Data Infrastructure Development Project

Architecture & Requirements Document

Table of Contents

Table of Figures

1 Background

In January 2011, the Ohio BEACON (Best Evidence for Advancing Childhealth in Ohio Now) Council commissioned a project to develop a data infrastructure that was intended to "assist the BEACON Council in realizing its mission to enhance the use of improvement science to support quality initiatives through increased transparency and efficiency of data collection, management and analyses across all of the improvement projects[1]." Developing a standard informatics platform for all the BEACON projects would reduce the time and cost of development, allow for components to be reused and shared across projects, and lower the education and training burden for all of the BEACON project staff. Phase 1 of the Data Infrastructure Development Project was to be a planning phase, the culmination of which is this architecture design document. It provides a high-level system overview that describes the proposed architecture. It will be instantiated based on the specific needs and requirements of new BEACON projects, as well as existing BEACON projects that are migrated to the new platform.

The approach used to execute this six-month planning project is shown in Figure 1. This approach was designed to involve and engage stakeholders across the entire spectrum of Ohio BEACON projects, to ensure the resulting design would provide a usable platform for current and future quality improvement projects in Ohio (a larger version of this document appears in Appendix 1 (Section 9).

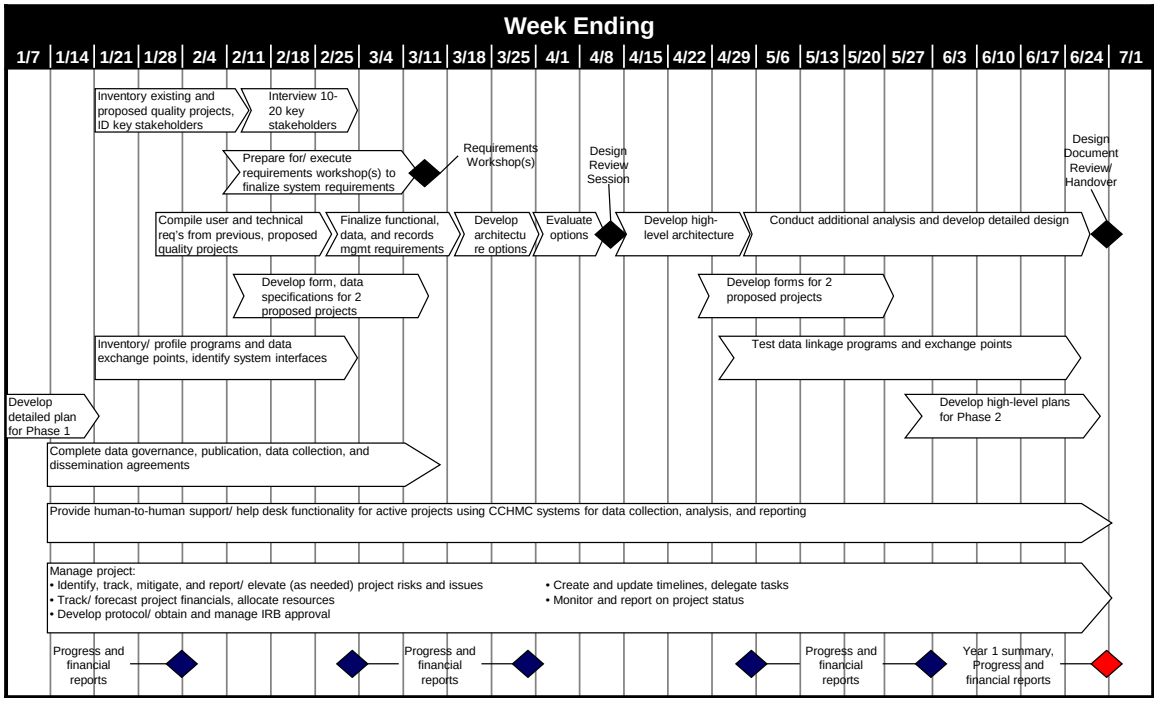


Figure 1: Approach to the planning phase of the data infrastructure development project.

The team conducted interviews with 23 stakeholders representing nine BEACON projects to understand their projects' needs and technical requirements (See Appendix 2 [Section 10] for the interview guide and Appendix 3 [Section 11] for a listing of the

[1] Perinatal Quality Improvement Project and Data Infrastructure Development Project Request for Applications

projects and stakeholders interviewed). The team also compiled the technical requirements from a variety of recent quality improvement projects with which they were involved and compared/combined these requirements with those captured through the interview process. The team rated each requirement according to its importance to the projects (based on the feedback from the interviewees) and cost/technical complexity (based on the team's technical assessment). Based on these ratings, the requirements were divided into functional categories and recommended release bundles (e.g., requirements needed for initial release, requirements needed for future releases, and non-essential requirements). These recommendations were reviewed and validated with participants of the BEACON HIT Technical Subcommittee, held on April 4, 2011, and the resulting functional requirements served as the key inputs into the design process (see Appendix 4 [Section 12] for a list of the must-have system requirements grouped by function).

The team developed several architecture options and selected a high-level architecture based on the approved requirements bundles. This conceptual architecture was vetted with the full BEACON Council during a meeting in Columbus on May 13, 2011. Based on the feedback received during this session, the team took the conceptual architecture and approved requirements and developed them into the high-level design presented here.

2 Overview

This document describes the requirements of an architecture that is to be developed to support the multi-center quality improvement and research networks that are part of the state of Ohio's BEACON initiative. The architecture will comprise a common informatics infrastructure with components that can be shared across projects. Each project will have its own set of unique constraints, but there will be crosscutting commonalities. The proposed system must allow for each project to create customized workflows as well as a customized look-and-feel, but the underlying components should remain the same.

The approach laid out in this design document is influenced by the systems developed for the Ohio Perinatal Quality Collaborative (OPQC) and Solutions for Patient Safety (SPS) projects. Due to the ad hoc nature of how these projects arose, they were largely developed as "one-off" productions, without much effort put into standardization or reusability. The requirements detailed here build on the learning from those projects, but address the shortcomings in the existing system when it comes to reusability. The infrastructure to be developed is intended for any new BEACON networks, but it must be possible to migrate existing networks into the system as new data collection projects arise.

The intent of the infrastructure development process is to build out functional components as the need arises, avoiding the time and expense of building software that is not currently needed. At a minimum, it must be possible to share forms and reports, along with common templates and site layouts that can ease the learning and training burden for users who participate in several projects. The system should be modular so that components can be swapped or upgraded without affecting the rest of the architecture.

3 Project Goals

From the perspective of the Ohio BEACON Council, the goals of the project are to develop a shared data management infrastructure to support statewide collaborative outcomes improvement in Ohio. Specifically, this will result in a standardized informatics platform that can be leveraged by all BEACON projects, resulting in a lower time and cost of development, the reuse of common system modules, and a lower education and training burden for all of the BEACON project staff.

4 Scope

This document describes the expected and "must have" features and requirements of the system. It provides a roadmap on how to meet those requirements, though it is certainly possible to design a system that meets the requirements using a different conceptual model. A detailed discussion is provided to explain certain requirements relating to data collection, user and system administration, as well as reporting and analysis. The BEACON data infrastructure will also require functionality relating to communication and collaboration. Those requirements are not discussed in detail, but a list of must have features is provided in Appendix 4 (Section 12.4) for reference purposes.

Not included in this document are any requirements relating to the construction of a BEACON-specific data warehouse, or strong links to external data sources like medical record systems. The state's implementation of a Health Information Exchange are still in the initial stages, so it is not possible to plan the design of BEACON's informatics needs more than two or three years into the future. As details emerge, specific designs will be drafted.

This document does not detail the specific forms and reports that would be part of the first build phase of the data infrastructure. At the time this document was written, the networks and corresponding projects had not been finalized, and as such, it was not possible to determine the number of forms that would need to be developed. Those requirements will be provided in a separate attachment.

Also out of scope are the elements of the project related to data management, including validation and cleaning rules, edit checks, and manuals of operations. It will be necessary to create those documents in collaboration with the data management staff of each project and network.

5 Conceptual Architecture

This conceptual architecture for the BEACON data infrastructure is derived from the infrastructure that was developed to support OPQC and SPS. It builds on the successful deployment and implementation of those projects and aims to develop a standardized approach for future projects.

To satisfy the must-have requirements of the project, the BEACON infrastructure should, at a minimum, contain the following components:

- Electronic Data Capture (EDC) – web-based forms, with client-side validation, along with a forms authoring tool that can be used by non-developers.

- Content Management System (CMS) – a collaboration space where users can create project websites (internal and external), post documents, create calendars, host discussion boards, launch forms and much more.
- Database – a robust database back-end that can support both large and small projects as well as the CMS.
- Reporting Tool – a reporting platform that can be used by moderate-to-expert users to create graphical and tabular reports that can be pushed to users or generated on demand.
- Integration Tool – a set of tools that can be used to integrate data in multiple formats from multiple sources. Database administrators can create graphical workflows that illustrate the integration process, providing a human readable approach to the extraction, transformation and loading (ETL) of data.

A high-level diagram of the system is shown in the figure below. It can be broken down into components based on functional area: data collection, user administration, reporting and data analysis and secondary use. The requirements of those areas are discussed in more detail in the following section.

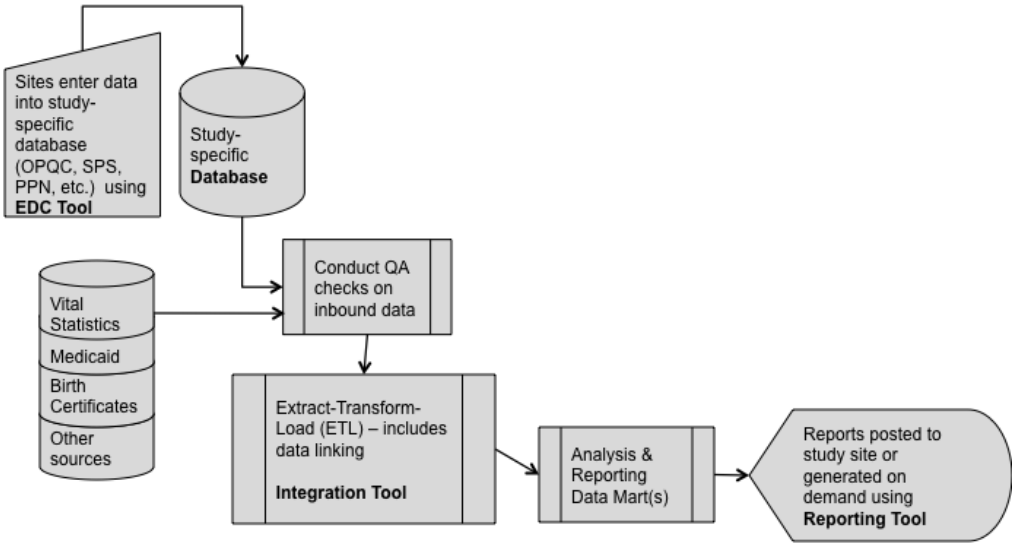


Figure 2: Conceptual system architecture.

5.1 Data Collection

Almost all BEACON project users are familiar with the use of web-based data collection forms. They allow users to submit data remotely, and when coupled with data validation, can be used to ensure high data quality at the time of submission. Based on the requirements listed below, an enterprise-scale solution will be needed in order to handle statewide projects with the capability for future growth.

5.1.1 Data collection requirements

Forty-seven of 66 data collection requirements were classified as must-have (scoring 4 or higher on a scale of 5). Notable features derived from these requirements include:

- Web-based data entry on multiple platforms/browsers (DC1, DC56)
- Form-level data validation -- before data is saved as opposed to after (DC21, DC25)
- Forms with skip patterns/branching and repeating sections (DC8)
- Ability to customize the look-and-feel for each project, both for the questionnaires and the overall web user interface (DC61, DC70)
- Pre-population of form data (DC7)
- Automated import of secondary data and integration with primary data (DC19, DC20, DC26, DC58)
- Integration with other collaborative features to create a user-friendly project portal (DC64, DC65)
- Audit trails for data access, entry, modification and export (DC43, DC46, DC48)
- High availability (DC66) and high performance (DC9, DC10)
- Ability to scale to 100 networks (DC57)
- Ability of intermediate users to build/modify forms and maintain the portal without custom programming (DC70)

A complete list of must-have data collection requirements may be found in Appendix 4 (Section 12.1).

5.1.2 Satisfying must-have requirements

The EDC and CMS components of the system should be able to satisfy all must-have data collection requirements. They should be met primarily with out-of-the-box functionality, though it is expected that some features may require the use of custom programming and add-on technologies. The number of third-party add-ons and in-house customizations should be limited as a way to mitigate risks associated with licensing, maintenance/upgrades and support. It is recognized, however, that there are also risks in forgoing the possibilities of custom development and in ignoring the vast contributions of the third-party community.

5.1.2.1 Web-based data entry

There are several advantages of web-based data entry as opposed to other methods such as paper-based data collection, use of Microsoft Excel to store/exchange data or use of client-side database systems such as Microsoft Access. First, especially for projects spanning multiple institutions/sites, there is no need to have users install special software on their systems; all they need is a computer with Internet access. This virtually eliminates the need for desk side IT support, which is expensive and not

portable from site to site. Increased data integrity and security are other key advantages. When data are stored on individual user's computers, there is greater risk of data loss (for example, when a hard drive fails) or breech (for example, when a laptop is lost or stolen). Using web-based forms that submit to a database server, on the other hand, ensures that all data are secured in a single location that is backed up.

