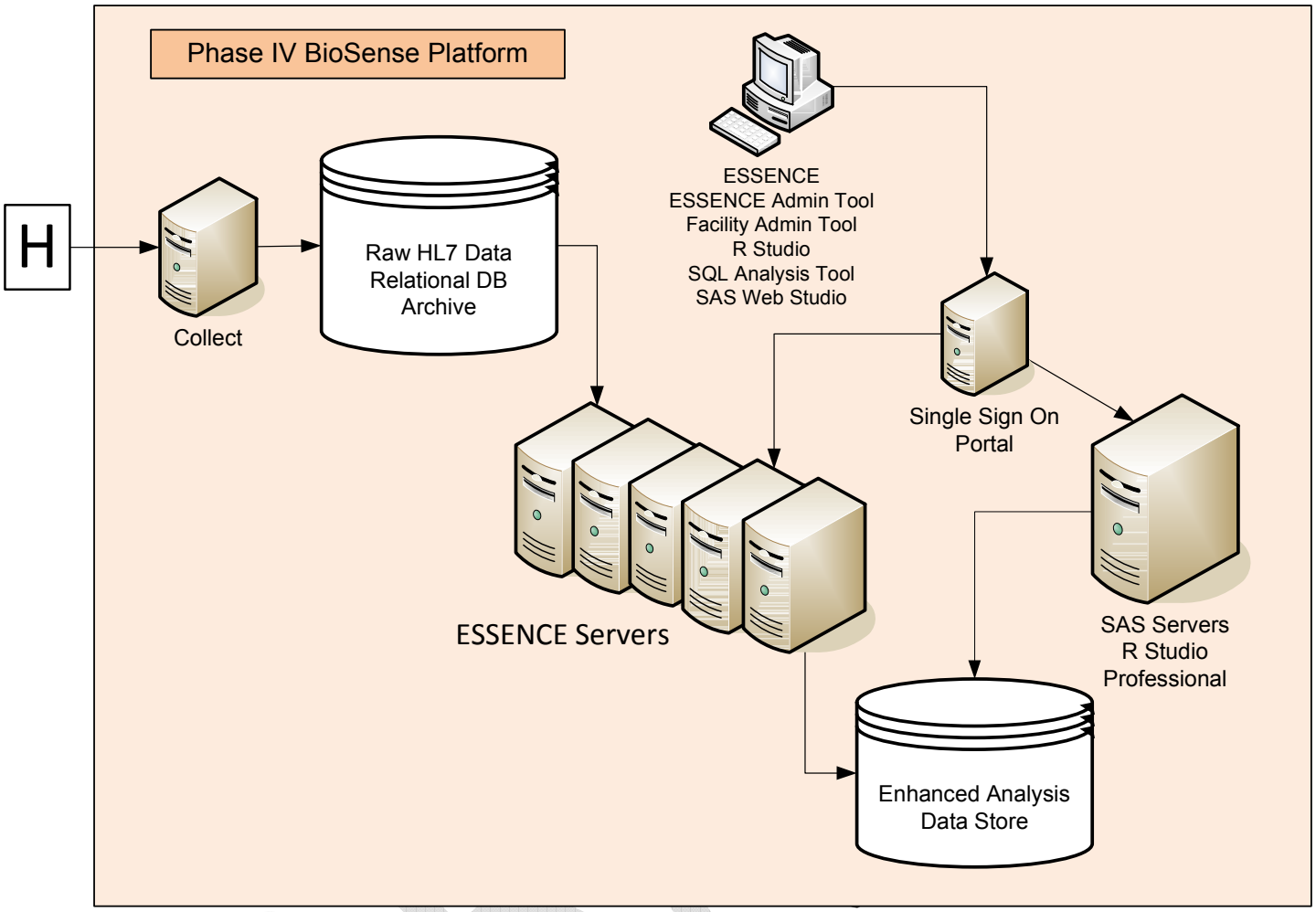


Figure 3 – Phase IV BioSense Platform



Separate Transactional Data for Data Analysis and New Analysis Tools

The existing BioSense data analysis tools (RStudio and phpMyAdmin) have been implemented in the current BioSense architecture such that users utilizing the tools connect to MySQL views to analyze their data. When a user attempts to analyze their data in the current BioSense, their queries place a database lock on the production data sometimes resulting in a delay in the processing of incoming data. This database lock can freeze a table for multiple days, prevent data processing for a jurisdiction, and require manual monitoring of the current BioSense on a daily basis to ensure timely data processing. Additionally the RStudio and phpMyAdmin tools are currently out of date and do not provide the data analysis abilities needed by users.

