Data Infrastructure Development Project

Conceptual Architecture Design Document
# Table of Contents

1 Background ................................................................................................. 5

2 Overview ..................................................................................................... 6

3 Project Goals ............................................................................................. 7

4 Critical Success Factors ............................................................................. 7

5 Scope .......................................................................................................... 7

6 Conceptual Architecture .......................................................................... 7

## 6.1 Data Collection ................................................................................... 9

### 6.1.1 Data collection requirements ......................................................... 9

### 6.1.2 InfoPath/SharePoint .................................................................... 9

### 6.1.3 Satisfying must-have requirements ............................................. 10

#### 6.1.3.1 Web-based data entry .............................................................. 10

#### 6.1.3.2 Client-side data validation ..................................................... 12

#### 6.1.3.3 Skip logic/branching and repeating sections ....................... 13

#### 6.1.3.4 User interface customization .................................................. 13

#### 6.1.3.5 Reusability .......................................................................... 14

#### 6.1.3.6 Data pre-population ............................................................... 15

#### 6.1.3.7 Secondary data integration .................................................... 15

#### 6.1.3.8 Audit trails ........................................................................... 15

#### 6.1.3.9 High availability/performance ............................................... 16

#### 6.1.3.10 Scalability .......................................................................... 16

### 6.1.4 Data storage methodology ........................................................... 17

#### 6.1.4.1 InfoPath XML ................................................................... 17

#### 6.1.4.2 Web service model ............................................................... 18

#### 6.1.4.3 Database model ................................................................. 19

#### 6.1.4.4 Advantages of web service/database model ....................... 20

### 6.1.5 Supporting documentation .......................................................... 21

## 6.2 System/user administration ............................................................. 24

### 6.2.1 System/user administration requirements .................................. 24

### 6.2.2 Identity and access management (IAM) ...................................... 24

### 6.2.3 SharePoint .................................................................................. 25

### 6.2.4 Satisfying must-have requirements ............................................. 26

#### 6.2.4.1 Streamlined user account creation ...................................... 26

#### 6.2.4.2 Multiple layers of security .................................................... 27

#### 6.2.4.3 Direct control of user access/permissions ............................. 27

#### 6.2.4.4 Account/permissions/usage tracking .................................... 28

## 6.3 Reporting ........................................................................................... 30

### 6.3.1 Reporting Requirements ............................................................ 30

### 6.3.2 SQL Server Reporting Services ............................................... 31

### 6.3.3 Satisfying must-have requirements ............................................. 31

#### 6.3.3.1 Reporting Authoring ............................................................ 32

#### 6.3.3.2 Report Templates ................................................................. 32

#### 6.3.3.3 Access Control ................................................................. 38

#### 6.3.3.4 Report Deployment Process ............................................... 38

## 6.4 Data Analysis and Secondary Use ................................................ 39

### 6.4.1 Data Analysis and Secondary Use Requirements ..................... 39

### 6.4.2 SQL Server & SQL Server Integration Services ....................... 39
Table of Figures

FIGURE 1: APPROACH TO THE PLANNING PHASE OF THE DATA INFRASTRUCTURE DEVELOPMENT PROJECT. 5

FIGURE 2: CONCEPTUAL SYSTEM ARCHITECTURE. DATA IS ENTERED VIA INFOPath AND STORED IN SQL Server. THE DATA FROM DIFFERENT SOURCES IS PULLED TOGETHER USING SSIS AND DISPLAYED USING SSRS. 8

FIGURE 3: INFOPath FORM BUILDER 10

FIGURE 4: EXAMPLE INFORMED CONSENT FORM BUILT USING INFOPath 11

FIGURE 5: EXAMPLE DATA VALIDATIONS ON AN INFOPath FORM 12

FIGURE 6: TOOLTIP EXPLANATION OF FIELD-LEVEL DATA VALIDATION 12

FIGURE 7: EXAMPLE OF A FORM WITH A REPEATING SECTION 13

FIGURE 8: A SharePoint site with a customized layout 14

FIGURE 9: UPDATING FIELDS IN AN INFOPath XML DATA SOURCE 17

FIGURE 10: WEB SERVICE WORKFLOW 19

FIGURE 11: SIMPLIFIED DIAGRAM OF THE DATA MODEL USED TO SUPPORT DATA COLLECTION 20

FIGURE 12: TECHNICAL FLOW OF DATA DURING THE DATA COLLECTION PROCESS 21

FIGURE 13: TECHNICAL FLOW OF DATA DURING DATA COLLECTION PROCESS (EXPANDED VIEW) 23

FIGURE 14: ARCHITECTURE OF CCHMC’s NEW RESEARCH NETWORK 25

FIGURE 15: SharePoint CONTEXT MENU 26

FIGURE 16: WORKFLOW TO CREATE AND PROVISION USER ACCOUNTS 27

FIGURE 17: ASSIGNING GROUP PERMISSIONS TO USERS IN SharePoint 28

FIGURE 18: 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 33

FIGURE 19: RUN CHART WITH DATA TABLE. THIS REPORT SHOWS A CHART WITH VARIOUS MEASURES, ALONG WITH THE DATA THAT COMPRISE EACH LINE 34

FIGURE 20: CONTROL CHART WITH ANNOTATIONS. THIS GRAPH ILLUSTRATES HOW TEXT ANNOTATIONS CAN BE ADDED TO A GRAPH IN ORDER TO DENOTE SPECIAL CAUSES (MARKED BY YELLOW ASTERISK) 35

FIGURE 21: SITE SPECIFIC DASHBOARD. THE 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 35

FIGURE 22: 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 36

FIGURE 23: SSRS REPORTS AND DATA CAN BE EXPORTED IN A NUMBER OF FORMATS. CHARTS CAN ALSO BE CONFIGURED TO DISPLAY ONLY CERTAIN TIME PERIODS (FUNCTIONALITY NOT SHOWN) 37

FIGURE 24: TABULAR REPORTS CAN BE CREATED AS A MECHANISM TO ALLOW FOR ON DEMAND DATA EXTRACTS TO AUTHORIZED USERS 37

FIGURE 25: I2B2 WORKBENCH DEVELOPED BY CCHMC. SCREEN SHOWS VIEW OF AN OBFUSCATED ACCESS USER 47

FIGURE 26: ELEVATED ACCESS MODE OF THE WORKBENCH 48

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.” 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 12).

![Figure 1: Approach to the planning phase of the data infrastructure development project.](image)

The team conducted interviews with 23 stakeholders representing nine BEACON projects to understand their projects’ needs and technical requirements (See Appendix 2 [Section 13] for the interview guide and Appendix 3 [Section 14] 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 15] 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 detailed design presented here.

2 Overview

This document describes an architecture designed to support 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. While it is understood that each project will have its own set of unique constraints, there will be crosscutting commonalities. The proposed system allows for each project to create customized workflows as well as a customized look-and-feel, but the underlying components remain the same. Using plumbing as a metaphor, everyone in the building shares the same pipes, but residents are free to choose their own fixtures.

The approach laid out in this design document is influenced by the electronic data capture (EDC) systems developed by the Division of Biomedical Informatics (BMI) at the Cincinnati Children’s Hospital Medical Center (CCHMC) for the Ohio Perinatal Quality Collaborative (OPQC) and Solutions for Patient Safety (SPS) projects. Due to the ad hoc nature of how these projects ended up in the BMI development pipeline, and the lack of clarity about whether there would be similar projects in the future, they were largely developed as “one-off” productions, without much effort put into standardization or reusability. The architecture proposed here builds on the learning of those projects, but addresses the shortcomings in the existing system when it comes to reusability. The proposed infrastructure is intended for all new BEACON networks, but existing networks can be migrated into the system as new data collection projects arise.