Users should be able to open the forms in any mainstream platform (Windows, Mac, Linux) or browser (Internet Explorer, Firefox, Safari). It should be possible to save and re-open forms later for additional data entry. The connection between the user's browser and the server will need to be encrypted through the use of the secure sockets layer (SSL) protocol, and all data are stored on a centralized server that is backed up nightly, so the security and confidentiality of data is assured.

An example data entry form and corresponding data dictionary is provided in Appendices 5 & 6 (Sections 13 and 14).

During the stakeholder interviews, web-based data entry emerged as one of the key "must have" data collection requirements. While certain areas of the state still lack reliable, high-speed broadband connections, it was felt that most areas where data collection occurs (hospitals, physician offices, etc.) would have some sort of connection available. In addition, statewide efforts to bring high-speed connectivity to underserved areas should mitigate this issue in the future.

If it is deemed critical, off-line data collection options may need to be developed. These include the ability to fax paper forms to the project staff for re-entry, or through some other mechanism (like Teleforms). It should be noted that certain off-line methods lack client-side data validation (see below), making it difficult to catch data entry errors as they occur, worsening data quality.

5.1.2.2 Client-side data validation

Validation is a strategy for mitigating errors, omissions and inconsistencies in a data set. For example, checks may be implemented to verify that all required values have been collected, that values are the proper type (e.g., numerals instead of text) or that values are in an expected range (e.g., 1-100). These sorts of checks may be executed before the data has been stored or after the data has been stored. In a web-based data entry system, the former method is called client-side validation, since it occurs within the user's browser. The principal advantage of this approach is that as many errors as possible are caught before they occur. Although quality checks may still be necessary once the data is stored, the number of ongoing query resolutions should be minimized.

Project staff – without the assistance of IT staff – should be able to add validation checks to forms by using an intuitive interface similar to a Microsoft Office application. For example, one or more fields may be flagged as required, which means users will not be able to submit the form without completing them:

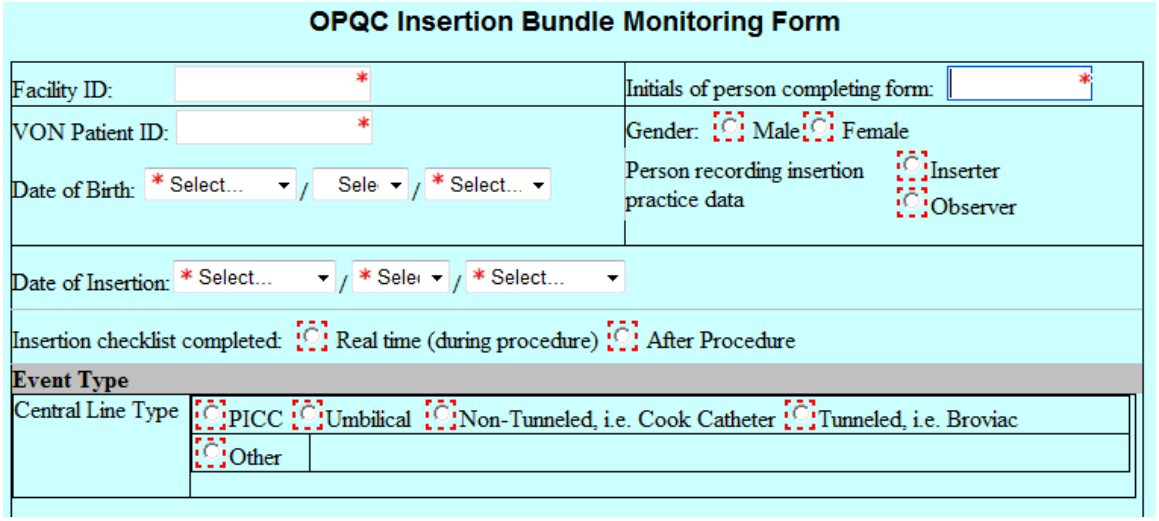


Figure 3: Example data validations on a web form.

Specific data type requirements may be added to each field as well, such as in this example where only an integer is allowed:

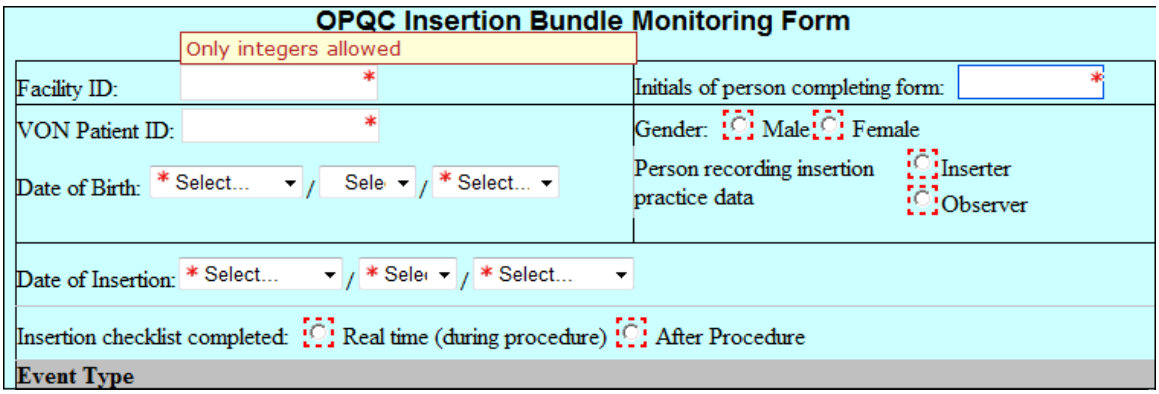


Figure 4: Tooltip explanation of field-level data validation.

Other validation options should include the ability to add soft and hard checks on data ranges. In a soft check of values between 1 and 100, for example, if a user enters 110 an alert will appear but the user will be able to save the data anyway. In a hard check of the same scenario, the user would not be able to save the value. The ability to add both hard and soft checks to the BEACON data collection forms must be available.

5.1.2.3 Skip logic/branching and repeating sections

The ability to add skip and repeat patterns is essential to any questionnaire building tool for complex data collection projects. The EDC tool should enable the creation of such patterns, again through the use of the user interface rather than custom programming. In this example questionnaire, the same fields are repeated as many times as necessary to enable a given user to input all available data, an approach that minimizes screen clutter for users with fewer inputs:

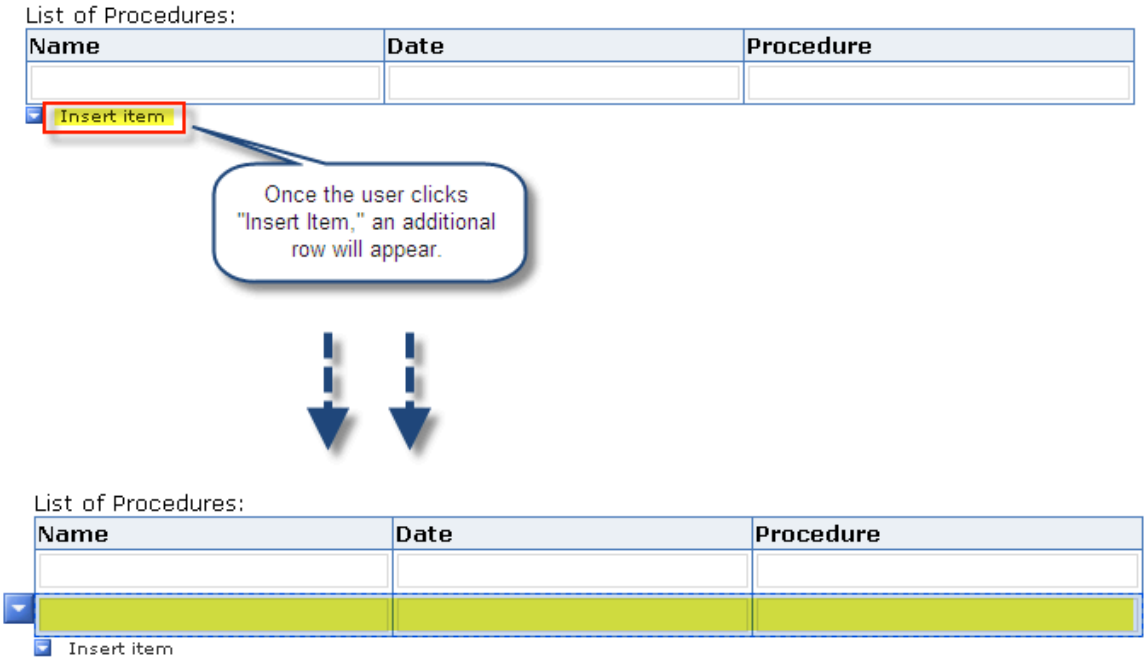


Figure 5: Example of a form with a repeating section.

5.1.2.4 User interface customization

A customized look-and-feel is often more important to data collection projects than may be expected. The ability to add a logo, change fonts and colors and manipulate other qualities of the user interface to reinforce the study or project brand must be accounted for in the selection of a data collection tool. Within the EDC form-authoring tool, it should be possible to drag and drop elements on a canvas and modify colors and other qualities using the menus/commands similar to those in Microsoft Word.

It must also be possible to customize the look-and-feel of the CMS site where the forms are hosted/launched. The CMS must offer several out-of-box themes, or templates, with the ability to install other themes generated by third parties or even create a custom theme (which may require some knowledge of web design technologies but no actual programming). The CMS must enable users to manipulate page layout by dragging and dropping objects – including the forms themselves. Below is an example of a CMS site customized to support a clinical trial:

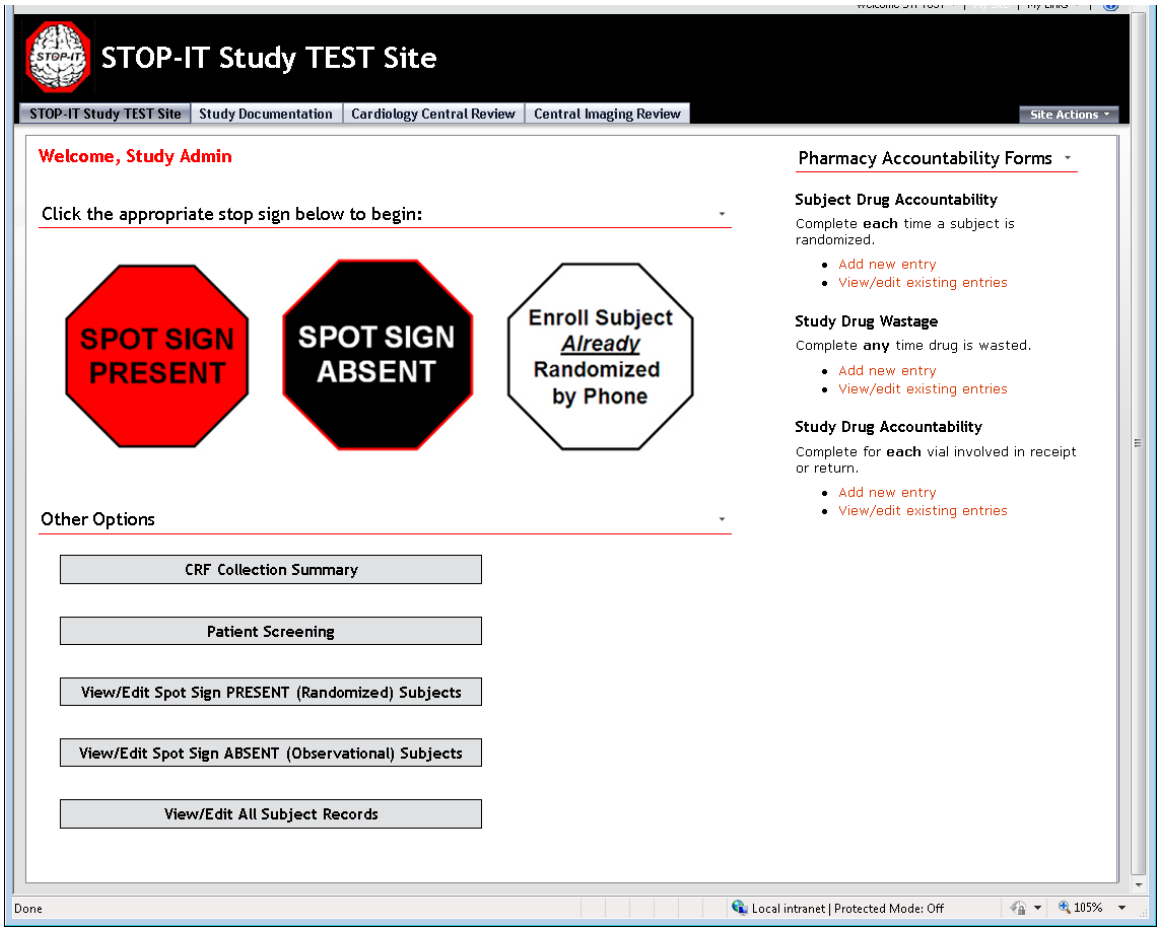


Figure 6: A study site with a customized layout.

5.1.2.5 Reusability

The need to be able to reuse components such as questionnaires across multiple projects is an important cost savings requirement. Assuming agreement on the makeup of standardized instruments – for example, a demographics form – is reached, corresponding templates will be developed. Even if a template does not contain all the fields a given project needs, it must be possible to use it as a starting point so that every form in every project is not built "from scratch." Alternative solutions such as custom-coded web forms do not place this kind of reusability directly in the hands of project staff. Moreover, with a CMS available as part of the central BEACON infrastructure, there must be no additional costs to provide the ability to share form templates across the network. It should be possible to upload them to a central library where they can be downloaded and used by any BEACON project.