In order to better support data analysis, enable timely data processing and provide better analysis tools, phase II of the BioSense platform development will include the creation of a new local data store. In phase II, DHIS will be working with SAS tools and the enhanced data store to support data quality activities. During Phase III, the new local data store, hosted on a Microsoft SQL Database will provide users with access to their existing Production BioSense data (copies of the MUB and BinLocker) as well as access to views of their facility table and the expanded PHIN Messaging Guide v2.0 – table 4.2 based data. Other data analysis databases and views will also be designed to allow users to access ESSENCE data elements. Secure access to the enhanced data store with new tools including SAS Web Studio, RStudio Professional and a

replacement for phpMyAdmin will be developed. During the transition period to the new local data store and tools, access to all the existing tools and databases will be maintained for the user community.

ESSENCE Jurisdiction Administration Tool

To support a national ESSENCE implementation for security, data access, user account administration and data sharing a new web based ESSENCE Jurisdiction Administration tool will be created. Access to the new Jurisdiction Admin tool will be limited to designated jurisdiction administrators and their alternates through ActiveDirectory permission schemes built with the single-sign-on portal in mind. This admin tool will enable administrators to perform all needed user account administration tasks including but not limited to account creation, deletion and password tasks. The admin tool will also provide the administrator the tools required to control ESSENCE data access for their staff as well as control data sharing with other individuals or entire jurisdictions. The ESSENCE jurisdiction administration tool will also include a data sharing/access reporting function and email notification for specific functions.

Research Single Sign-On Approaches

A new NSSP web portal is planned as a one stop shop for NSSP and the broader NSSP CoP. During this implementation plan, research will be done on approaches for a Single Sign-on solution to provide the BioSense Platform user community a single URL for syndromic surveillance resources and to provide a single access to all of the applications hosted in the BioSense platform. Eventually users will be able to log onto the portal with a secure 2 factor authorization protocol, and once authenticated will be able to access all of the applications and tools hosted for which they are authorized. This portal will replace the existing multiple URLs, user names and passwords users must currently maintain. The NSSP portal will also have public spaces for resources, announcements and other features desired by the user community and in line with the governance group communications workgroup plans.

ESSENCE Settings

The NSSP team will work with ESSENCE's JHPL staff to implement a new Site Id variable and adjust ESSENCE settings to support the BioSense platform ESSENCE transition. The concept and development of Site Id is important to address existing deficiencies in current BioSense with regards to jurisdictions, data and facility ownership, reporting and data segregation and security. Each jurisdiction (federal, state, county, local as appropriate) will be assigned a Site Id and the Site Id will be associated with all of the facilities that jurisdiction has a relationship with regardless of physical location. This means that facilities can be tied to jurisdictions even if the facility resides in different state or county than the jurisdiction (which can be appropriate and happens quite regularly). Data access and data sharing will be keyed to the new Site Id which will also be associated with all of a jurisdiction's staff or users. ESSENCE data viewing, data mapping and reporting will also be tied to the Site Id and the Site Id concept will be propagated at the database level. The CDC NSSP team will also work with JHPL staff to implement changes to default ESSENCE settings to control data element displays and drill down functionality; work with jurisdiction administrators to identify desired data access and sharing settings and implement highly valued improvements to the application as resources allow.

ESSENCE Technical and User Documentation

The NSSP team will work with JHPL to develop technical documentation for the ESSENCE data flow, processing and technical details to support and answer questions from the user community as resources allow. The program will also work to create documentation to enable training and support for the new ESSENCE Jurisdiction Administration tool and other new tools as they are deployed onto the BioSense platform. End user guides and non-technical training documentation will also be generated to provide a complete set of documentation to support the user community. All of these resources will be made available on the NSSP portal when available.

Appendix D

Jurisdictional Transition Order

June		
Status	**Jurisdiction**	**ED Size estimate**
BioSense Only	Illinois (includes Cook)	3894098
BioSense Only	Massachusetts	2644442
BioSense Only	Kentucky	2411060
BioSense Only	Arizona	2357660
BioSense Only	Mississippi	1770436
BioSense Only	Arkansas	1392101
BioSense Only	West Virginia	1234689
BioSense Only	Kansas	1133050
BioSense Only	Houston, TX	998772

July		
Status	**Jurisdiction**	**ED Size estimate**
BioSense Only	Nevada	909949
BioSense Only	Utah	895150
BioSense Only	New Mexico	839473
BioSense Only	Denver Public Health	736064
BioSense Only	Riverside, CA	609569
BioSense Only	Idaho	554512
BioSense Only	North Dakota	424789
BioSense Only	Montana	371086
BioSense Only	Alaska	290711