The idea behind this infrastructure 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 will 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 is designed to 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 Critical Success Factors

The most critical element to the success of this system is the requirement, or at least the encouragement, that all BEACON projects use the platform, unless there are legitimate reasons why it would not be sufficient. While the data infrastructure could be considered a success even if it was only used by one project, the true value of this endeavor comes from the economies of scale gained from supporting multiple projects, specifically from the reuse of common elements like forms, site layouts and report templates.

5 Scope

The general system features and overall structure and design are considered to be in scope for this phase, as illustrated using examples from OPQC and SPS. Out of scope is the implementation plan, i.e., the specific steps needed to build the system described here. In order to prepare such a document, detailed project requirements are needed, such as field definitions, and the layout, content and logic of the project’s forms and reports. That information was not collected during this planning phase and would be obtained during the infrastructure build phase. 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. Those documents would be drafted as the system was developed. A detailed discussion on the collaboration tools provided by SharePoint is also out of scope. Since most of the BEACON projects were already using a collaboration space of some kind, the focus of this planning phase was more on the other data management features (data collection, reporting, data linking) that lacked standardization.

6 Conceptual Architecture

The proposed system architecture 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. It takes advantage of several Microsoft technologies:

- InfoPath – a forms authoring tool with an Office-like form builder.
- SharePoint – a collaboration space where users can create project websites (internal and external), post documents, create calendars, host discussion boards and much more.
• SQL Server – a robust database back-end that can support both large and small projects as well as SharePoint.
• SQL Server Reporting Services – 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.
• SQL Server Integration Services – 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.

In addition, the system includes custom software developed by BMI to support other electronic data collection projects and can also support a data warehouse based on the i2b2 platform, if there is a desire to allow BEACON end users to perform de-identified cohort identification. Due to the cost and complexity of building the data warehouse, and since most of the existing BEACON IRB protocols do not allow such data linking and integration, it is not necessary to include it in the first few versions of the system, but the option remains for future phases of development.

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 description of each component follows the same general outline: a list of “must-have” requirements as approved by the BEACON Technical Committee, an explanation of how the proposed technology meets those requirements, and a greater description of certain features of the underlying technology.

Figure 2: Conceptual system architecture. Data is entered via InfoPath and stored in SQL Server. The data from different sources is pulled together using SSIS and displayed using SSRS.
6.1 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, we propose an enterprise-scale solution that can be used to handle statewide projects with the capability for future growth.

6.1.1 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 15.1).

6.1.2 InfoPath/SharePoint

To satisfy these requirements, two widely supported technologies from Microsoft will be employed: InfoPath, a tool for building questionnaires, and SharePoint, a platform for hosting web-based InfoPath questionnaires and integrating with other features to create interactive web sites.

InfoPath is a component of the Microsoft Office Professional Plus Suite. It enables users without programming knowledge or abilities to build online and offline forms. Using a graphical user interface similar to Microsoft Word, users can drag and drop common controls – for example, text boxes, check boxes or radio buttons – onto a canvas, where they can add labels and adjust fonts, background colors and other user interface elements. Here is an example of the InfoPath form-building canvas:
Figure 3: InfoPath Form Builder

InfoPath forms can be published to SharePoint, a server-based technology that enables users – again, without any programming skills – to create secure, team- or project-oriented web sites. In most cases, users must enter a user name and password to enter a SharePoint site. There they may find a mix of content, some of which may be read-only and others to which they can actively contribute. For example, the site may contain a message board where members can post and answer questions. Or it may contain a library of documents, which members can collaboratively develop by checking out files, making changes and checking them back in. For data entry-oriented projects, users may find a list of InfoPath questionnaires, which they may open, complete and submit. A list of completed and partially completed forms may also be displayed so that combined with configurable email alerts, users may easily track the progress of their work.

6.1.3 Satisfying must-have requirements

InfoPath and SharePoint are able to satisfy all must-have data collection requirements of the BEACON data infrastructure. Many of these requirements will be met by out-of-box functionality; others will be met with the use of custom programming and add-on technologies. Indeed, Microsoft has developed both InfoPath and SharePoint in a way that enables them to be easily extended by IT professionals, and as a result a wide array of third-party tools and services are available to augment out-of-box functionality. While we propose to limit the number of third-party add-ons and in-house customizations as a way to mitigate risks associated with licensing, maintenance/upgrades and support, we recognize from several years of experience working with these technologies there are also risks in forgoing the possibilities of custom development and in ignoring the vast contributions of the third-party community.

6.1.3.1 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.

Although InfoPath forms are created offline and may be used offline by users with select versions of Microsoft Office (containing a product known as InfoPath Filler), they also can be published to a SharePoint site. Users can open InfoPath browsers from any mainstream platform (Windows, Mac, Linux) or browser (Internet Explorer, Firefox, Safari). Forms may be saved and re-opened later for additional data entry. The connection between the user’s browser and the SharePoint server is 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. For more details about security and compliance, see the section on system/user administration (Section 0).

Here is an example of an informed consent form developed in InfoPath, published to SharePoint and loaded in Internet Explorer:

![Figure 4: Example informed consent form built using InfoPath.](image-url)
6.1.3.2 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.

Using InfoPath, project staff – without the assistance of IT staff – can add validation checks to forms by using an intuitive interface similar to other Microsoft Office applications. 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 5: Example data validations on an InfoPath form.](image)

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

![Figure 6: Tooltip explanation of field-level data validation.](image)
Other validation options 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. Both hard and soft checks will be added to the BEACON data collection forms, although careful consideration will be made to strike an appropriate balance between data integrity and system usability. Past experience has shown that too much hard validation, in particular, may frustrate users and lead to decreased responsiveness.

6.1.3.3 **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. InfoPath enables the creation of such patterns, again through the use of the InfoPath 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 7: Example of a form with a repeating section.](image-url)

6.1.3.4 **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. One of the main advantages of InfoPath over alternative form-building technologies is the ease with which form presentation can be changed. For example, a form can be designed in Microsoft Word and then imported directly into InfoPath, which will retain any formatting. Within InfoPath,
elements can be dragged and dropped around a canvas, and fonts, colors and other qualities can be changed using the same menus/commands as in Microsoft Word.

It may also be desirable to customize the look-and-feel of the SharePoint site where InfoPath forms are hosted. SharePoint offers 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 requires some knowledge of web design technologies but no actual programming). Like InfoPath forms, SharePoint enables users to manipulate page layout by dragging and dropping objects – including InfoPath forms themselves. Below is an example of a SharePoint site customized to support a clinical trial:

![SharePoint site with a customized layout](image)

**Figure 8: A SharePoint site with a customized layout.**

### 6.1.3.5 Reusability

The need to be able to reuse components such as questionnaires across multiple projects is an important cost savings requirement that InfoPath supports well. Assuming agreement on the makeup of standardized instruments – for example, a demographics form – may be reached, corresponding InfoPath templates will be developed. Even if a template doesn't contain all the fields a given project needs, it may be used 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 SharePoint already available as the central
BEACON platform, there are no additional costs to providing the ability to share form templates across the network; they simply may be uploaded to a central library where they can be downloaded and used by any BEACON project.

6.1.3.6 Data pre-population

In some more advanced forms, it may be desirable 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 is built into InfoPath and may be implemented by any user, the latter functionality requires some customization. However, through the use of a simple web service written by a developer (read more about InfoPath and web services in section 6.1.4.2) a non-developer may retrieve values from anywhere in the data set and display dynamically on their InfoPath forms.

6.1.3.7 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 6.4). The proposed InfoPath/SharePoint 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 InfoPath/SharePoint. This is made possible by the fact that, as discussed in section 6.1.4, InfoPath form data is stored in a relational SQL Server database, from which it can easily be exported to other systems – for example, a SAS program or the i2b2 research data warehouse – and combined with other data sets. Such exports could be manually executed or could be automated using SQL Server Integration Services. Another option is made possible by what is known as the Business Connectivity Services (BCS) within SharePoint, which essentially provides the capability to surface and even manipulate data from external systems without actually importing it. For example, a SharePoint project web site could be configured to connect to an external database, fetch the data and display it to the user in a browser – perhaps alongside other data collected via different means (e.g., through InfoPath forms).