5.1.2.6 Data pre-population

In some more advanced forms, it is desirable to have the ability to pre-populate information – for example, a value captured on one page of a multi-page form may be referenced elsewhere in the form. Or a value collected on one form may be referenced on another form. While the former functionality should be built native to the EDC tool, it is recognized that the latter functionality may require some customization. However, through the use of a simple web service, it should be possible for a non-developer to

retrieve values from anywhere in the data set and display them dynamically on their form.

5.1.2.7 Secondary data integration

A key component of the proposed BEACON informatics infrastructure will be the ability to integrate primary data collected using web-based forms with secondary data collected sets such as Vermont Oxford Network data and vital statistics data (e.g., birth certificates, death certificates, see Section 5.4). The proposed solution will enable collection of primary patient-level data while also providing multiple avenues for linkage and integration with secondary patient-level data. One option is to integrate the data outside of the EDC/CMS platform. This is made possible by the fact that the form data is stored in a relational database, from which it can easily be exported to other systems – for example, a SAS program – and combined with other data sets. Such exports could be manually executed or could be automated using the Integration Tool.

5.1.2.8 Audit trails

The ability to trace the complete history of database records – from creation through each modification – is an essential requirement of data collection projects, especially those involving sensitive data such as protected health information (PHI). Out-of-box, the CMS should provide the ability to retain version histories of all documents including the EDC forms.

In addition to the CMS-level audit trails, a separate audit trail should be created in the database where, for each form, a database trigger will be implemented to capture any changes to the data (including the date/time, operator, old value and new value). This dual approach to audit trails will satisfy regulatory requirements such as 21 CFR Part 11, the standard guidance for FDA-regulated clinical trials.

5.1.2.9 High availability/performance

To satisfy the needs of statewide projects, the BEACON data collection system(s) must have minimal downtime. In addition, data collection forms, although often complex in nature, must be responsive to the needs of busy health care and research professionals. Accordingly, aggressive standards have been set for system availability and performance. The proposed infrastructure must meet these standards by including multiple application and database servers in a load-balanced/failover configuration. This level of redundancy has two implications. First, traffic will be equally distributed among the servers, ensuring that at no point is any one server under undue stress. Second, if one server goes offline, intentionally (as part of a system upgrade process) or unintentionally (as an episode of failure), all connections to this server automatically will be transferred to another server. In combination, these two measures should maximize the responsiveness of the servers and should minimize downtime to the 60 seconds or less it takes for one server to fail over to another.

5.1.2.10 Scalability

A final requirement of the BEACON data infrastructure is to make it scalable so that it supports at least 100 networks while meeting the same performance and availability standards. The ability to cluster both application servers and database nodes is one approach to satisfying this requirement. Cost considerations notwithstanding, it should be possible to add additional application and database servers to the infrastructure to

increase capacity. To start, the infrastructure should consist of three application servers and a clustered database environment with at least five nodes (servers), all of which will exceed the minimum recommendations for processing, RAM, and storage. It is expected that this configuration, summarized in the following table, will be able to support at least two dozen medium-sized networks.

BEACON SHAREPOINT HARDWARE			
Machine	**CPU**	**RAM**	**Disk**
Application Server 1	2 Intel Xeon CPU 3.00 GHz	8GB	240GB
Application Server 2	2 Intel Xeon CPU 3.00 GHz	8GB	240GB
Application Server 3	2 Intel Xeon CPU 3.00 GHz	8GB	240GB
Database Servers 1-5 (clustered)	2 Intel Xeon CPU 3.00 GHz	32GB	1TB

5.1.3 Supporting documentation

A final component of the BEACON infrastructure development project is to draft supporting documentation for form development. This documentation will include user guides covering basic and advanced form design. In addition, a set of form design standards will be developed covering topics such as the following:

- Use of repeating sections/tables – while an effective device for capturing a large number of data points without cluttering the user interface, overly complex repeating sections may have performance implications
- Use of conditional formatting – consider usability gains against potential performance degradation
- Use of rules – the more cross-references among fields, the more difficult it is to make changes to forms over time
- Naming of variables – consultation with the data manager/statistician is critical; ideally, a common convention will be established across all BEACON projects to facilitate analysis
- Use of web service calls – too many “round trips” to the server may lead to slower performance; it is advisable to make clear to the user which actions lead to server calls (e.g., button clicks) rather than make “hidden” calls (e.g., when a user tabs from one field to another)

These standards, combined with design strategy sessions before development whenever possible, will further contribute to the goals of excellent system performance and usability.

5.2 System/user administration

While not highly visible to the end user, having a system where project staff can easily administer both users and the project infrastructure goes a long way towards creating a usable system. Having a system that facilitates many of the administrative workflows can allow the project staff to focus on more important matters.

5.2.1 System/user administration requirements

Seven of nine system/user administration requirements were classified as must-have (scoring 4 or higher on a scale of 5). These requirements include:

- Streamlined process for user account creation (SA1)
- Multiple layers of security (SA9)
- Ability of project staff to manage users/permissions without assistance from IT staff (SA1)
- Ability for users to request additional permissions from project staff without assistance from IT staff (SA6)
- Ongoing user account status monitoring (SA6)
- Detailed user permissions monitoring (SA7)

A detailed list of highly important requirements may be found in Appendix 4 (Section 12.2)

Two technologies should be employed to meet these requirements: an Identity and Access Management (IAM) system and a CMS.

5.2.2 Satisfying must-have requirements

Installing a CMS to support data entry, report viewing and collaboration on a highly configurable yet highly secure network that includes a web-based identity and access management system will ensure that all must-have system/user administration requirements are met.

5.2.2.1 Streamlined user account creation

Conventional methods for setting up user accounts, especially for external affiliates, involve paper-based forms, hand signatures, faxes, inter-office or postal mail, even hand delivery. Especially for projects such as the BEACON quality improvement initiatives, which involve multiple institutions and potentially hundreds of users statewide, such methods pose significant administrative burdens and may even jeopardize collaborations; a busy clinician or researcher understandably may have little tolerance for yet another bureaucratic process. At the same time, collection of basic user information, including a signature indicating agreement with terms of use, is essential for meeting organizationally and legally defined security standards.

The IAM system, which will sit in front of the BEACON infrastructure, provides a means of streamlining the user account creation process while strengthening checks of user identity and compliance. As Figure 7 illustrates, via a self-service web portal one or more designated members of a given BEACON project's staff should be able to initiate a new account request on behalf of a user (or a batch of users, for that matter) by submitting an electronic form. The user will then be emailed a secure link for setting a password, agreeing to the terms of use and gaining access to designated resources on the

network. This will remove the logistical nightmares of collecting and routing paperwork, but retain and even enhance the security and authenticity of the transaction: the project staff essentially vouches for new users by initiating the process on their behalf, and all user information is captured and stored via a secure, encrypted connection instead of through paper documents, faxes or electronic scans, all of which can easily be misplaced or intercepted by an unauthorized party.

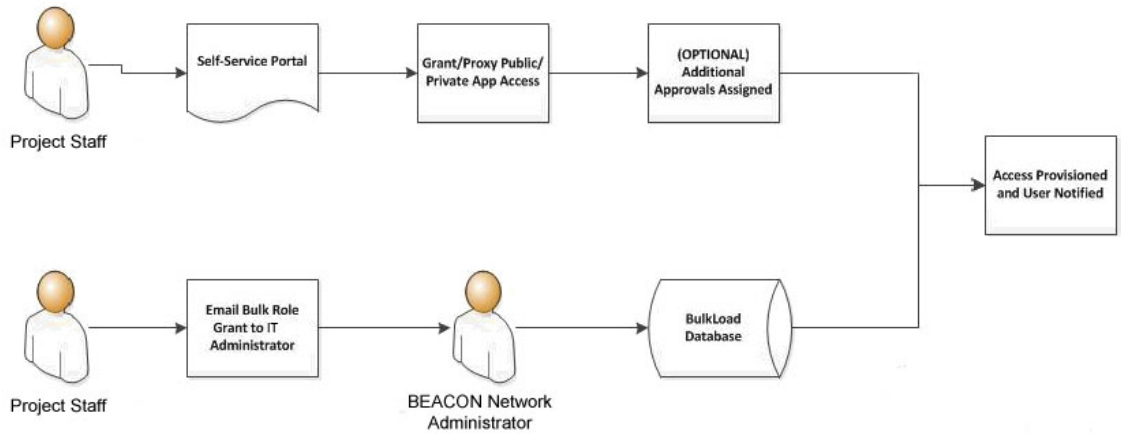


Figure 7: Workflow to create and provision user accounts.

5.2.2.2 Multiple layers of security

Because the BEACON data infrastructure will need to support a large number of projects on a common network, the ability to ensure that each project's members have access only to their own area of the network is critical. The IAM system will fulfill this requirement by intercepting all web traffic and verifying that the requesting user has access to the target resource – say, a CMS site. If the user is not authorized to access the site, the request should be denied and logged before it ever reaches the site.

On top of this very strong initial layer is the security of the CMS itself. By default, all sites should be password-protected, so even if a user were able to bypass the IAM gateway and make a request directly to the CMS server, access would only be granted if the user were a member of the site. Within the CMS platform, user access will be controlled by setting up groups – for instance, one for each participating collaborative site – and defining permissions for each group. These permissions may apply to the entire site or to specific areas such as a folder or even a specific file. So a given group by default might have read-only access to the site, with an exception to give them write access to a collaborative-wide message board (the ability to post items), along with unique access to a folder containing their institution's data, reports and other sensitive information that should not be accessible to other institutions.

5.2.2.3 Direct control of user access/permissions

In the past, setting up user access to web sites and controlling permissions within sites has required the involvement of IT staff such as network and system administrators. As the number of site users increases along with the need for a more personalized experience for each user, involving IT staff in every access or permissions request becomes untenable. Both the IAM system and CMS must enable non-IT staff to

manage user settings. Granting a user access to a given BEACON project, for example, may take several forms: in the preferred scenario, the project administrator will open the IAM web site and grant the user access directly via an electronic form. Alternatively, the project administrator may enable users to submit access requests directly, in which case they will be routed to the project staff, who will review and grant/deny the request.

Just as non-technical project staff are able to create accounts using IAM, they also can add/remove users and permissions within the CMS. Via a simple web form, for example, it should be possible to add a user to a site and assign either group or direct permissions.

5.2.2.4 Account/permissions/usage tracking

The ability to track who has access to what resources – for example, a data set or report – is an essential requirement, especially those under regulatory scrutiny. Conventional methods have included a mix of emails, spreadsheets, paper files – and in many cases have been absent altogether. For an initiative as large as the BEACON data infrastructure, these methods are inadequate. Both the IAM system and CMS platform should include functionality for monitoring and tracking user accounts and related permissions. The IAM system should serve as an authoritative source of user information including key identifiers (name, email address, user ID), account status and history, and currently assigned resources (e.g., web sites, databases). It must include default review policies that are enforced with automated workflows, as well as the ability to define custom policies/workflows. For example, the central staff of a given BEACON project should automatically receive periodic reports (e.g., annually) on users with access to their web site or database and will be required to verify, using a simple web interface, whether each user should continue to have access. This workflow, which may need to be adjusted if necessary to meet additional regulatory requirements, ensures that users have access only to the resources they should, and then only as long as necessary. Along those lines the IAM system should also enable project administrators to specify expiration dates for user access.

The CMS should provide basic site usage reporting (total page hits, frequent visitors per page) and enable site administrators to track users and permissions by browsing the membership of the various groups on the site. It should also be possible to achieve more granularity – for example, detailed histories of user access or listing of user permissions. In combination with the functionality provided by the IAM system, project staff will have on-demand access to all the information they need about their users, and also have the ability to make changes to user settings without involving IT staff.

5.3 Reporting

For many of the existing BEACON projects, the generation of data and quality reports has been a manual affair. The data is collected and stored in a database, from which monthly extracts are generated. An analyst will execute SAS procedures against the data, generating a set of reports, which are then posted back to a collaboration space. While there is nothing inherently wrong with this process, it is not an efficient use of an analyst's time. All of the standing monthly reports can be automated, freeing the analyst to work on ad hoc queries or other research. The creation of an automated reporting platform can be one of the biggest efficiency gains over the existing BEACON project infrastructure.

5.3.1 Reporting Requirements

Twenty-seven of 36 data reporting/analysis requirements were classified as must-have (scoring 4 or higher on a scale of 5). Fifteen of those requirements relate specifically to reporting. A selection of the must-have requirements are listed below:

- System to allow aggregation or drilldown of data by an intermediate user (e.g., site-specific resource, QIC, etc.) (DAR2)
- System to allow creation of clinician-, site-, or patient-specific reports, including patient and aggregate data (DAR5)
- System to allow export of data to external/ third party analytical packages (e.g., SAS), expert user to be able to create/ modify format/ structure of the export file (DAR7)
- Reports/analysis available in multiple formats (e.g., .pdf and others) that can be read on multiple platforms (computer, smartphone) (DAR12)
- System to allow automated publication/ push of approved reports to email, website (DAR13)
- System to allow presentation of graphical and tabular data (to include display of data over time, e.g., run charts) (DAR14)
- System to allow comments/ narrative to accompany graphical data representations ("write-on" capabilities) (annotation) (DAR15)
- System to allow creation/ modification of custom reports by expert user (graphical and tabular display with commentary/ narrative) (DAR16)
- System to allow graphical presentation of performance vs. benchmarks or standards (predetermined or calculated benchmarks) (DAR18)
- System to allow creation/ modification of tabular/graphical reports by an expert user (graphical and tabular display with commentary/ narrative) (DAR21)
- System to allow for creation of standard report templates (DAR29)
- Reports are printable from system in a standard format (DAR33)

A complete list of the must-have reporting requirements may be found in Appendix 4 [Section 12.3].