August		
Status	**Jurisdiction**	**ED Size estimate**
BioSense Only	Stanislaus, CA	230410
BioSense Only	Linn County, IA	100287
BioSense Only	Santa Clara, CA	63596
BioSense Only	Nevada, CA	51391
Non BioSense	Florida	8194073
Non BioSense	Ohio	6660935
Non BioSense	Pennsylvania	6276403
Non BioSense	New York	4706521
Non BioSense	North Carolina	4644060

September		
Status	**Jurisdiction**	**ED Size estimate**
Non BioSense	Georgia	4216071
Non BioSense	New York City	3713757
Non BioSense	New Jersey	3705912
Non BioSense	Indiana	3434886
Non BioSense	Tarrant County TX	3309208
Non BioSense	Missouri	3051606
Non BioSense	Louisiana	2583184
Non BioSense	Maryland	2552154
Non BioSense	Washington	2405502

October		
Status	**Jurisdiction**	**ED Size estimate**
Non BioSense	Oklahoma	1959867
Non BioSense	Minnesota	1898936
Non BioSense	Connecticut	1643476
Non BioSense	South Carolina	1451782
Non BioSense	Oregon	1403126
Non BioSense	Maine	793797
Non BioSense	Nebraska	682945
Non BioSense	New Hampshire	652190
Non BioSense	Rhode Island	590464

November		
Status	**Jurisdiction**	**ED Size estimate**
Non BioSense	Boston Public Health Commission	524804
Non BioSense	County of Sacramento, CA	509352
Non BioSense	District of Columbia	499154
Non BioSense	Delaware	437706
Non BioSense	San Diego, CA	433873
Non BioSense	Hawaii	399575
Non BioSense	Vermont	302030
Non BioSense	South Dakota	289200
Non BioSense	San Mateo, CA	247889

Appendix E

Detailed Description of Phase III Activities

During Phase III, DHIS will work with each jurisdiction for approximately 40 days to ensure a smooth transition into the ESSENCE production environment. The 40 day transition period will be broken into (1) Communications (2) Master Facility Table Review (3) Move jurisdictional data into the staging ESSENCE environment (4) Conduct ESSENCE orientation, BioSense Platform Training, and User Acceptance Testing (5) Migration to Production ESSENCE and (6) Load Jurisdictional back data into ESSENCE production environment. Here are the details of what can be expected during this phase.

1. **Communications**: DHIS staff will schedule an initial conference call with a jurisdiction so that all involved in this process (NSSP and Jurisdiction) can discuss roles and responsibilities. Additionally, the staff will review, and modify if required, the timeline associated with the 40 day window. Other items that will need to be discussed early on include, but are not limited to, items like technical connections such as sFTP, PHIN-MS, and Mirth to Mirth connections as appropriate to that jurisdiction.
2. **Master Facility Table Review**: All jurisdictions will go through a Master Facility Table review process during Phase II of the Implementation plan. During Phase II, the clean-up process is to include enhancing facility master tables so they consistently capture universal metadata about facilities that includes but is not limited to identifying NSSP-sending entity, parent facility, facility type, open/close/changed status, date added, date ended, and electronic medical record system associated with the facility's data feed. During this phase of the plan, the intent will be to review/modify the work that has already been accomplished and agree that the jurisdictions master facility table is ready for ESSENCE within the staging environment.
3. **Move Jurisdictional data into the staging ESSENCE environment**: During this step, DHIS staff will move 3 months of data (most current 3 months) into the jurisdictions ESSENCE staging environment and turn on a data feed for new incoming data (prospective data feed into BioSense will continue to flow during this step).
4. **Conduct ESSENCE Orientation, BioSense Platform Training, and User Acceptance Testing**: The NSSP team will provide the jurisdiction with User Guides for ESSENCE, BioSense Platform Single-Sign On as it is made available, and Jurisdictional Administration Tool (The Jurisdictional Administration Tool will be developed during Phase II of the Implementation Plan). Additionally, DHIS will schedule at least one, maybe two, interactive orientation sessions for each jurisdiction. Once the jurisdiction, in coordination with the DHIS staff, agree that all data connections are operational, that data is being displayed properly, and that the administration tool functions correctly, we will move to step 5.
5. **Migration to production ESSENCE**: Once user Acceptance testing has occurred, DHIS will move the jurisdictions staging environment settings, users, and data to the ESSENCE production environment. The jurisdictions and DHIS staff will conduct a final check to ensure that everything is functioning properly in the production environment.
6. **Load jurisdictional back data into the ESSENCE production environment**: DHIS staff will begin migrating all jurisdictional back data into the production environment. The time to complete this step will vary between jurisdictions. Once complete DHIS staff will inform the jurisdiction that all data has been loaded into the ESSENCE Production environment.