The preferable method of integration may vary depending on the requirements of a given project. But a rule of thumb is that the SharePoint BCS will be used for simple integration in the user interface and that SSIS and/or the i2b2 research data warehouse will be used for more complex integration to facilitate advanced analysis.

6.1.3.8 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, InfoPath/SharePoint provide the ability to retain version histories of all documents including InfoPath forms. In addition, several third-party add-ons are available to capture more detailed audit trails of InfoPath/SharePoint content. One example is the DocAve Auditor from AvePoint ([http://www.avepoint.com/sharepoint-auditing-docave/](http://www.avepoint.com/sharepoint-auditing-docave/)), which documents all user interactions with a SharePoint site including InfoPath forms.
To fulfill the compliance requirements of the BEACON infrastructure, this product or a similar competitor will be purchased and installed on the BEACON SharePoint server.

In addition to the SharePoint-level audit trails, a separate audit trail also will be created in SQL Server where, for each InfoPath form, a database trigger will be implemented to capture any changes to the data (including the date/time, operator, old value and new value). See section 6.1.4.3 for more details about database-level auditing. This dual approach to audit trails will satisfy regulatory requirements such as 21 CFR Part 11, the standard guidance for FDA-regulated clinical trials.

6.1.3.9 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 SharePoint infrastructure, or farm, will 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 in the farm, 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 the farm. In combination, these two measures should maximize the responsiveness of the servers – for example, when an InfoPath form is being loaded or submitted – and should minimize downtime to the 60 seconds or less it takes for one server to fail over to another.

Maximizing performance will also be a consideration in design of the InfoPath forms. On the basis of several years of working with this technology, several techniques have been developed to minimize the number of post-backs, or calls to the server, without sacrificing functionality. These measures are discussed in greater detail in section 6.1.5.

6.1.3.10 Scalability

A final challenging 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. Once again, the ability to cluster both SharePoint application servers and SQL Server database nodes will meet this requirement. Cost considerations notwithstanding, additional application and database servers can always be added to the farm to increase capacity. To start, the farm will consist of three application servers and a clustered SQL Server environment with at least five nodes (servers), all of which will exceed the minimum Microsoft recommendations for processing, RAM, and storage (see http://technet.microsoft.com/en-us/library/cc262485.aspx#section2). It is expected that this configuration, summarized in the following table, will be able to support at least two dozen medium-sized networks.
6.1.4 Data storage methodology

Ultimately the data collection forms are only as effective as the data storage methods behind them. Accordingly, an approach will be used to maximize the security, integrity and portability of the BEACON primary data set(s) while also maximizing options for integrating the primary data with secondary data such as Vermont Oxford Network (VON) and vital statistics data. Understanding this methodology first requires a basic understanding of how InfoPath natively stores form data.

6.1.4.1 InfoPath XML

As users develop forms in InfoPath, they map each field to an element in what is called the data source, which is essentially a graphical depiction of the form’s XML schema. During this process, form designers also give each field/XML node a more meaningful name than the default name of fieldN. In the following example, a field is renamed “Patient_Name,” which updates the corresponding node in the XML schema:

![Figure 9: Updating fields in an InfoPath XML data source.](image)
By default, when an InfoPath form is submitted, an XML document is generated that conforms to the defined schema, and this document is stored in a document library on a SharePoint site. For basic data collection projects that do not require extensive reporting or analysis, this default storage method is sufficient. Security is accomplished by the fine-grained permissions management abilities within SharePoint (see section 0 for additional details), and audit trails can be accomplished through the use of versioning or the installation of third-party add-ons (see section 2.3.8 for details).

For projects involving multiple, larger data sets and advanced reporting, integration or analysis a more robust data storage method is recommended. Accordingly, BEACON data collected using InfoPath forms will be stored in a SQL Server relational database. Specifically, the InfoPath XML document will be submitted to a web service, which will populate the data in the appropriate database table(s).

6.1.4.2 Web service model

A web service is software that enables interaction between two systems that are otherwise distinct. In the case of InfoPath, a web service may be used to link a form to a SQL Server database. Specifically, the web service will interpret the InfoPath XML, match the form schema to the database schema and copy the data to the appropriate database tables/fields.

Submitting InfoPath forms to such a web service has several advantages, especially for data collection projects of a longitudinal nature. First, it is easier to generate reports of the data set (using technologies such as SQL Server Reporting Services, described in section 6.3). Second, it is easier to export the data – for example, to a third-party analysis tool such as SAS. Third, it is easier to combine the data with other sources, which may be accomplished directly in SQL Server (by using SQL Server Integration Services to create a data mart and SQL Server Reporting Services to create reports – see section 6.4 for details) or by importing the data into a warehouse environment such as i2b2 (see section 6.5 for details).

While it is true that a developer generally must write a web service, once the web service is deployed to a given SharePoint site, any user may submit a form to it. Moreover, to minimize developer involvement, the BEACON web service model will be generalized so that it may be used for multiple forms across multiple projects with little additional customization.

This approach is in contrast to a more traditional approach where a database is designed and then a web service is written to match the database schema. A major limitation of this approach is that whenever a change is made to the InfoPath form that affects the data structure (for example, the addition of a new field), a developer must make two additional changes, first to add the field to the database and then to update the web service to match the new database schema. Given the continuously evolving nature of forms in most research projects, this approach becomes untenable.

In contrast, the BEACON web service will automatically generate database tables and fields based on the InfoPath XML schema. When a change is made to the XML schema, a corresponding change automatically will be made to the SQL Server
database schema. As illustrated in the diagram below, when an InfoPath form is submitted, the web service first checks to see if a matching database table exists. If not, the table is created. Next the web service iterates through the XML nodes, checking each to see if matching columns exist in the database table. If not, the columns are created. Finally, with the XML and database schemas aligned a new record is added to the database table or an existing record is updated. There are additional nuances such as the creation of secondary tables for data collected in a repeating section. But overall this method offers a simple, efficient way to transform InfoPath-generated XML documents into SQL Server database records while minimizing custom programming. Already this method has been used on more than a dozen high-impact clinical research studies including two FDA-regulated phase II clinical trials, and data integrity has been consistently maintained.

Figure 10: Web service workflow.

6.1.4.3 Database model

The database model resulting from the use of a generalized web service is simple and, accordingly, accommodates a range of reporting, integration and analysis. As depicted below, each InfoPath form corresponds to at least one database table; each repeating section within a given form corresponds to a child table of its own. Each field within the form corresponds to a database column. As a result, when a form is submitted, a new row is added to the main database, and one or more rows are added to any secondary databases.
Note that a database trigger is also used to generate an audit table. Each time a form is updated and a corresponding change is made to the database, a new row is added to the audit table with the following information: form name, field name, date/time, operator, old value, new value. This level of detail is sufficient to satisfy most regulatory requirements including 21 CFR Part 11.

6.1.4.4 Advantages of web service/database model

The complete data entry model is illustrated in Figure 12 below (a larger view can be found in Figure 13, located at the end of this section). Some additional details are noteworthy. For one, this model works for both web-based InfoPath forms, which are expected to be used exclusively for BEACON projects, and client-side InfoPath forms. The same model also may be used with a custom .NET web form. In exceptional cases, a form may be so complex that it cannot be effectively built in InfoPath (for example, if the form includes extensive audio/video integration). In such a circumstance the form will instead be developed using C#.NET code, which will submit the data to the same web service, which in turn will update the database.
In addition to being usable by both web-based InfoPath forms and other data entry instruments, the web service/database model has another key advantage: the possibility of generating reports using SQL Server Reporting Services. As is illustrated, these reports may be deployed to the report server and viewed either there or in a SharePoint web part, or they may be published directly to a SharePoint site where SQL Server Reporting Services integrated mode is enabled. In this configuration, access to the report is controlled within SharePoint itself, which greatly simplifies the management of users and permissions. More details on generating and publishing reports may be found in section 6.3, and more details on SharePoint security may be found in section 0.