5.3.2 Satisfying must-have requirements

The BEACON Reporting Tool should support all of the must-have reporting requirements. Some of the requirements, such as the ability to aggregate or drill-down into the data, or to provide identified or de-identified data to authorized users may

depend on the report configuration, but there should be nothing in the tool to prevent the creation of such a report.

5.3.2.1 Reporting Authoring

End users should be able to develop reports for the BEACON infrastructure using a self-service authoring tool. A Microsoft Office-like authoring environment that allows users to embed features such as sparklines, maps, the data bar, and indicator data visualizations is preferable. It should include the ability to create reports containing aggregates of aggregates, as well as enhanced support for expressions (complex calculations or function calls that can be used to compute/organize data or to change the report appearance. Easy-to-use wizards should be able to walk users through the steps of creating a table, matrix, chart, or map.

In the preferred model, reporting is not done against the transactional production database used for data collection. Complex analyses can adversely affect system performance; therefore, reporting should be done against either a copy of the database (snapshot or extract), or a view. The reporting databases are typically refreshed once a day, but if more real-time reporting is required, triggers will need to be written to update the data on a more frequent basis.

The data needed for a given report will be produced by directly querying the database, calling a stored procedure, or by performing a set of expressions or calculations. In the event where a report cannot be produced using these methods, if it relied on the result of a complex analysis that can only be produced using SAS, for instance, the general workflow should be as follows. The source dataset would be produced, either as a database or a flat file. The SAS procedure would be executed (manually or as a batch job), producing the results needed for the report, which would then be read as another data source by the Reporting Tool, meaning it would function simply as a graphical interface to the data.

Since many existing reports exist as SAS procedures, the above approach may be used in the short-term, but over time, the intent would be to convert SAS procedures into the native report language used by the Reporting Tool, allowing the report production to be completely automated.

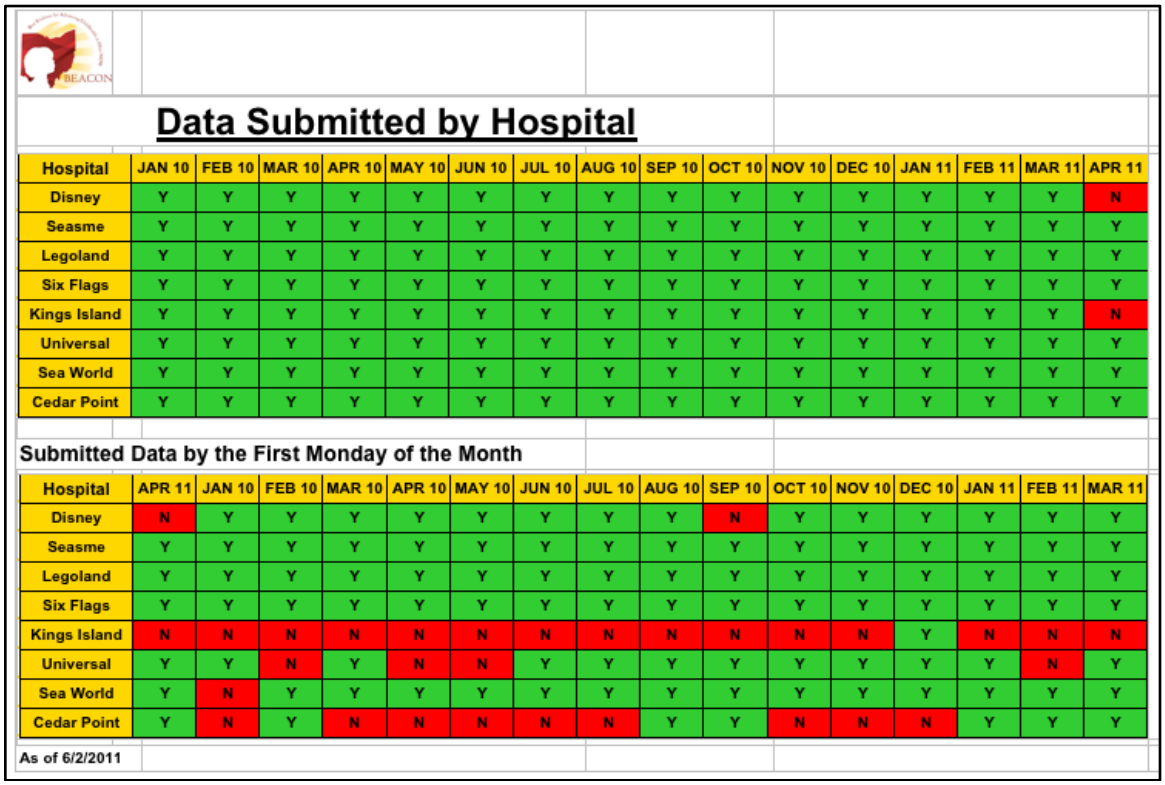
5.3.2.2 Report Templates

To help standardize the data and quality reports used by the BEACON community, each project using the data infrastructure are encouraged to adopt a set of template reports that provide information on items like whether data was submitted on time, whether it meets the quality standards of the project, and how the performance of a given site compares against a specific benchmark. The look-and-feel of each report can be customized, but the underlying logic behind each one would remain the same. Projects must be free to create additional reports if they have specialized needs, and each project is also encouraged to share their reports as templates with the larger BEACON community.

The following set of reports are an example of the templates that would be used by the BEACON projects. They are modeled after the reports developed by the Solutions for Patient Safety collaborative, as they are the only BEACON project that utilizes

automated reporting. The form used to collect the data in these reports can be found in Appendix 5 (Section 13). It is envisioned that these templates will evolve over time as projects find more efficient and effective ways of displaying the data. Once a report has been validated and vetted by the larger community, the changes can be propogated to the different projects.

The report in Figure 8 provides an example of a data submission report. The top table denotes whether a given site has submitted their data for a given month. The table in the bottom shows whether the data was submitted on time, where the on-time deadline is uniquely defined for each project.



Data Submitted by Hospital

Hospital	JAN 10	FEB 10	MAR 10	APR 10	MAY 10	JUN 10	JUL 10	AUG 10	SEP 10	OCT 10	NOV 10	DEC 10	JAN 11	FEB 11	MAR 11	APR 11
Disney	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Seasme	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Legoland	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Six Flags	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Kings Island	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Universal	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sea World	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cedar Point	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

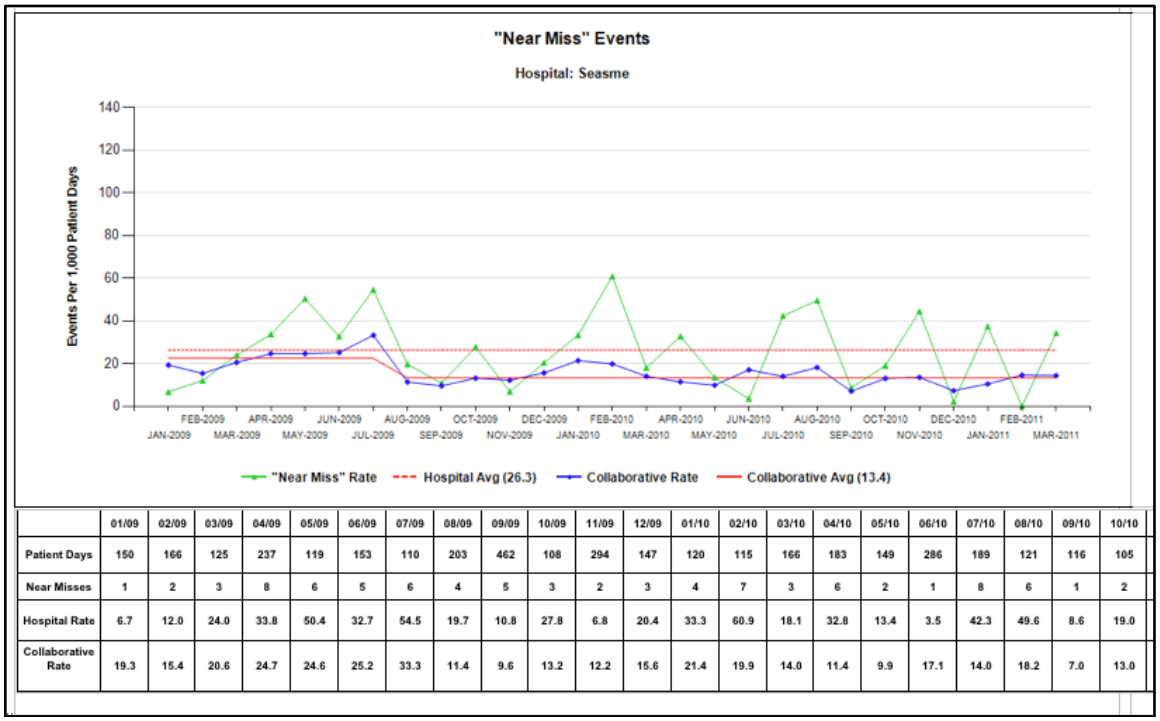
Submitted Data by the First Monday of the Month

Hospital	APR 11	JAN 10	FEB 10	MAR 10	APR 10	MAY 10	JUN 10	JUL 10	AUG 10	SEP 10	OCT 10	NOV 10	DEC 10	JAN 11	FEB 11	MAR 11
Disney	N	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y
Seasme	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Legoland	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Six Flags	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Kings Island	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	N	N
Universal	Y	Y	N	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	N	Y
Sea World	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cedar Point	Y	N	Y	N	N	N	N	N	Y	Y	N	N	N	Y	Y	Y

As of 6/2/2011

Figure 8: Status of submitted data. The table at the top illustrates whether a site has submitted their data for a given month. The bottom shows whether the submission was on time.

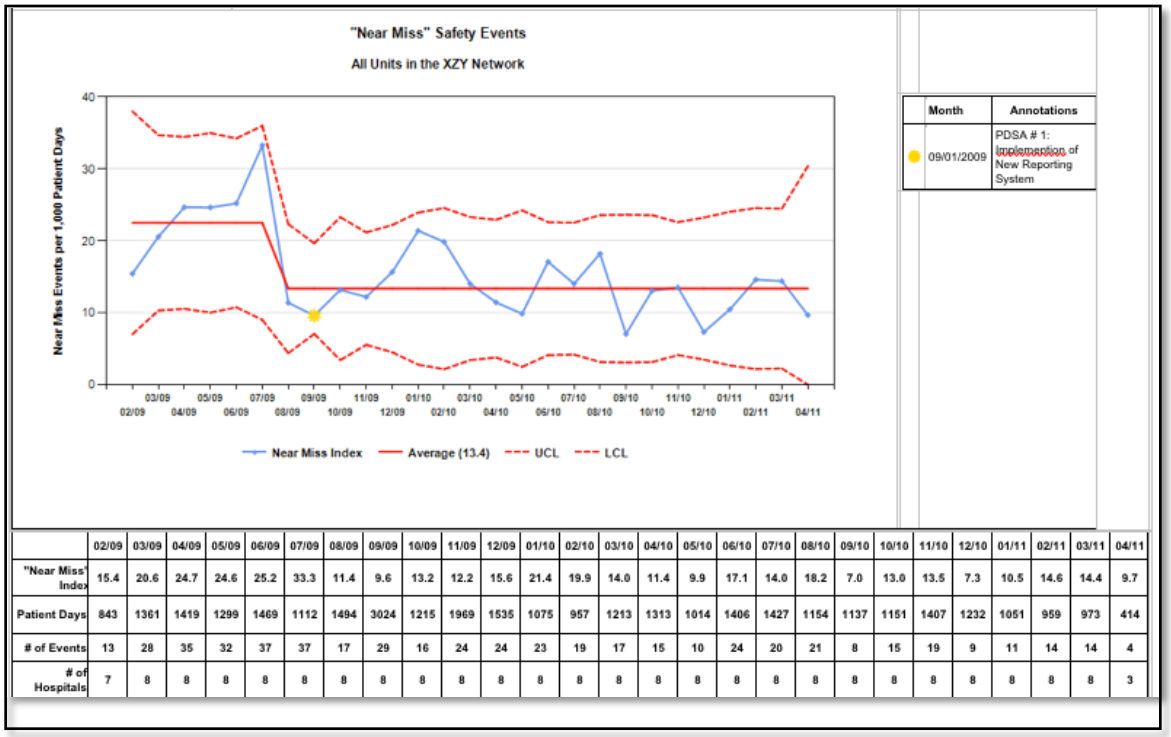
Figure 9 provides an example of a run chart with a data table. Several peformance metrics are shown on the chart. These metrics include the specific performance of a site along with that of the whole collaboraive. While it is not shown in the figure, it is possible for an end user to set the start and end time points before generating the chart.