6.1.5 Supporting documentation

A final component of the BEACON data collection plan is to develop supporting documentation for InfoPath form development. Based on more than four years of experience working with InfoPath, 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.
Figure 13: Technical flow of data during data collection process (expanded view).
6.2 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.

6.2.1 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 15.2)

Two technologies will be employed to meet these requirements: an Identity and Access Management (IAM) system and Microsoft SharePoint.

6.2.2 Identity and access management (IAM)

The IAM system is a major component of a new research network infrastructure that has been developed at CCHMC and will be leveraged for the BEACON data infrastructure. This network has been designed with several needs in mind:

- **Collaboration**: an increasing number of research projects span multiple institutions; users at these institutions need an easy way to get access to shared tools such as web applications and databases.
- **Security**: the need to share resources must be balanced against organizational and regulatory standards for securing systems and data.
- **Compliance**: the ability to adhere to regulations such as 21 CFR Part 11 can no longer be delayed and requires both technical solutions (e.g., detailed monitoring/auditing capabilities) and the ability to define and enforce policies.
- **Availability**: to attract and retain collaborators, uptime of web sites, applications and databases must be maximized.
- **Scalability**: resources must be able to support a constantly growing research community.

The new network’s multi-tier architecture, with redundancies and firewall protection in each layer, provides inherent security advantages while ensuring high availability and the ability to expand capacity as user demand increases.
To effectively balance the needs of research collaboration with compliance, an additional network component has been developed: an identity and access management (IAM) system based on an Oracle framework and customized by CCHMC system administrators and developers. This system offers several compliance-oriented features such as the ability for network administrators to define and enforce policies (e.g., annual account review/renewal) and for administrators and web application owners to monitor users and permissions. At the same time, the system accommodates multi-site collaboration by enabling users, either directly or via a proxy at CCHMC, to request new accounts and new permissions online. As a result, outdated and insecure paper-based processes involving fax machines and manual processing have been replaced by secure, automated workflows.

6.2.3 SharePoint

On top of the user management abilities of the IAM system, SharePoint, which will be used as the platform for hosting the BEACON data collection forms, reports and other collaboration features (see sections 6.1 and 6.3 for details), has its own built-in user management capabilities, adding another layer of administrative control to the BEACON data infrastructure. For example, within each SharePoint site an unlimited number of groups may be created, each with assigned permissions that are inherited by their members. In addition, group and user permissions may be set at any level – from the web site as a whole to folders and even specific items within a folder. In this example, unique permissions are being set for an individual Word file:
This type of fine-grained access control makes it possible for users from multiple sites to go to a common web site and get access to only the information and data they should.

6.2.4 Satisfying must-have requirements

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

6.2.4.1 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 network, provides a means of streamlining the user account creation process while strengthening checks of user identity and compliance. As Figure 16 illustrates, via a self-service web portal one or more designated members of a given BEACON project’s staff will 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. Gone are the logistical nightmares of collecting and routing paperwork, but retained and even enhanced is 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.
6.2.4.2 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 SharePoint site. If the user is not authorized to access the site, the request will be denied and logged before it ever reaches SharePoint.

On top of this very strong initial layer is the security of SharePoint itself. By default, all SharePoint sites are password-protected, so even if a user were able to bypass the IAM gateway and make a request directly to the SharePoint server, access would only be granted if the user were a member of the site. Within SharePoint, 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.

6.2.4.3 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 SharePoint will enable non-IT staff to manage user settings. Granting a user access to a given BEACON project, for example, can take a couple different forms: in the recommended 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 SharePoint. Via a simple web form, for example, a user can be added to a site and assigned either group or direct permissions, as the following screen indicates:

Figure 17: Assigning group permissions to users in SharePoint.

6.2.4.4 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 for research studies, 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 woefully inadequate. Both the IAM system and SharePoint, however, include functionality for monitoring and tracking user accounts and related permissions. IAM, for example, serves 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 includes 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 automatically will 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 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 IAM also enables project administrators to specify expiration dates for user access.

Out-of-box, SharePoint provides basic site usage reporting (total page hits, frequent visitors per page) and enables site administrators to track users and permissions by
browsing the membership of the various groups on the site. But to achieve more granularity – for example, detailed histories of user access or listing of user permissions – an additional tool or service is required. Thanks to the existence of a large, active community of third-party developers, there are a number of SharePoint add-ons specifically designed to make access, usage and permissions monitoring an easy task that can be performed directly by BEACON project staff instead of IT staff. Examples include:

- Axceler ControlPoint
  (http://www.axceler.com/SharePointAdministration/ControlPoint.aspx)
- Quest Administrator for SharePoint
  (http://www.quest.com/site-administrator-for-sharepoint/)
- DocAve SharePoint Administrator
  (http://www.avepoint.com/sharepoint-administration-docave).

One of these products, or a suitable competitor, will be installed on the BEACON SharePoint farm. In combination with the functionality provided by IAM, these tools will provide project staff with not only on-demand access to all the information they need about their users, but also the ability to make changes to user settings without involving IT staff.
6.3 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.

6.3.1 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==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 15.3].

As a reporting platform for the various BEACON initiatives, we propose the use of Microsoft SQL Server Reporting Services (SSRS).
6.3.2 SQL Server Reporting Services

SQL Server Integration Services is a full-featured reporting toolset that integrates well with the rest of the Microsoft technology stack (including SharePoint), supports graphical and tabular reports, and allows moderate-to-expert end users to create their own reports, removing the need to go to a developer for every request. In fact, we envision that the study staff or data managers would create most new reports, with the infrastructure team only being involved during the final QA and deployment.

6.3.3 Satisfying must-have requirements

SQL Server Reporting Services natively supports 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 depends on the report configuration, but there is nothing within SSRS that would prevent the creation of such a report.

The list below contains a set of standard SSRS features, as provided by Microsoft. They are taken more or less verbatim from the Microsoft website. The intention of providing this information is to illustrate the capabilities of the tool. Even the existing BEACON projects that have used automated reporting have only scratched the surface.

- **Tabular, matrix, chart, and free-form report layouts** – Create tabular reports for column-based data, matrix reports for summarized data, chart reports for graphical data, and free-form reports for everything else. The tablix data region can be used for reports that combine table and matrix layouts. Reports can embed other reports and charts, along with lists, graphics, and controls for dynamic Web-based applications.
- **Presentation formats** – Choose a presentation format before or after opening. Formats include HTML, MHTML, PDF, XML, CSV, TIFF, Word, and Excel.
- **Aggregations** – Data can be aggregated and summarized using controls and expressions.
- **Relational, multidimensional, XML, and custom data sources** – Build reports using data from any data source type that has a Microsoft .NET Framework-managed data provider, OLE DB provider, or ODBC data source. Reports can be created using relational and multidimensional data from SQL Server and Analysis Services, Oracle, Hyperion, and other databases. An XML data processing extension can be used to retrieve data from any XML data source.
- **Ad hoc reports** – Ad hoc reporting is supported through report models and report templates using an application called Report Builder. Model-based ad hoc reports provide infinite click-through functionality so users can explore the data that is available to them.
- **Drill-through reports and interactivity** – It is possible to add interactive features by providing links to sub-reports and drill-through reports. Parameters can be used to filter data for customized views.
- **Graphical elements** – It is possible to embed or reference images and other resources, including external content.

In terms of deployment, and access control, SSRS offers the following benefits:

---

• **Server-based platform** – By centralizing report storage and management, SSRS provides secure access to reports and folders and controls how reports are processed and distributed.

• **Integration with SharePoint** – SSRS can be configured to run in SharePoint integrated mode, providing access to report server content and operations from within a SharePoint site or farm.

• **Report Manager** – It is possible to configure role-based security and manage report server content by setting properties on data sources, reports, folders, resources, and report models. It is also possible to configure report execution, report history, and to set limits on processing time, monitor and cancel pending or in-process reports, and create and manage data source connections and schedules independently of the reports that they are associated with.

• **Role-based security** – Role-based security can control access to folders, reports, and resources. Security settings follow an inheritance pattern through the folder structure. Security can be varied at any branch to redefine user access down to the item level.

• **On-demand access over Web connections** – A web browser can be used to navigate a folder hierarchy to find and work with reports and other items.

• **SharePoint Web parts** – Reporting Services supports different levels of integration. If deployed in SharePoint integrated mode, reports can be accessed from libraries on a SharePoint site. If a native mode report server is used, Web parts can be embedded to view reports and navigate the report server folder hierarchy.

• **Subscriptions for E-mail or File share Delivery** – Report delivery can be automated through standard subscriptions and by setting report presentation preferences. Users who prefer to view a report in Microsoft Excel, for example, can specify that format in a subscription. Rendered reports can be delivered to an e-mail inbox. Delivery options control whether the report is delivered as a link or attachment.

6.3.3.1 **Reporting Authoring**

SQL Server Reporting Services reports can be developed using a free Microsoft download called Report Builder. Now on version 3.0, Report Builder is a report authoring tool that features a Microsoft Office-like authoring environment and allows users to embed features such as sparklines, maps, the data bar, and indicator data visualizations. New features include the ability to create reports containing aggregates of aggregates, as well as enhanced support for expressions. Easy-to-use wizards walk users through the steps of creating a table, matrix, chart, or map.

6.3.3.2 **Report Templates**

To help standardize the data and quality reports used by the BEACON community, each project using the data infrastructure would be 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 could be customized, but the underlying logic behind each one would remain the same.

---

Projects would be free to create additional reports if they had specialized needs, and each project would also be 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. 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 propagated to the different projects.

The report in Figure 18 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**

<table>
<thead>
<tr>
<th>Hospital</th>
<th>JAN 10</th>
<th>FEB 10</th>
<th>MAR 10</th>
<th>APR 10</th>
<th>MAY 10</th>
<th>JUN 10</th>
<th>JUL 10</th>
<th>AUG 10</th>
<th>SEP 10</th>
<th>OCT 10</th>
<th>NOV 10</th>
<th>DEC 10</th>
<th>JAN 11</th>
<th>FEB 11</th>
<th>MAR 11</th>
<th>APR 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disney</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SeaWorld</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Legoland</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Six Flags</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Kings Island</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Universal</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Sea World</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Cedar Point</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
</tbody>
</table>

**Submitted Data by the First Monday of the Month**

<table>
<thead>
<tr>
<th>Hospital</th>
<th>APR 11</th>
<th>JAN 10</th>
<th>FEB 10</th>
<th>MAR 10</th>
<th>APR 10</th>
<th>MAY 10</th>
<th>JUN 10</th>
<th>JUL 10</th>
<th>AUG 10</th>
<th>SEP 10</th>
<th>OCT 10</th>
<th>NOV 10</th>
<th>DEC 10</th>
<th>JAN 11</th>
<th>FEB 11</th>
<th>MAR 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disney</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>SeaWorld</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Legoland</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Six Flags</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Kings Island</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Universal</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Sea World</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Cedar Point</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
</tbody>
</table>

As of 6/2/2011

Figure 18: 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 19 provides an example of a run chart with a data table. Several performance 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.
Figure 19: Run chart with data table. This report shows a chart with various measures, along with the data that comprise each line.

Figure 20 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. Annotations are added using Report Builder, and would be added by the project's data manager or data analyst, if a special cause were determined.
Figure 20: 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 21. 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 21: 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.
SSRS also allows users to embed multiple charts into a single page, as shown in Figure 22. Displaying all 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.

**Figure 22: 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.

As mentioned above, SSRS provides the ability to export data in several common formats. A list of the standard export formats is shown in Figure 23. Two of the formats are data-specific (CSV, XML), meaning only the data behind the report is downloaded, not the graphics of the report itself. The remaining formats, PDF, MHTML, Excel, TIFF and Word provide more of a WYSIWYG (What You See Is What You Get) feel. Please note that by default, the reports might not “pretty print.” This means that without configuration, a report may end up split across multiple pages. If pretty printing is desired, through Report Builder it is possible to ensure proper formatting during the report design.
Figure 23: SSRS reports and data can be exported in a number of formats. Charts can also be configured to display only certain time periods (functionality not shown).

The proposed data infrastructure provides several ways of providing data extracts to analysts and investigators. One of those methods is through an SSRS report. Figure 24 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 SharePoint to ensure that only authorized users have access.