	01/09	02/09	03/09	04/09	05/09	06/09	07/09	08/09	09/09	10/09	11/09	12/09	01/10	02/10	03/10	04/10	05/10	06/10	07/10	08/10	09/10	10/10
Patient Days	150	166	125	237	119	153	110	203	462	108	294	147	120	115	166	183	149	286	189	121	116	105
Near Misses	1	2	3	8	6	5	6	4	5	3	2	3	4	7	3	6	2	1	8	6	1	2
Hospital Rate	6.7	12.0	24.0	33.8	50.4	32.7	54.5	19.7	10.8	27.8	6.8	20.4	33.3	60.9	18.1	32.8	13.4	3.5	42.3	49.6	8.6	19.0
Collaborative Rate	19.3	15.4	20.6	24.7	24.6	25.2	33.3	11.4	9.6	13.2	12.2	15.6	21.4	19.9	14.0	11.4	9.9	17.1	14.0	18.2	7.0	13.0

Figure 9: Run chart with data table. This report shows a chart with various measures, along with the data that comprise each line.

Figure 10 presents an example of a control chart with annotations. The upper and lower control limits can be automatically determined, or set by the end user, whichever behavior is preferred. The ability to add annotations is a must, using the report-authoring tool, for example. In this case, the project's data manager or data analyst would add them if a special cause were determined.



	02/09	03/09	04/09	05/09	06/09	07/09	08/09	09/09	10/09	11/09	12/09	01/10	02/10	03/10	04/10	05/10	06/10	07/10	08/10	09/10	10/10	11/10	12/10	01/11	02/11	03/11	04/11
"Near Miss" Index	15.4	20.6	24.7	24.6	25.2	33.3	11.4	9.6	13.2	12.2	15.6	21.4	19.9	14.0	11.4	9.9	17.1	14.0	18.2	7.0	13.0	13.5	7.3	10.5	14.6	14.4	9.7
Patient Days	843	1361	1419	1299	1469	1112	1494	3024	1215	1969	1535	1075	957	1213	1313	1014	1406	1427	1154	1137	1151	1407	1232	1051	959	973	414
# of Events	13	28	35	32	37	37	17	29	16	24	24	23	19	17	15	10	24	20	21	8	15	19	9	11	14	14	4
# of Hospitals	7	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	3

Figure 10: Control chart with annotations. This chart illustrates how text annotations can be added to a graph in order to denote special causes (marked by yellow asterisk).

An example of a site-specific quality dashboard is shown in Figure 11. The dashboard is used to present a roll-up of the data quality and performance metrics, illustrating the site's performance as well as the network as a whole.

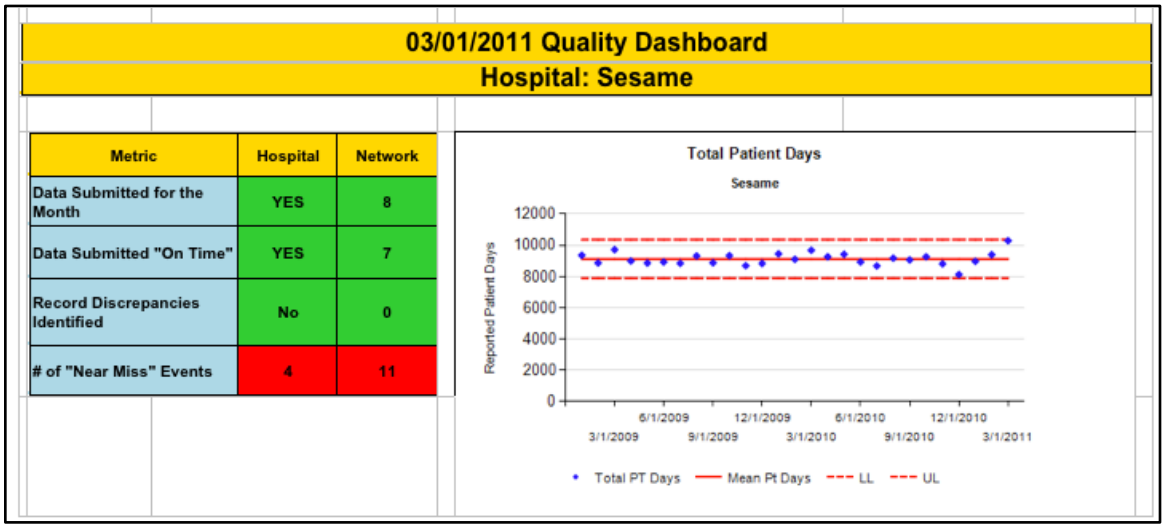


Figure 11: Site-specific monthly dashboard. This dashboard provides a summary for each site as well as the overall network: whether data was submitted and on time, whether there were any issues with the data, as well as the results of certain outcome measures.

The Reporting Tool should allow users to embed multiple charts into a single page, as shown in Figure 12. Displaying several metrics on a report allows users to get a general sense of performance at a glance. This particular example shows a series of run charts, but it is also possible to embed different graphical elements, such as sparklines. It is not necessarily to include data tables, but they can be included, if desired.

Disney	JAN 10	FEB 10	MAR 10	APR 10	MAY 10	JUN 10	JUL 10	AUG 10	SEP 10	OCT 10	NOV 10	DEC 10	JAN 11	FEB 11	MAR 11
Patient Days	164	93	97	192	101	243	109	89	87	118	119	87	24	154	95
Events	2	4	1	1	1	1	3	2	1	2	0	0	0	2	0
Hospital Rate	12.2	43.0	10.3	5.2	9.9	4.1	27.5	22.5	11.5	16.9	0.0	0.0	0.0	17.5	0.0
Collab Rate	21.4	19.9	14.0	11.4	9.9	17.1	14.0	18.2	7.0	13.0	13.5	7.3	10.5	14.6	14.4

Sesame	JAN 10	FEB 10	MAR 10	APR 10	MAY 10	JUN 10	JUL 10	AUG 10	SEP 10	OCT 10	NOV 10	DEC 10	JAN 11	FEB 11	MAR 11	APR 11
Patient Days	129	115	166	183	149	286	189	121	116	105	135	476	107	101	175	217
Events	4	7	3	6	2	1	8	6	1	2	6	1	4	0	6	2
Hospital Rate	33.3	60.9	18.1	32.8	13.4	3.5	42.3	49.6	8.6	19.0	44.4	2.1	37.4	0.0	34.3	9.2
Collab Rate	21.4	19.9	14.0	11.4	9.9	17.1	14.0	18.2	7.0	13.0	13.5	7.3	10.5	14.6	14.4	9.7

Lego	JAN 10	FEB 10	MAR 10	APR 10	MAY 10	JUN 10	JUL 10	AUG 10	SEP 10	OCT 10	NOV 10	DEC 10	JAN 11	FEB 11	MAR 11	APR 11
Patient Days	158	140	112	81	101	167	113	119	157	113	119	110	72	82	80	109
Events	3	1	4	2	2	2	1	1	0	3	0	2	0	0	1	1
Hospital Rate	19.0	7.1	35.7	24.7	19.8	12.0	8.8	8.4	0.0	26.5	0.0	18.2	0.0	0.0	12.5	9.2
Collab	21.4	19.9	14.0	11.4	9.9	17.1	14.0	18.2	7.0	13.0	13.5	7.3	10.5	14.6	14.4	9.7

Six Flags	JAN 10	FEB 10	MAR 10	APR 10	MAY 10	JUN 10	JUL 10	AUG 10	SEP 10	OCT 10	NOV 10	DEC 10	JAN 11	FEB 11	MAR 11	APR 11
Patient	88	75	86	99	85	81	77	113	62	84	66	85	78	99	104	88
Events	1	4	1	2	0	2	4	5	0	1	0	1	0	5	2	1
Hospital Rate	11.4	53.3	11.4	20.2	0.0	24.7	51.9	44.2	0.0	11.9	0.0	11.8	0.0	50.5	19.2	11.4
Collab	21.4	19.9	14.0	11.4	9.9	17.1	14.0	18.2	7.0	13.0	13.5	7.3	10.5	14.6	14.4	9.7

Figure 12: Thumbnail view of site performance by outcome measure. Each chart represents the performance of a site for a specific outcome measure. Multiple charts can be displayed on a single page.

The Reporting Tool should provide the ability to export data in several common formats. The standard formats could be data-specific (CSV, XML), meaning only the data behind the report is downloaded, not the graphics of the report itself. Or they could be more of a WYSIWYG (What You See Is What You Get) type – PDF, MHTML, Excel, TIFF and Word, for example. It is not expected that the reports will "pretty print" by default. This means that without configuration, a report may end up split across multiple pages. If pretty printing is desired, it should be possible to ensure proper formatting during the report design.

The proposed data infrastructure should provide several ways of providing data extracts to analysts and investigators. One of those methods must be through a report. Figure 13 provides such an example. The data is presented in a table and can be downloaded via Excel. An advantage to this approach is that a project can leverage the access and security controls native to the CMS to ensure that only authorized users have access.

Data Extraction Example

Hospital ID	Patient Tracking Number	Patient ID	Current Weight	Admission Date	Discharge	Age (in Days)	Inpatient Unit
XYZ	67	XYZ - 67	39.5	12/7/2009 12:00:00 AM	12/10/2009 12:00:00 AM	156	Med - Surg
XYZ	68	XYZ - 68	41	12/12/2009 12:00:00 AM	12/22/2009 12:00:00 AM	108	Med - Surg
XYZ	69	XYZ - 69	13.0	12/7/2009 12:00:00 AM	12/10/2009 12:00:00 AM	24	Med - Surg
XYZ	70	XYZ - 70	23.6	11/29/2009 12:00:00 AM	12/1/2009 12:00:00 AM	120	Critical Care
XYZ	71	XYZ - 71	15.4	12/18/2009 12:00:00 AM	12/21/2009 12:00:00 AM	36	Medical
XYZ	72	XYZ - 72	7.4	12/6/2009 12:00:00 AM	12/8/2009 12:00:00 AM	9	Medical
XYZ	73	XYZ - 73	50	11/25/2009 12:00:00 AM	12/6/2009 12:00:00 AM	168	Med - Surg
XYZ	74	XYZ - 74	31.3	12/23/2009 12:00:00 AM	12/25/2009 12:00:00 AM	132	Critical Care
XYZ	75	XYZ - 75	22.8	12/26/2009 12:00:00 AM	12/28/2009 12:00:00 AM	96	Med - Surg
XYZ	76	XYZ - 76	22.2	12/1/2009 12:00:00 AM	12/3/2009 12:00:00 AM	96	Med - Surg
XYZ	77	XYZ - 77	12.9	12/8/2009 12:00:00 AM	12/11/2009 12:00:00 AM	12	Med - Surg
XYZ	78	XYZ - 78	17.2	11/28/2009 12:00:00 AM	12/1/2009 12:00:00 AM	72	Critical Care
XYZ	79	XYZ - 79	48.8	12/4/2009 12:00:00 AM	12/13/2009 12:00:00 AM	132	Med - Surg
XYZ	80	XYZ - 80	26	11/23/2009 12:00:00 AM	12/2/2009 12:00:00 AM	96	Critical Care
XYZ	81	XYZ - 81	16.8	12/6/2009 12:00:00 AM	12/16/2009 12:00:00 AM	36	Med - Surg
XYZ	82	XYZ - 82	38	12/4/2009 12:00:00 AM	12/7/2009 12:00:00 AM	288	Specialty Practice
XYZ	83	XYZ - 83	50.9	12/8/2009 12:00:00 AM	12/15/2009 12:00:00 AM	216	Acuity Adaptable
XYZ	84	XYZ - 84	3.3	11/19/2009 12:00:00 AM	12/11/2009 12:00:00 AM	1	NICU
XYZ	85	XYZ - 85	4.6	11/2/2009 12:00:00 AM	12/9/2009 12:00:00 AM	4	Critical Care
XYZ	86	XYZ - 86	46.8	12/11/2009 12:00:00 AM	12/15/2009 12:00:00 AM	120	Critical Care
XYZ	358	XYZ - 358	22	1/26/2010 12:00:00 AM	1/30/2010 12:00:00 AM	60	Specialty Practice
XYZ	359	XYZ - 359	46.9	1/9/2010 12:00:00 AM	1/15/2010 12:00:00 AM	192	Psych
XYZ	360	XYZ - 360	57.6	1/21/2010 12:00:00 AM	1/26/2010 12:00:00 AM	192	Psych
XYZ	361	XYZ - 361	20	1/10/2010 12:00:00 AM	1/13/2010 12:00:00 AM	72	Med - Surg
XYZ	362	XYZ - 362	52.4	1/21/2010 12:00:00 AM	1/24/2010 12:00:00 AM	180	Med - Surg
XYZ	363	XYZ - 363	20.7	1/5/2010 12:00:00 AM	1/7/2010 12:00:00 AM	96	Onocology
XYZ	364	XYZ - 364	2.14	1/22/2010 12:00:00 AM	1/30/2010 12:00:00 AM	0	NICU
XYZ	365	XYZ - 365	57.5	1/2/2010 12:00:00 AM	1/4/2010 12:00:00 AM	204	Med - Surg
XYZ	366	XYZ - 366	38.6	1/3/2010 12:00:00 AM	1/9/2010 12:00:00 AM	72	Surgical
XYZ	367	XYZ - 367	6.41	1/25/2010 12:00:00 AM	1/30/2010 12:00:00 AM	2	Med - Surg
XYZ	368	XYZ - 368	11	12/9/2009 12:00:00 AM	1/6/2010 12:00:00 AM	24	Med - Surg
XYZ	369	XYZ - 369	17.8	1/13/2010 12:00:00 AM	1/15/2010 12:00:00 AM	72	Critical Care
XYZ	370	XYZ - 370	4.05	1/21/2010 12:00:00 AM	1/24/2010 12:00:00 AM	0	NICU
XYZ	371	XYZ - 371	11.3	1/19/2010 12:00:00 AM	1/22/2010 12:00:00 AM	12	Medical
XYZ	372	XYZ - 372	50.3	1/25/2010 12:00:00 AM	1/28/2010 12:00:00 AM	144	Med - Surg
XYZ	373	XYZ - 373	2.3	12/22/2009 12:00:00 AM	1/13/2010 12:00:00 AM	1	Critical Care
XYZ	374	XYZ - 374	74.4	12/29/2009 12:00:00 AM	1/1/2010 12:00:00 AM	216	Surgical
XYZ	375	XYZ - 375	50	1/5/2010 12:00:00 AM	1/8/2010 12:00:00 AM	156	Med - Surg
XYZ	376	XYZ - 376	45	1/4/2010 12:00:00 AM	1/10/2010 12:00:00 AM	168	Onocology
XYZ	377	XYZ - 377	101.3	1/13/2010 12:00:00 AM	1/16/2010 12:00:00 AM	276	Specialty Practice

Figure 13: Tabular reports can be created as a mechanism to allow for on-demand data extracts to authorized users.