<table>
<thead>
<tr>
<th>Hospital ID</th>
<th>Patient Tracking Number</th>
<th>Patient ID</th>
<th>Current Weight</th>
<th>Admission Date</th>
<th>Discharge</th>
<th>Age (In Days)</th>
<th>Inpatient Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>XYZ</td>
<td>XYZ - 67</td>
<td>67</td>
<td>59.5</td>
<td>12/7/2009 12:00:00 AM</td>
<td>12/10/2009 12:00:00 AM</td>
<td>56</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 68</td>
<td>68</td>
<td>61</td>
<td>12/2/2009 12:00:00 AM</td>
<td>12/22/2009 12:00:00 AM</td>
<td>108</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 69</td>
<td>69</td>
<td>4.8</td>
<td>12/5/2009 12:00:00 AM</td>
<td>12/10/2009 12:00:00 AM</td>
<td>21</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 70</td>
<td>70</td>
<td>13.6</td>
<td>11/29/2009 12:00:00 AM</td>
<td>12/2/2009 12:00:00 AM</td>
<td>120</td>
<td>Critical Care</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 71</td>
<td>71</td>
<td>5.4</td>
<td>12/18/2009 12:00:00 AM</td>
<td>12/20/2009 12:00:00 AM</td>
<td>16</td>
<td>Medical</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 72</td>
<td>72</td>
<td>4</td>
<td>12/26/2009 12:00:00 AM</td>
<td>12/8/2009 12:00:00 AM</td>
<td>66</td>
<td>Medical</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 73</td>
<td>73</td>
<td>60</td>
<td>11/26/2009 12:00:00 AM</td>
<td>12/6/2009 12:00:00 AM</td>
<td>168</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 74</td>
<td>74</td>
<td>11.3</td>
<td>12/21/2009 12:00:00 AM</td>
<td>12/25/2009 12:00:00 AM</td>
<td>54</td>
<td>Critical Care</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 75</td>
<td>75</td>
<td>26.2</td>
<td>12/26/2009 12:00:00 AM</td>
<td>12/28/2009 12:00:00 AM</td>
<td>56</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 76</td>
<td>76</td>
<td>22.2</td>
<td>12/2/2009 12:00:00 AM</td>
<td>12/3/2009 12:00:00 AM</td>
<td>56</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 77</td>
<td>77</td>
<td>22.9</td>
<td>12/5/2009 12:00:00 AM</td>
<td>12/1/2009 12:00:00 AM</td>
<td>72</td>
<td>Critical Care</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 78</td>
<td>78</td>
<td>17.2</td>
<td>11/28/2009 12:00:00 AM</td>
<td>12/1/2009 12:00:00 AM</td>
<td>72</td>
<td>Critical Care</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 79</td>
<td>79</td>
<td>88.8</td>
<td>12/4/2009 12:00:00 AM</td>
<td>12/19/2009 12:00:00 AM</td>
<td>14</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 80</td>
<td>80</td>
<td>26</td>
<td>11/23/2009 12:00:00 AM</td>
<td>12/2/2009 12:00:00 AM</td>
<td>56</td>
<td>Critical Care</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 81</td>
<td>81</td>
<td>16.8</td>
<td>12/6/2009 12:00:00 AM</td>
<td>12/16/2009 12:00:00 AM</td>
<td>56</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 82</td>
<td>82</td>
<td>58</td>
<td>12/1/2009 12:00:00 AM</td>
<td>12/7/2009 12:00:00 AM</td>
<td>56</td>
<td>Specialty Practice</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 83</td>
<td>83</td>
<td>50.9</td>
<td>12/18/2009 12:00:00 AM</td>
<td>13/5/2009 12:00:00 AM</td>
<td>216</td>
<td>Acute Admitte</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 84</td>
<td>84</td>
<td>3</td>
<td>11/10/2009 12:00:00 AM</td>
<td>11/11/2009 12:00:00 AM</td>
<td>57</td>
<td>NCU</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 85</td>
<td>85</td>
<td>6.6</td>
<td>1/1/2009 12:00:00 AM</td>
<td>12/9/2009 12:00:00 AM</td>
<td>4</td>
<td>Critical Care</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 86</td>
<td>86</td>
<td>66.8</td>
<td>12/12/2009 12:00:00 AM</td>
<td>12/15/2009 12:00:00 AM</td>
<td>30</td>
<td>Critical Care</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 87</td>
<td>87</td>
<td>5.2</td>
<td>1/26/2010 12:00:00 AM</td>
<td>1/20/2010 12:00:00 AM</td>
<td>8</td>
<td>Social Practice</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 88</td>
<td>88</td>
<td>62</td>
<td>1/1/2010 12:00:00 AM</td>
<td>1/30/2010 12:00:00 AM</td>
<td>30</td>
<td>Social Practice</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 89</td>
<td>89</td>
<td>45</td>
<td>11/1/2010 12:00:00 AM</td>
<td>1/6/2010 12:00:00 AM</td>
<td>24</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 90</td>
<td>90</td>
<td>78</td>
<td>1/1/2010 12:00:00 AM</td>
<td>1/15/2010 12:00:00 AM</td>
<td>216</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 91</td>
<td>91</td>
<td>10.3</td>
<td>11/1/2010 12:00:00 AM</td>
<td>1/15/2010 12:00:00 AM</td>
<td>216</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 92</td>
<td>92</td>
<td>12.8</td>
<td>1/1/2010 12:00:00 AM</td>
<td>1/15/2010 12:00:00 AM</td>
<td>216</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 93</td>
<td>93</td>
<td>8.7</td>
<td>1/1/2010 12:00:00 AM</td>
<td>1/15/2010 12:00:00 AM</td>
<td>216</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 94</td>
<td>94</td>
<td>0.77</td>
<td>1/1/2010 12:00:00 AM</td>
<td>1/15/2010 12:00:00 AM</td>
<td>216</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 95</td>
<td>95</td>
<td>13.1</td>
<td>1/1/2010 12:00:00 AM</td>
<td>1/15/2010 12:00:00 AM</td>
<td>216</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 96</td>
<td>96</td>
<td>15.1</td>
<td>1/1/2010 12:00:00 AM</td>
<td>1/15/2010 12:00:00 AM</td>
<td>216</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 97</td>
<td>97</td>
<td>20.7</td>
<td>1/1/2010 12:00:00 AM</td>
<td>1/15/2010 12:00:00 AM</td>
<td>216</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 98</td>
<td>98</td>
<td>30.7</td>
<td>1/1/2010 12:00:00 AM</td>
<td>1/15/2010 12:00:00 AM</td>
<td>216</td>
<td>Med - Surg</td>
</tr>
<tr>
<td>XYZ</td>
<td>XYZ - 99</td>
<td>99</td>
<td>40.7</td>
<td>1/1/2010 12:00:00 AM</td>
<td>1/15/2010 12:00:00 AM</td>
<td>216</td>
<td>Med - Surg</td>
</tr>
</tbody>
</table>

Figure 24: Tabular reports can be created as a mechanism to allow for on-demand data extracts to authorized users.
6.3.3.3 Access Control

SQL Server Reporting Services can be configured to run in Native Mode, where a stand-alone application server is installed and that server provides all the viewing, management, processing and delivery of reports and report models. The alternative is in SharePoint integrated mode, where the report server runs within a SharePoint server farm. The SharePoint site provides the front-end access to the report server content and operations. The report server provides all the report processing and rendering\(^4\). The benefit of using SharePoint integrated mode is that the database administrators and report developers do not need to configure a separate security model for the user roles and access rights to each report. They can leverage the controls provided by SharePoint, easing the administrative burden on the project staff.

The original SPS reports were delivered by an SSRS reporting server deployed in native mode. BEACON reports will be deployed using SharePoint integrated mode.

6.3.3.4 Report Deployment Process

The deployment of SSRS reports is fairly straightforward. A simplified version of the process is described below. To deploy reports for a multi-center study when SSRS is in Native mode, one must:

1. Create a SQL Server service account with permission to access the data.
2. Create separate Reports folders for each site.
3. Upload the report to SSRS.
4. Assign user permission to their respected folders/Reports.
5. Provide the report URL to the end user.

When SSRS is in SharePoint integrated mode, the process is much more straightforward and can leverage existing user accounts and permissions.

1. Create a reports library on the project’s SharePoint site
2. Upload the report
3. Assign user permission to the respective reports using SP Audience Targeting
4. Create a web part page in SharePoint to display all available reports and a report viewer web part to view the actual reports.

6.4 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.

6.4.1 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 15.3].

These requirements can be met using SQL Server and SQL Server Integration Services (SSIS).

6.4.2 SQL Server & SQL Server Integration Services

SQL Server Integration Services comprises 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. It can be used to generate extracts, trigger process execution upon file receipt and produce error logs.
SQL Server is an enterprise-class database platform that supports audit trails, user roles and security.

6.4.3 Satisfying must-have requirements

All of the must-have requirements can be met using SQL Server and SSIS. Each project is likely to have different cleaning and validation rules, as well as different data input and output formats, so meeting the 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.

6.4.3.1 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 why this is so. 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.

As part of the planning phase of the data infrastructure project, a small study was undertaken to test the feasibility of linking outside data sources with OPQC data. This project is currently underway, but due to issues with the IRB and the time needed to identify a specific research question that would allow the secondary data to be released, it was not possible to finish the project before the completion of this document.

6.4.3.2 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 SQL procedures and ETL processes created in SSIS.
   b. Once the data is linked, it will be provided back to the project staff as a set of datamarts housed in a SQL Server 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.
6.5 BEACON-specific Data Warehouse

Finding the right subset of data, particularly when linking in massive secondary sources, can be a challenge, especially in the early stages of hypothesis generation. Individual requests for basic information, such as the number of patients that meet certain inclusion/exclusion criteria, may not take an inordinate amount of time for the project staff to fulfill, but given enough requests, it will become impossible for the staff to complete any “real” work. With appropriate data obfuscation, it is possible to allow these rudimentary requests to be completed directly by the end user, freeing staff to handle more advanced requests.

The following is a proposal to construct a BEACON-specific de-identified patient-level data warehouse that links data collected directly by the BEACON projects with data obtained from other secondary sources. Based on the open-source i2b2 framework\(^5\) this warehouse is designed around cohort identification, allowing users to query large patient populations in order to identify small subsets based on certain inclusion and exclusion criteria.

It is noted that a BEACON-specific data warehouse is currently outside the proposed scope and budget of the BEACON informatics infrastructure. The main use of secondary data for BEACON projects is, and is likely to remain, simply for reporting and analysis purposes, making a warehouse unnecessary. In future phases of the project however, there may be a desire to create a warehouse that allows end users to perform de-identified cohort identification, presenting a clear business need for such a tool.

Users would interact with the warehouse in two major ways, both originating with the CCHMC-developed web-based i2b2 workbench. The first way, which is the default method of access, is to generate cohort queries through the web-based workbench. The results are returned in the form of obfuscated patient counts. The second method of access is reserved for “elevated” requests. After running a query and receiving the number of patients in a cohort, a user can fill out a form and request de-identified access to the longitudinal records of the patients in the specified cohort. These records can be returned to the user as an extract, or by viewing them through the workbench in “chart review” mode. All requests for identified data will be reviewed and approved following the data use policies established by the BEACON Council.

6.5.1 Background - i2b2

i2b2 (Informatics for Integrating Biology and the Bedside) is an NIH-funded National Center for Biomedical Computing based at Partners HealthCare System. The i2b2 Center is developing a scalable informatics framework that will bridge clinical research data and the vast data banks arising from basic science research in order to better understand the genetic bases of complex diseases. i2b2 has been released as a collection of open source software tools that facilitate the collection and aggregation of patient data, both clinical and research, into a translational research repository.

The basic function of the software is to allow the user to identify cohorts of patients from a large repository of observations. The core code manages projects, files, query

ontology, data and workflows. It is designed to allow developers to add functional modules, called cells, to the core code. Cells have been developed to extend the core to a multitude of activities like natural language processing, annotation of genomic and image data, data analysis, concept mapping, and linkage to other research systems.

The functional design of i2b2 is based around a central ontology of query terms consisting of standard and local terminology to describe clinical and research observations. The i2b2 ontology is fully extensible to local terminology sets and contains terms for diagnoses, procedures, demographics, lab and microbiology tests, visits, service providers, genomics, proteomics and imagery. Users select terms from the ontology in logical combinations (like a Venn diagram), and i2b2 generates a query against its underlying database to deliver cohorts of patients meeting the selected criteria.

i2b2 employs a data model based on the data-warehousing-standard star schema and runs on common database engines like Oracle and SQL Server. All of the patient observations are contained in a central database table. Each observation is linked to a concept which links back to the central ontology.

The data model embedded in i2b2 uses a variation of the entity-attribute-value model, where a number of attributes, each with a corresponding value, are stored about an entity (patient). Attributes are facts, measures or observations about a patient, such as demographics, diagnoses, or laboratory results. It is also possible to define de novo attributes or create variables that are derivations or combinations of existing ones. Observations are tagged with one or more concepts, which can come from standard terminologies like SNOMED or LOINC or from custom definitions specific to a project. These codes are used when querying the warehouse. Collectively these concept codes are the items that can be used in a query. For example, when a user submits a query asking for all patients with a diagnosis of ‘Asthma’, the i2b2 query engine will look for anyone with an observation that has the code associated with ‘Asthma.’ It is also possible to assign synonym codes, so that a search for ‘Asthma’ would return all patients with a corresponding ICD or SNOMED diagnosis.

Communication in i2b2 occurs through a middleware of web services. These services pass information between the different components, from the database layer up to the front-end application. Currently, users access i2b2 through one of two different approaches. One is by using a derivative of the open source Eclipse workbench that runs as a workstation-based client, and the other is through a browser-based Ajax (JavaScript and XML) client, like the CCHMC-developed i2b2 toolkit (described below).

Data can be loaded into i2b2 from disparate sources including EMR, research data systems, clinical trials systems, laboratory systems, genomics and proteomics stores, and image banks, creating a rich mineable repository of patient observations. Processes are written for each unique data source to extract, transform and load (ETL) data from source systems to the i2b2 star schema.
A major strength of networks that build upon the i2b2 warehousing platform is the existence of a very active and diverse academic users group, consisting of sites throughout the US, as well as internationally. The users build and share software modules, which interface with i2b2 and meet diverse needs of translational researchers. As a result, this open source platform is surrounded by a wide and growing variety of interfaced services (cells), which are rapidly enhancing the ability of i2b2 to support many types of translational research.

6.5.2 Data Processing

The data to be loaded into the BEACON data warehouse may contain identifiers. These identifiers will be used to create a single, integrated set of data for each patient. They will also be used to help remove any spurious or duplicate information. All procedures to be used for the extraction, transformation and loading (ETL) of data would be performed on a separate server. The production instance of the de-identified warehouse would be deployed on a database that is physically and logically separate from the ETL database.

To allow for the historical linking of records across multiple data sources, and for the re-identification of patients at a later date, patient and physician numbers would be replaced with a BEACON identifier. This will allow a connection across records without revealing any identifying information. The numbers would be generated randomly and would not be based on any underlying patient or provider information. There is no algorithm that would allow a user to generate any of a patient’s identifiers based on that patient’s internal BEACON identifier.

There would be a mapping table on the production database that is used to store encrypted patient identifiers and the corresponding BEACON identifier. This table would not accessible by the end user and the decryption routines will only be accessible to the BEACON warehouse staff. Keeping the encrypted identifiers in the production database will significantly reduce the amount of processing needed to refresh the database, and greatly ease the work of the warehouse staff when running mediated queries.

Access to the production database would be controlled by username and password and be limited to the warehouse staff. End users would access the database through the web-based workbench using their BEACON username and password. The ETL databases are controlled by a separate set of usernames and passwords that are only available to members of the BEACON warehouse team.

6.5.3 Access to the BEACON i2b2 workbench

Depending on the level of access, a user would be able to either generate obfuscated patient counts or view/download the de-identified records for an entire patient cohort. The proposed permission levels are as follows:

- Obfuscated Access – It is proposed that this represent the default level of access and be open to any user participating in a BEACON project who has signed the appropriate data use/access agreements. With obfuscated access, a user could run queries, but the results are limited to obfuscated (approximate) count data, with “<5” returned if there are fewer than 5 patients in a given cohort. In addition, if a user ran the same query more than 7 times within 7 days, their account would be locked and they would be unable to log in to the warehouse without contacting the warehouse staff to have the suspension lifted. The primary purpose for this
level of access would be to generate hypotheses and to conduct reviews preparatory to research.

- Elevated Access – After running a query, a user may request “elevated” access to the patient cohort. Within the i2b2 workbench, a user could right-click on a previous query and “request elevated access.” That would bring up a form that requests the following information:
  o Description of use/reason for access
  o User’s primary BEACON project
  o Responsible party if member is not an investigator on the project
  o Type of access (chart review or extract)
  o Required fields (if extract)
  o Format (Excel or csv)
  o Additional query metadata will automatically be pulled from the system (number of patients, query criteria, query id)
  o Before submitting the form, the user will need to agree to a second data use agreement (to be drafted by BEACON, though a sample is listed in the box below)

Upon submission, if the requestor were not a project investigator, the request would be routed to the responsible member for approval. If they approve, or if the original requestor is a project investigator, the request would then be routed to the <BEACON Data Use Committee> for final approval.

If the request were approved, either a data extract or a cohort-specific i2b2 data mart (for chart review) would be created. Extracts would be placed on the user’s BEACON project web site, where they will be able to securely download the file. If the user requests a cohort-specific i2b2 data mart, a separate i2b2 mart would be created containing only those patients in the specified cohort. The requestor would be able to access the data mart through the i2b2 workbench and will have the ability to view the longitudinal records of the patients in the cohorts. They would also be able to generate and download reports run against the cohort. These data marts would be active for 6 months, after which they will be decommissioned, unless the requestor asks for an additional 6-month extension. Users would be allowed one extension, resulting in a maximum chart review period of one year.

By default, users would not be able to request a chart review for patient cohorts larger than XXX patients or extracts on more than YYY patients (limits to be defined later). If a request were outside these limits, it would need to undergo additional review by the <BEACON Data Use Committee>. All requests must contain 10 patients or more, however.