5.4 Data Analysis and Secondary Use

While the primary use of the data collected by the BEACON projects is to drive improvement in outcomes of each project, linking the study-specific data to other sources such as the Vermont Oxford database, birth/death certificates, Medicaid claims, and medical record/discharge data, enables investigators to use the combined datasets for additional research and analysis.

5.4.1 Data Analysis and Secondary Use Requirements

Twenty-seven of 36 data reporting/analysis requirements were classified as must-have (scoring 4 or higher on a scale of 5). Twelve were related specifically to the analysis of data. They are as follows:

- System to allow complex analysis of data at patient and/or practice level (Desired system capabilities: sort patients by due date, severity classification, date of last visit; compare performance by practice/clinician; 1-system w. separate db required to handle both levels of data; second db is likely source to collect secondary data collected (e.g. survey monkey) (DAR1)
- Ability to generate extract to look at root-level data (DAR3)
- System to allow analysis/manipulation/presentation of qualitative and quantitative data (DAR6)
- System to allow export of reports for upload into external systems (e.g., EMRs); expert user to be able to create/modify format/structure of the export file (DAR8)
- System tracks and documents revisions to data (within database) - audit trail (DAR11)
- System to allow automated execution of pre-defined analysis/manipulation of a dataset (create 'standard' report upon receipt of clean data) (DAR19)
- System to allow analysis/presentation of current and/or archived data (DAR22)
- System to allow validation/checking of data by authorized users (DAR23)
- System to allow modification of data in database by authorized users with proper training (DAR24)
- Logs and creates reports on errors (validation, import, export, etc.) (DAR25)
- System tracks and documents which designated user inputs/modifies data for each entry (DAR28)
- System can identify and store a snapshot of data used for analysis for a predetermined amount of time for a subset for the dataset (DAR31)

A complete list of the must-have data analysis requirements may be found in Appendix 4 [Section 12.3].

5.4.2 Database & Data Integration Tools

Any tools used for the linking of data should be able to support the integrate data in multiple formats from multiple sources. Database administrators should have the ability to create graphical workflows that illustrate the integration process and provide a human readable approach to the extraction, transformation and loading (ETL) of data. The tools should have the ability to generate extracts, trigger process execution upon file receipt and produce error logs. All BEACON data should be housed in an enterprise-class database platform that supports audit trails, user roles and security.

5.4.3 Satisfying must-have requirements

Each BEACON project is likely to have different cleaning and validation rules, as well as different data input and output formats, so meeting the must-have requirements will require the creation of unique configurations. As a result, this is likely to be one of the more labor-intensive areas of the project. Since no BEACON project is currently linking data in an automated fashion using these tools, it is not possible to show "real-world" examples. A general strategy for linking data is described below.

5.4.3.1 Linking different data sources

When it comes to the linking and integration of different data sources, there are few specifics as to "how" the integration is to proceed. For the most part, it will have to be done on a project-by-project basis. There are two main reasons. The first is that because of the differences in the data collected by each project, sources will not always be linked using the same fields, though there is expected to be a core set that will be used in most cases. The second reason is that due to the legal/privacy restrictions on many secondary sources, barring a blanket agreement allowing full access for any project under the BEACON umbrella, each project is going to have to request specific fields for a specific, defined research purpose. The use of an "honest broker," who can link and de-identify data that is then to be provided to others, can help alleviate some of the regulatory issues, but such a setup requires its own IRB protocol.

5.4.3.2 General Workflow for Linking Data

The general process for linking data is described below.

1. Data is collected or generated using one of the following methods:
    a. Direct data entry (using the approaches described in previous sections).
    b. File uploads using pre-defined formats. The files themselves will be one of the traditional formats like comma-separated values (csv), XML, or Excel. The internal formatting of the file contents (i.e., field definitions) will be driven by the specific requirements of each project.
    c. Data is obtained from external sources (Medicaid claims, birth/death certificates, EMR, etc.). The collection methods for these sources are outside the purview of the BEACON projects, and as such, BEACON would have little to no influence on the collection practices. The data from these sources may be uploaded into the system as individual files, or they may be obtained via a direct database connection or web service, depending on the access methods provided by each source.
2. Data validation:
    a. It is recommended that the BEACON projects adopt a common set of data cleaning/data quality procedures that can be used across all projects. These would include both project-level and site-level procedures. However, even with BEACON-wide scripts, some project-specific validation rules will also be necessary.
    b. The data validation rules can be triggered before the data is submitted. In the case of a web form, a field can be fixed before a form is submitted. With file uploads, a set of exception reports can be generated listing the results of hard and soft edit checks. Data elements that fail the soft checks (warnings) may be allowable, while those that fail hard checks (errors) would need to be corrected before the upload is accepted.

c. With external data, it will not really be possible to correct any spurious or missing data values. Therefore, it must be decided whether a certain level of data quality is required before linking is allowed to occur. Otherwise, the validity of the analysis may be compromised. The investigators and data managers involved with the project will be responsible for making these decisions.

3. Data integration:
    a. After execution of the different data cleaning and validation procedures, the sources will be integrated by joining on common fields. These fields are likely to vary by project, though there will be some overlap. For instance, when linking a BEACON project to Medicaid claims, a patient's gender, date of birth and the Medicaid ID of their provider may be enough to identify a patient with a high degree of accuracy. But since some of those identifiers qualify as PHI under HIPAA, collecting that information may require IRB amendments (for existing projects) or special waivers of consent and/or data use agreements for new projects. The validation scripts and data integration will be executed using a combination of database programming and ETL processes.
    b. Once the data is linked, it will be provided back to the project staff as a set of datamarts housed in a database. These datamarts can then be used to produce data for reports using the workflows and technologies described in the previous sections, or as extracts, since much of the secondary research and analysis is likely to be conducted using advanced statistical programming languages like SAS, R or STATA. The ideal approach would be to have most of the analysis done against the datamarts, and not on the extracts. Keeping the data on the server allows for central administration and auditing, making it possible to take advantage of the user permissions and row/column-level security. Once an extract is provided, it is up to the end user to ensure that proper data security procedures are followed.

5.5 Records Retention

To safeguard BEACON systems and data, all components required to reconstruct or reestablish any given system or application should be stored within two separate and secure sites at all times. The first site is a platform for source control that can be used throughout the software development lifecycle and which will serve as a general repository beyond development and throughout the useful life of the production system. This repository will preserve the integrity of system and application elements, including a degree of versioning, and includes application modules, files, documentation, and any other aspects needed to reproduce or modify the product. This site is in addition to the existing server backup and recovery plan that would be in place for all BEACON servers, but is not an unnecessary redundancy, as the two systems have different purposes.

The second site where complete systems and applications and data will be archived is an offsite storage system. This permanent archival storage is implemented after the life of the application or system and, unlike the source repository, includes all relevant data. The data and systems should be retained indefinitely, or until notified otherwise.

This structure of file and data retention will satisfy not only the immediate production needs to assure ongoing access and reliance, but also the 21 CFR Part 11 compliance requirements for FDA-regulated studies. These FDA requirements include the need for continued access controls and audits for the useful life of the system. To this end, the source control system must allow control policies that govern who and how access and changes are allowed, records such actions, and optionally alerts named users when events occur.

6 System Benefits & Drawbacks

The system described in this document must provide a number of benefits, laid out in the list below:

- The system must be capable of meeting all of the “must-have” requirements, as specified by the BEACON Technical Committee.
- It must provide the ability for end users to perform a large amount of the work when it comes to user configuration and report and form development. This lessens the need to rely on applications for minor or straightforward changes. Developers may still be needed to perform tasks like quality assurance, template development and form deployment, but there must be a role for end users.
- The system must be scalable, high-performance and high-availability. It must have the ability to handle statewide projects, with hundreds or thousands of users.
- It must meet or exceed the requirements for security, auditing, and protection of patient privacy.
- The proposed components of the system must be easily integrated.
- While there is a focus on creating common templates and reusable components, it should be possible to create customized workflows and tailor the look-and-feel so that is it unique to every project.
- The system should require only a handful of staff to administer, and it should be possible to amortize this cost among projects.

While the benefits of the proposed architecture far outweigh the drawbacks, there are a few that should be noted:

- The design and development of reports and forms can get complex. It is expected that developers will handle these tasks initially as BEACON users undergo training and become familiar with the tools. Also, developers will be needed to perform quality assurance on the forms in order to ensure acceptable system performance.
- Linking different data sources will require database programming, especially as new sources are added. It is likely that the system will be unable to "automatically" accept new data without configuration.
- Form development and deployment is not expected to be instantaneous. It takes some time to go through the process. The ability to create studies "on the fly" was sacrificed in order to create a system that is expandable and can handle future, as yet undefined requirements.
- The underlying technologies are likely to require a license.

7 Assumptions & Constraints

This document was drafted under the following assumptions and constraints. It was assumed that the document would provide a high-level overview of the system architecture. Specific project requirements would come later during the build phase, and would be used to create an implementation plan. Data validations, cleaning rules, edit checks, etc. will be defined by data management staff, and would not be defined until the specific project requirements are defined. Finally, it was assumed that system documentation and a manual of operations would be developed as part of the deployment process, in conjunction with data management staff.

In terms of constraints, the underlying system requirements were derived from the experience of developing and using the system for OPQC and SPS. In addition, it was decided that the system must be enterprise-scale and be able to handle undefined future extensions. Users need the ability to customize certain elements while also having the ability to create common reusable templates. Finally, the system design must be modular so that it can be built in phases as funding and projects allow.

8 Project Risks & Mitigations

Taking on a project as large as the BEACON data infrastructure requires careful planning for, and mitigation of, risks. The vendor should follow defined methods and processes for software development – specifically, a modified version of the agile development methodology. Experienced project managers and business analysts should be involved to ensure proper implementation of this methodology, in which system development occurs in short, iterative cycles. Each cycle will include testing and other quality assurance checks. At the end of each cycle, the development staff and project staff/stakeholders will review progress on the system, thus minimizing the risk that it fails to meet defined requirements/specifications. The lessons learned in each cycle feed the next cycle until a final deliverable is achieved. At this point, defined change management processes take over and are followed to ensure that any modifications to code, settings, hardware, etc., undergo careful planning and impact analysis.

9 Appendix 1 – Data Infrastructure Development Project: Approach

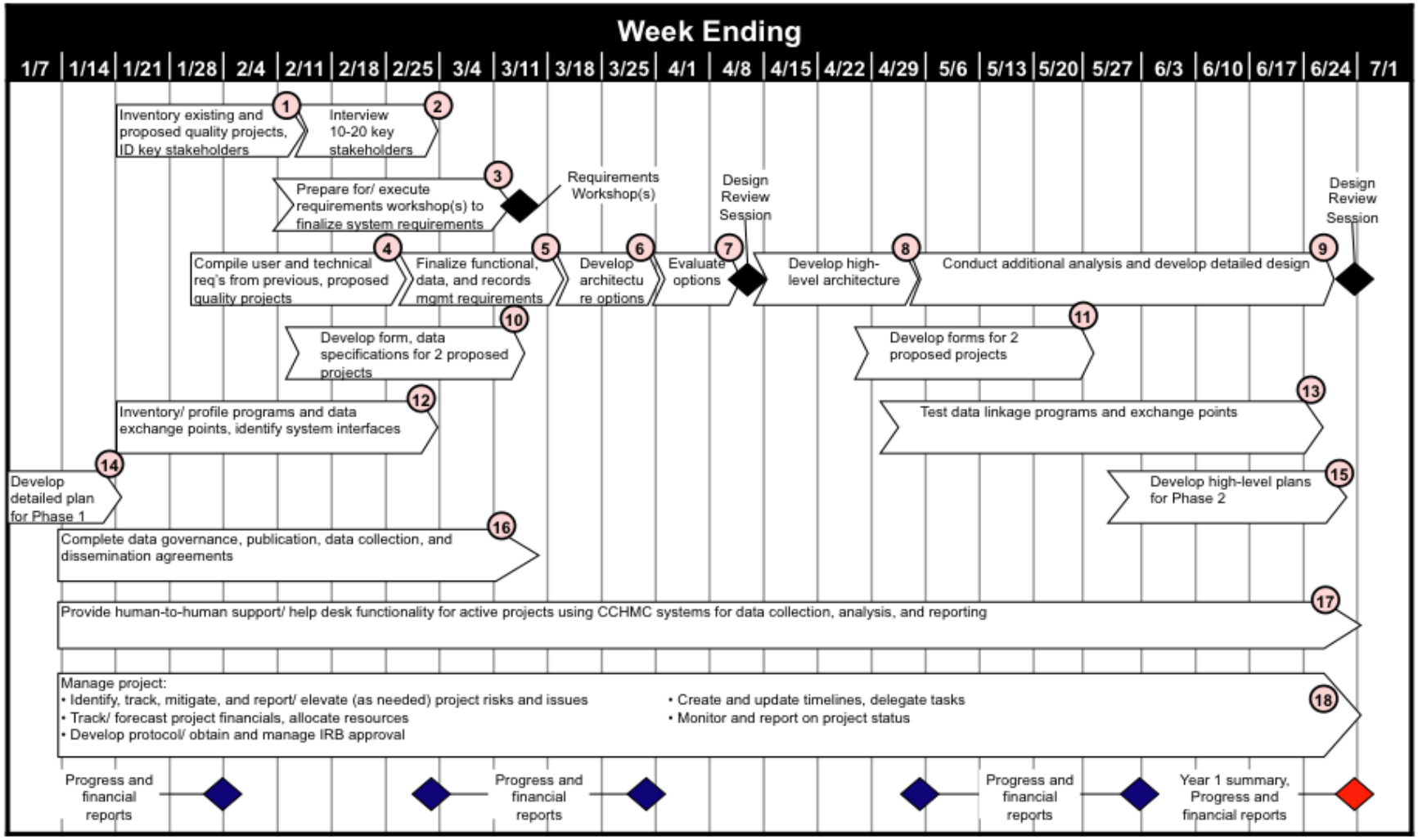


Figure 14: Project plan/ approach to Phase 1 of the Data Infrastructure Development Project.

10 Appendix 2 – Interview Guide

Data Infrastructure Development Project
Key Informant Interview Script

Background

1. *Remind the interviewee of the purpose for the interview.*
    a. We are developing a standard set of IT tools and processes to build a system that supports everyone and where each project is able to learn from others.
    b. We'd like to gather your input on system needs/ requirements, based on your experience with an improvement project in Ohio
        i. Because resources are limited and this is a system being developed for all projects, certain requirements may be compromised. We will take your requirements into consideration and grade them on a scale with all other requirements. The resulting scale will be used to determine which requirements will be included in the initial system.
2. Background: Interviewee
    a. Name
    b. Organization
    c. Title
3. On your BEACON project(s), what has been your level of involvement with the following groups of activities (High, Medium, Low – use to understand which sets of activities to focus on)
    a. Data Collection
    b. Analysis and Report Generation
    c. Communication and Collaboration

Understanding Interviewee's Project(s)

1. What project(s) do you work on?
2. What is your role on the project(s)?
3. For each project, what is/ are the
    a. Overall purpose
    b. Aim(s)
    c. Target population
    d. Start date
    e. (anticipated) End date
    f. Data sources
4. What tools does your project use for each set of activities?
    a. Data Collection
    b. Analysis and Report Generation
    c. Communication and Collaboration
5. Do you require that data from different sources be linked at the individual or provider levels?

6. Do you share tools with other statewide improvement projects in Ohio?

Known Requirements / Current Experiences *(repeat the following questions for each set of activities with which the interviewee is involved)*

1. What IT requirements/tools do you have right now that you wouldn't want to lose in the new system?
2. What are you using that you want/need to be included?

Future Requirements

1. What current IT needs on your projects that are currently not supported?

Follow-up

1. What are your concerns for establishing a standard set of tools and processes to support a statewide learning system in Ohio?
2. Do you have any questions for us?
3. Is there anything else we haven't discussed you think we should be thinking about?
4. Is there anyone else we should interview? If yes, who?
5. Is it OK to share the information/responses that you have provided with other interviewees and/or be posted online for others to review?

11 Appendix 3 – Interview Schedule

	Project	Interviewee - Key Stakeholder	Key Stakeholder Organization	Invitation Sent	Interview Date	Interviewee - Operational	Interviewer(s)	Invitation Sent	Scheduled (completed) date
1	CHIPRA Medicaid Encounter Data	Gerry Fairbrother	CCHMC	1/21/2011 1/27/2011	Cmpleted 2/16/10 2pm-3pm				
2	Dept. of Mental Health/ Psych	Beth Ferguson		1/27/2011 2/2/2011 2/10/11	2/24/11 1-2pm			1/27/11	
3	Solutions for Patient Safety	Steve Meuthing	CCHMC	1/27/11	Completed 2/17/11 10-11am	Missy Shepherd (PM)	James, Tim	1/27/11	Completed 2/3/11, 10-11am
						Ray Pruett (Analyst)	Brooke	1/21/11	Completed 2/1/2011, 3pm-4pm
4	OPQC - Neo	Ed	CCHMC		Completed 2/14/11	Matt Short (PM)	Tim	1/21/11	Completed 1/27/11, 2-3pm
						Ralph Brueggemann (Analyst)	Brooke	1/21/11	Completed 2/2/11, 9am-10am
						Vic Tawde (Developer)	Tim	1/21/11	Completed 1/31/11, 10am-11am
						Ron Bryson (Developer)	Tim	1/20/11	Completed 1/27/2011, 10am-11am
5	OPQC - OB					Divvie Powell (QIC)	Brooke	1/21/11	Completed 2/2/11, 1-2pm
6	Medtapp Development (Autism)	Carole Lannon	CCHMC	1/27/11	Completed 2/1	Kevin Stanford	James, Tim	1/21/2011 1/27/2011	Completed 2/3/11, 4-5pm
		John Duby	Akron	1/27/2011 2/10/11	Completed 3/1/11 1-1:30pm	Sandy Fuller	James, Tim	1/27/11	Completed 2/14/11, 10am-11am
		Melissa Arnold	AAP	1/27/2011 2/2/11 2/10/11 2/21/11	Completed 3/2/11 10am	Sang Sam	James, Tim	1/27/2011 2/10/11	Completed 2/17/11 2-3pm
7	Medtapp HKO (obesity)	Carole Lannon	CCHMC		Completed 2/1	Stacy Kramer	James, Tim	1/27/11	Completed 1/31/11
8	OCHA HKO OP	Melissa Arnold	AAP	1/27/11	Completed 3/2/11 10am				
		Amy Sternstein	Nationwide	1/27/11	Deferred to Bob Murray				
		Bob Murray	Nationwide	2/2/2011 2/10/11 2/21/11	Completed 2/23/11 10am				
		Samantha Anzeljc	Nationwide	2/2/2011 2/10/11 2/21/11	Completed 2/23/11 10am				
9	Asthma	Melissa Arnold	AAP		Completed 3/2/11 10am				

12 Appendix 4 – Must-Have System Requirements

12.1 Data Collection

Requirement ID	Category	Subcategory	Requirement Description
DC1	Data collection	Functionality	Permits Internet/ web-form entry into network DB
DC12	Data collection	Functionality	System allows group/ bulk submission of multiple forms (up to 50)
DC19	Data collection	Functionality	System to allow upload of a standard format file (export from EMR or other enterprise application).
DC20	Data collection	Functionality	System to allow import standard information (e.g., standard flat file upload template) Designate specified file types
DC21	Data collection	Functionality	System allows for data to be checked for accuracy, format, and completeness as it is entered; and prompts user to enter the correct data
DC24	Data collection	Functionality	System provides feedback on data that was just uploaded
DC25	Data collection	Functionality	System validates data upon upload and prompts user to update incorrect fields
DC26	Data collection	Functionality	Data extracted from EMR/ enterprise system/ secondary data source is translated to correct format and checked (values within range, unit conversion, completion, data type)
DC41	Data collection	Functionality	System tracks and documents which designated user inputs data for each entry
DC56	Data collection	Functionality	System can operate on multiple platforms
DC16	Data collection	Functionality	System to automatically save data upon input at 15 second intervals
DC30	Data collection	Functionality	System to disallow local modification of network-common forms/ fields
DC34	Data collection	Functionality	System to create unique identifiers for designated users
DC45	Data collection	Functionality	System has ability to monitor status of all data fields against validation rules

DC49	Data collection	Functionality	System allows for data to be checked for accuracy, format, and completeness after it is uploaded to the network database
DC50	Data collection	Functionality	System to preserve data integrity during upload
DC51	Data collection	Functionality	System can verify new data against stored data for accuracy
DC54	Data collection	Functionality	System automatically saves and backs-up all data stored to the network database
DC58	Data collection	Functionality	System supports automated collection, storage, retrieval, and transfer
DC67	Data collection	Functionality	System to allow local or network forms to access/ make use of standard or common measures definitions from reference materials
DC68	Data collection	Functionality	Logs and creates reports on errors (validation, upload, etc.)
DC69	Data collection	Functionality	System to capture/ link patient-level data (unique identifier for each patient that links multiple entries to a single patient)
DC14	Data collection	Functionality	System allow save of partially completed forms (allows return to complete at a later date)
DC17	Data collection	Functionality	System to allow creation of standard format file for data extraction from secondary data source or EHR
DC18	Data collection	Functionality	Ability to add upload modules to import secondary data to registry/database
DC32	Data collection	Functionality	System to limit changes to network forms to authorized users only
DC47	Data collection	Functionality	System allows designated resources to revise fields and entries after upload
DC64	Data collection	Functionality	System to include Message Board or Communication space where network managers can post user- or site-specific information (to be displayed upon login, must be acknowledged to move to Functionality, archived/ stored so users can view/ respond at a later date)

DC65	Data collection	Functionality	System to allow users to submit questions/ respond to questions from network manager (e.g., data definition, submission date, etc.)
DC71	Data collection	Functionality	System to allow standard forms to be pulled and placed for use by any project
DC9	Data collection	Performance	System response lag (e.g., refresh on webform) < 5 seconds
DC10	Data collection	Performance	Form submission < 15 seconds
DC11	Data collection	Performance	System to indicate delay/lag in time frame to execute
DC57	Data collection	Performance	System can support multiple networks (up to 100)
DC66	Data collection	Performance	System availability/ up-time to be > 95%, with outages resolved within 4 hours of report (std business hours)
DC33	Data collection	Security/compliance	System to limit access to authorized users or groups of users for Functionality and form modification
DC37	Data collection	Security/compliance	System to disallow simultaneous login/ use of the same account for a single network
DC43	Data collection	Security/compliance	System tracks and documents all uploads
DC44	Data collection	Security/compliance	System tracks and documents all validation checks and cleaning activities (entry specific)
DC46	Data collection	Security/compliance	System tracks and documents all exports (field-level detail)
DC48	Data collection	Security/compliance	System tracks and documents revisions to data (within database)
DC60a	Data collection	Security/compliance	System has ability to be in accordance with appropriate regulations for HIPAA
DC7	Data collection	User interface	System pre-populates fields based on earlier entries (form to form, pre-populates with info from network db)
DC8	Data collection	User interface	System provides selective questioning, based on earlier field entry (e.g., removes options/ fields that are no longer relevant) (skip patterns)
DC15	Data collection	User interface	System to notify/ remind authorized users of incomplete forms (saved, not submitted)

DC61	Data collection	User interface	System provides ability to modify the "look" of the user interfaces: color schemes, fonts, layouts, logos/graphics
DC70	Data collection	User interface	System provides ability to modify the "look" of the user interfaces: color schemes, fonts, layouts, logos/graphics by an intermediate user

12.2 User/System Administration

Requirement ID	Category	Subcategory	Requirement Description
SA1	System administration	Functionality	System to allow administration duties to be performed by authorized resources (creating/ deleting user accounts, modifying permissions). Administrative duties to be performed by intermediate users
SA3	System administration	Functionality	System to provide information (to designated users) on usage (last time logged in/ data entered, volume/ frequency, by site/ institution)
SA4	System administration	Security/compliance	System to allow user account permissions to be assigned for a finite or unlimited period of time
SA5	System administration	Security/compliance	System to capture requests for user/ permission changes and forward to authorized resources
SA6	System administration	Security/compliance	System to automatically notify designated resources about upcoming account expirations
SA7	System administration	Security/compliance	System shows all current users and their permissions
SA9	System administration	Security/compliance	System to support multiple layers of permissions (admin, intermediate user (edit content), read only)

12.3 Reporting & Analysis

Requirement ID	Category	Subcategory	Requirement Description
DAR1	Data analysis	Functionality	System to allow complex analysis of data at patient and/ or practice level (Desired system capabilities: sort patients by due date,/ severity classification, date of last visit; compare performance by practice/ clinician; 1-system w. separate db required to handle both levels of data; second db is likely source to collect secondary data collected (e.g. survey monkey)
DAR5	Data analysis	Functionality	System to allow creation of clinician-, site-, or patient-specific reports, including patient and aggregate data
DAR6	Data analysis	Functionality	System to allow analysis/ manipulation/ presentation of qualitative and quantitative data
DAR12	Data analysis	Functionality	Reports/ analysis available in multiple formats (e.g., .pdf and others) that can be read on multiple platforms (computer, smartphone)
DAR14	Data analysis	Functionality	System to allow presentation of graphical and tabular data (to include display of data over time, e.g., run charts)
DAR15	Data analysis	Functionality	System to allow comments/ narrative to accompany graphical data representations ("write-on" capabilities) (annotation)
DAR16	Data analysis	Functionality	System to allow creation/ modification of custom reports by expert user (graphical and tabular display with commentary/ narrative)
DAR21	Data analysis	Functionality	System to allow creation/ modification of tabular/ graphical reports by an expert user
DAR22	Data analysis	Functionality	System to allow analysis/ presentation of current and/ or archived data
DAR3	Data analysis	Functionality	Ability to generate extract to look at root-level data
DAR7	Data analysis	Functionality	System to allow export of data to external/ third party analytical packages (e.g., SAS), expert user to be able to create/ modify format/ structure of the export file

DAR8	Data analysis	Functionality	System to allow export of reports for upload into external systems (e.g., EMRs); expert user to be able to create/ modify format/ structure of the export file
DAR10	Data analysis	Functionality	System to allow import/ upload of graphical outputs from external systems for inclusion in reports/ analysis for read-only inclusion
DAR13	Data analysis	Functionality	System to allow automated publication/ push of approved reports to email, web
DAR17a	Data analysis	Functionality	System to allow distribution of reports based on individually set preferences (identified/de-identified) to website
DAR18	Data analysis	Functionality	System to allow graphical presentation of performance vs. benchmarks or standards (predetermined or calculated benchmarks)
DAR19	Data analysis	Functionality	System to allow automated execution of pre-defined analysis/ manipulation of a dataset (create 'standard' report upon receipt of clean data)
DAR23	Data analysis	Functionality	System to allow validation/ checking of data by authorized users
DAR24	Data analysis	Functionality	System to allow modification of data in database by authorized users with proper training
DAR25	Data analysis	Functionality	Logs and creates reports on errors (validation, import, export, etc.)
DAR29	Data analysis	Functionality	System to allow for creation standard report templates
DAR30	Data analysis	Functionality	System to allow for design of standard report templates
DAR31	Data analysis	Functionality	System can identify and store a snapshot of data used for analysis for a predetermined amount of time for a subset for the dataset
DAR33	Data analysis	Functionality	Reports are printable from system in a standard format
DAR11	Data analysis	Security/compliance	System tracks and documents revisions to data (within database) - audit trail
DAR28	Data analysis	Security/compliance	System tracks and documents which designated user inputs/ modifies data for each entry
DAR2	Data analysis	User interface	System to allow aggregation or drilldown of data by an intermediate user (e.g., site-specific resource, QIC, etc.)

12.4 Communication & Collaboration

Requirement ID	Category	Subcategory	Requirement Description
CC22	Communication / collaboration	Functionality	System to allow static HTML pages
CC23	Communication / collaboration	Functionality	System to allow dyanmic HTML pages
CC24	Communication / collaboration	Functionality	System to allow wiki pages
CC25	Communication / collaboration	Functionality	System to allow calendars
CC29	Communication / collaboration	Functionality	System to allow task lists
CC30	Communication / collaboration	Functionality	System to allow discussion boards/blogs
CC33	Communication / collaboration	Functionality	System to allow project team and individuals to create customized alerts
CC37	Communication / collaboration	Functionality	System to allow file tagging/annotation
CC39	Communication / collaboration	Functionality	System to allow creation of metadata fields and require/ not require assignment of those fields to files
CC51	Communication / collaboration	Functionality	System to allow intermediate user to assign users to multiple networks
CC52	Communication / collaboration	Functionality	System to allow authorized users to delete/modify content posted on the message boards
CC53	Communication / collaboration	Functionality	System to allow authorized users to "rate" tools, documents, and comments posted to the site
CC59	Communication / collaboration	Functionality	System to allow users to set up profiles
CC61	Communication / collaboration	Functionality	System to allow for content to be searched
CC63	Communication / collaboration	Functionality	System to allow intermediate user to create distribution lists
CC5	Communication / collaboration	Functionality	System to create unique identifiers for designated users
CC6	Communication / collaboration	Functionality	Collaboration space to include functionality to check-in/ check-out documents (disallow simultaneous edits to files)
CC7	Communication / collaboration	Functionality	System automatically saves and backs-up all data stored to collaboration space with a specified frequency
CC18	Communication / collaboration	Functionality	System to allow large file storage/sharing (>100MB)
CC19	Communication / collaboration	Functionality	System to allow small file storage/sharing (<100MB)
CC20	Communication / collaboration	Functionality	System to allow file versioning

CC21	Communication / collaboration	Functionality	System to allow access control/permission management
CC34	Communication / collaboration	Functionality	System to allow custom workflows (e.g., form routing)
CC35	Communication / collaboration	Functionality	System to allow public/anonymous access
CC36	Communication / collaboration	Functionality	System to allow internal/secured access
CC49	Communication / collaboration	Functionality	System to check that a user does not already have an account
CC56	Communication / collaboration	Functionality	System allows links to be embedded
CC64	Communication / collaboration	Functionality	System to allow reports to automatically be posted (from database) to designated area in collaboration space
CC65	Communication / collaboration	Functionality	System to allow reports to be produced in different formats and distributed to user groups
CC66	Communication / collaboration	Functionality	System to allow user to select their preference for receiving reports (fax, email, internal email)
CC67	Communication / collaboration	Functionality	System to allow reports to be archived
CC70	Communication / collaboration	Functionality	System can be simultaneously accessed by up to 10000 individuals from up to 1000 institutions
CC71	Communication / collaboration	Functionality	System has ability to send calendar invites that are interoperable with Groupwise and Outlook (.ics file)
CC69	Communication / collaboration	Functionality	System to allow site information to be directly linked to within communications (i.e., emails, alerts, etc)
CC4	Communication / collaboration	Security / compliance	System to limit access to authorized users or groups of users for functionality and collaboration space modification
CC9	Communication / collaboration	Security / compliance	System to disallow simultaneous login/ use of the same account
CC13	Communication / collaboration	Security / compliance	System tracks and documents which designated user inputs data for each entry (audit trail)
CC15	Communication / collaboration	Security / compliance	System tracks and documents all uploads
CC57	Communication / collaboration	Security / compliance	System to allow permissions to be set by page and by file
CC1	Communication / collaboration	User interface	System to provide collaboration space with a pre-specified structure (e.g., folder structure, layout, look/ feel)

CC2	Communication / collaboration	User interface	System to allow modification of collaboration space by intermediate user (through graphical interface/ drag and drop; does not require coding)
CC31	Communication / collaboration	User interface	System to allow data collection forms to be accessible through the collaboration space
CC32	Communication / collaboration	User interface	System to allow RSS feeds
CC38	Communication / collaboration	User interface	System to allow file metadata analysis
CC44	Communication / collaboration	User interface	System to allow intermediate user to modify external website (look/ feel, colors, content updates such text, photos, etc.)
CC50	Communication / collaboration	User interface	System to allow users access to multiple networks
CC54	Communication / collaboration	User interface	System to display ratings/ frequency of access (most popular content)

13 Appendix 5 – Example Form

Harm Index

Click Here for Instructions

Month: MAY **Year:** 2011

* **All Numerators required. Please enter "0" if value is unknown.**

HAIs - Hospital Acquired Infections

	Numerator		Denominator
SSI - Neuro Primary Shunts		Neuro Primary Shunt Cases	
SSI - Orthopaedic		Orthopaedic Cases	
SSI - Cardiothoracic		Cardiothoracic Cases	
SSI - Any Others (Includes Neuro Revisions)		Any Other (Includes Neuro Revisions) Cases	
VAP - Ventilator Associated Pneumonia		Vent Days	
CA-BSI - Catheter Associated Blood Stream Infections		Line Days	
CA-UTI - Catheter Associated Urinary Tract Infections		Line Days	

Preventable Adverse Drug Events (Medication Errors are included)

	Numerator		Denominator
ADE - Level 5		Patient Days	
ADE - Level 6			
ADE - Level 7			
ADE - Level 8			
ADE - Level 9			

PU - Pressure Ulcers

	Numerator		Denominator
PU - Stage 2		Patient Days	
PU - Stage 3			
PU - Stage 4			

PIV - Infiltrates

	Numerator		Denominator
PIV - Grade 2		Line Days	
PIV - Grade 3			
PIV - Grade 4			
PIV - Grade 4D			
PIV - Grade 4V			

Medical Response Team Preventable Codes Outside the ICU

	Numerator		Denominator
MRT Preventable Codes			

Serious Falls

	Numerator		Denominator
Serious Falls			

SSE - Serious Safety Events (Preventable)

	Numerator		Adjusted Patient Days
SSE Level 1		Adjusted Patient Days	
SSE Level 2			
SSE Level 3			
SSE Level 4			
SSE Level 5			

CRF Status Update and Save Options

Save and Close | Close Without Saving

Form Version: 1.0.0.372

14 Appendix 6 – Example Data Dictionary

Database Column Name	Section Title	Form Label (Column)	Database Datatype	Validation / Comments
formid	na	na	int	Database row identity field (Primary Key)
hospitalid	na	na	varchar(50)	Three letter hospital/site identifier
msmonth	na	Month	varchar(50)	1-2 digit month number
msyear	na	Year	varchar(50)	4 digit year number
hihainum1	HAIs - Hospital Acquired Infections	SSI - Neuro Primary Shunts (Numerator)	int	
hihainum2	HAIs - Hospital Acquired Infections	SSI - Orthopaedic (Numerator)	int	
hihainum3	HAIs - Hospital Acquired Infections	SSI - Cardiothoracic (Numerator)	int	
hihainum4	HAIs - Hospital Acquired Infections	SSI - Any Others (Includes Neuro Revisions) (Numerator)	int	
hihainum5	HAIs - Hospital Acquired Infections	VAP - Ventilator Associated Pneumonia (Numerator)	int	
hihainum6	HAIs - Hospital Acquired Infections	CA-BSI - Catheter Associated Blood Stream Infections (Numerator)	int	
hihainum7	HAIs - Hospital Acquired Infections	CA-UTI - Catheter Associated Urinary Tract Infections (Numerator)	int	
hihaiden1	HAIs - Hospital Acquired Infections	Neuro Primary Shunt Cases (Denominator)	int	
hihaiden2	HAIs - Hospital Acquired Infections	Orthopaedic Cases (Denominator)	int	

hihaiden3	HAIs - Hospital Acquired Infections	Cardiothoracic Cases (Denominator)	int	
hihaiden4	HAIs - Hospital Acquired Infections	Any Other (Includes Neuro Revisions) Cases (Denominator)	int	
hihaiden5	HAIs - Hospital Acquired Infections	Vent Days (Denominator)	int	
hihaiden6	HAIs - Hospital Acquired Infections	Line Days (Denominator)	int	
hihaiden7	HAIs - Hospital Acquired Infections	Line Days (Denominator)	int	
hipadenum1	Preventable Adverse Drug Events (Medication Errors are included)	ADE - Level 5 (Numerator)	int	
hipadenum2	Preventable Adverse Drug Events (Medication Errors are included)	ADE - Level 6 (Numerator)	int	
hipadenum3	Preventable Adverse Drug Events (Medication Errors are included)	ADE - Level 7 (Numerator)	int	
hipadenum4	Preventable Adverse Drug Events (Medication Errors are included)	ADE - Level 8 (Numerator)	int	
hipadenum5	Preventable Adverse Drug Events (Medication Errors are included)	ADE - Level 9 (Numerator)	int	

hipadeden1	Preventable Adverse Drug Events (Medication Errors are included)	Patient Days (Denominator)	int	
hipunum1	PU - Pressure Ulcers	PU - Stage 2 (Numerator)	int	
hipunum2	PU - Pressure Ulcers	PU - Stage 3 (Numerator)	int	
hipunum3	PU - Pressure Ulcers	PU - Stage 4 (Numerator)	int	
hipuden1	PU - Pressure Ulcers	Patient Days (Denominator)	int	
hipivnum1	PIV - Infiltrates	PIV - Grade 2 (Numerator)	int	
hipivnum2	PIV - Infiltrates	PIV - Grade 3 (Numerator)	int	
hipivnum3	PIV - Infiltrates	PIV - Grade 4 (Numerator)	int	
hipivnum4	PIV - Infiltrates	PIV - Grade 4D (Numerator)	int	
hipivnum5	PIV - Infiltrates	PIV - Grade 4V (Numerator)	int	
hipivden1	PIV - Infiltrates	Line Days (Denominator)	int	
hinicanum1	Non-ICU Cardiac Arrests	Non-ICU Cardiac Arrests (Numerator)	int	
hinicaden1	Non-ICU Cardiac Arrests	(Denominator)	int	
hisfnum1	Serious Falls	Serious Falls (Numerator)	int	
hisfden1	Serious Falls	(Denominator)	int	
hissenum1	SSE - Serious Safety Events (Preventable)	SSE Level 1 (Numerator)	int	
hissenum2	SSE - Serious Safety Events (Preventable)	SSE Level 2 (Numerator)	int	
hissenum3	SSE - Serious Safety Events (Preventable)	SSE Level 3 (Numerator)	int	

hissenum4	SSE - Serious Safety Events (Preventable)	SSE Level 4 (Numerator)	int	
hissenum5	SSE - Serious Safety Events (Preventable)	SSE Level 5 (Numerator)	int	
hisseden1	SSE - Serious Safety Events (Preventable)	10K Adjusted Patient Days (Adjusted Patient Days)	int	
FormUniqueID	na	na	varchar(50)	UniqueID for form submitted xml
username	na	na	varchar(50)	AD display for login that created/modified entry
userlogin	na	na	varchar(50)	Network login for account that created/modified entry
lastmodified	na	na	datetime	datetime form was last modified
createdate	na	na	datetime	datetime form was first submitted to database
Notes:	Database Column naming convention is as follows: HI (Harm Index)+First Letters of Section Title+Numerator (num)/Denominator (den)+Order Number. For example, hihainum1 = Harm Index + Hospital Acquired Infections + Numerator + Row 1			