- Identified Access – At this level of access, a researcher could request identified patient data. They could file a request using the Elevated Access workflow as described above. Data are provided as an extract or a cohort-specific data mart. The only difference is that these requests must be approved using the IRB processes established by the BEACON Council.
SAMPLE DATA USE AGREEMENT

I understand that data contained in the research data warehouse is extracted from a
variety of systems and will reflect the same data, errors or omissions that may be
contained or absent from the source data.

I understand that data is audited upon extract and during processing to ensure the
highest level of accuracy possible in the data reports.

I understand my responsibility to report any use or disclosure not provided for by this
data use agreement.

I understand that all parties listed above with whom the data set will be shared will abide
by the same conditions and restrictions.

I understand that I am responsible for compliance with all BEACON privacy and security
policies and procedures including protecting the security and the confidentiality of the
information to which I have access using appropriate safeguards and due care.

I verify that I have the right to receive and access the information stated above and have
met all federal and state legal requirements for such access.

I assert that the information furnished is true and is the minimum necessary to meet the
needs of the request.

Requestor: <DATE>,<USER>
By submitting this request, I accept this agreement unconditionally. I Agree.

6.5.4 Workbench Features

The CCHMC i2b2 team has developed a web-based workbench that allows users to
either query the de-identified warehouse or perform a longitudinal chart review on a
specific cohort, depending on the level of access that has been granted to the user.
This code is open-source and available from the CCHMC i2b2 Subversion repository
(https://bmi.cchmc.org/svn/i2b2/i2b2/public/).

Users who have Obfuscated Access can use the workbench (see Figure 25 below) to
execute cohort queries. The workbench includes an autosuggest tool (1) that can be
used to find search terms within the ontology (2). These terms can be dragged and
dropped into a Venn diagram-like interface (3). Upon execution (4), the query will return
an obfuscated count of the number of patients that meet the specified criteria (5). If a
user wishes to gain elevated access to a query, they can request it by right clicking on a
previous query (6).
Figure 25: i2b2 workbench developed by CCHMC. Screen shows view of an Obfuscated Access user.

With Elevated/Identified Access, a user has access to the longitudinal records of a specific patient cohort, as shown in Figure 26. They can still use the general query tool to generate patient counts (though it will only return subsets of the approved cohort). Users can either select to view individual patients or general demographic information (1). In contrast to the count returned under the Obfuscated Access role, queries run with Elevated privileges return the exact number of patients (2). When accessing specific patients, a user can access a list of visits (3) and view the data collected during that encounter (4).
Finally, it is possible under the Elevated/Identified Access roles to view the same data through the workbench’s chart review mode (Figure 27). In this mode, a user can select a patient (1), and then view the different types of data available for that patient (2). In the “Encounter” mode, they can select a given encounter (3), (4), and view a list of data that was recorded at that particular visit.
6.5.5 Mediated Queries

Users also would have the ability to work directly with the BEACON warehouse team to meet their data needs. The BEACON team would typically handle complicated queries, queries involving specialized data sources, or those involving identified data. The BEACON team would record the same information required under the request for Elevated Access and will follow the same steps for the approval of mediated data requests.

<table>
<thead>
<tr>
<th>Start Date</th>
<th>End Date</th>
<th>Diagnosis</th>
<th>Type</th>
<th>Status</th>
<th>Class</th>
<th>Primary?</th>
<th>Chronic?</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013-09-16</td>
<td>2013-09-16</td>
<td>Nuevas/21.8(82)</td>
<td>NN</td>
<td>Deleted</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2013-09-16</td>
<td>2013-09-16</td>
<td>Acetaminophen 375 mg</td>
<td>Inhaled</td>
<td>Deleted</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2013-09-16</td>
<td>2013-09-16</td>
<td>Acetaminophen 375 mg</td>
<td>Inhaled</td>
<td>Deleted</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2013-09-16</td>
<td>2013-09-16</td>
<td>Acetaminophen 375 mg</td>
<td>Inhaled</td>
<td>Deleted</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th>Start</th>
<th>End</th>
<th>Location</th>
<th>In/Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2010-09-16</td>
<td>2010-09-16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 27: View of the chart review mode of the i2b2 workbench.
7 Records Retention

To safeguard BEACON systems and data, all components required to reconstruct or reestablish any given system or application will be stored within two separate and secure sites at all times. The first site is the Team Foundation Server (TFS), a mature and fully supported Microsoft product used for source control throughout the software development lifecycle and which serves 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 if required, and includes application modules, files, documentation, and any other aspects needed to reproduce or modify the product. This TFS 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 TFS system, includes all relevant data. There is currently no plan for deletion of these systems or data, which will be retained indefinitely.

This structure of file and data retention satisfies not only the immediate production needs to assure ongoing access and reliance, but also satisfies 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 TFS system does allow control policies that govern who and how access and changes are allowed, records such actions, and optionally alerts named users when events occur.

8 System Benefits & Drawbacks

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

- The system is capable of meeting all of the “must-have” requirements, as specified by the BEACON Technical Committee.
- It provides 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 are still needed to perform tasks like quality assurance, template development and form deployment, but there is a role for end users.
- The system is scalable, high-performance and high-availability. It has the ability to handle statewide projects, with hundreds or thousands of users.
- It meets or exceeds the requirements for security, auditing, and protection of patient privacy.
- The proposed technologies are easily integrated. InfoPath forms can be hosted and accessed remotely via SharePoint. SSRS reports can be accessed seamlessly through SharePoint. SharePoint lists can be manipulated in Microsoft Office. With SharePoint as the platform to host reports and forms, its collaboration tools come with no added cost.
- While there is a focus on creating common templates and reusable components, it is possible to create customized workflows and tailor the look-and-feel so that is it unique to every project.
• The system requires only a handful of staff to administer, and this cost can be amortized 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. The system will be unable to “automatically” accept new data without configuration. While this is largely true of any system that might be used for this purpose, it is important to note.
• Form development and deployment is not 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 require a license. If the system is developed and hosted at CCHMC, these costs are subsidized and largely absent, but other outside vendors may not have the same pricing structure.

9 Evaluation of Alternative Approaches

The software development group in BMI develops EDC systems and custom applications to support the research and QI endeavors of investigators at CCHMC. As such, BMI must constantly evaluate its existing offerings to make sure that the innovative needs of investigators are being met. If there is a gap in its offerings, BMI strives to find new technologies that will satisfy the new requirements.

As part of this process, in 2007, BMI performed an evaluation of several leading commercial technologies for enterprise-scale EDC. These technologies included Cardiff LiquidOffice (Teleforms/eForms), Adobe Acrobat/LiveCycle, and Microsoft InfoPath/SharePoint. BMI found the cost of the different platforms (overall system, per user) to be roughly equivalent. However, LiquidOffice and Acrobat/LiveCycle only provided capabilities for data collection. The combination of InfoPath and SharePoint offered data collection, as well as a rich collaboration space that came essentially “for free.” The integration with SSRS, a powerful reporting tool was especially appealing. The ability to manipulate SharePoint lists with Microsoft Office was also an attractive feature. Finally, the Microsoft tools are used elsewhere in the region, making it relatively easy to find qualified staff to administer and maintain the system. The other systems are not widely used locally, making support a challenge.

Since 2007, several other technologies have been released in this space, particularly Adobe AIR and REDCap (an open-access EDC tool released by Vanderbilt). Adobe AIR suffers the same drawbacks we found with LiquidOffice and LiveCycle – support is a challenge. REDCap, which is offered at CCHMC, is an excellent tool for small-to-medium-scale studies, but provides limited ability to create workflows, no ability to customize look-and-feel, and minimal reporting capabilities. When using REDCap, the cost of data collection is essentially free, but more time must be spent on data
integration and report development, as this must be done outside the system. This additional burden makes the data integration and reporting more complicated and neutralizes the data collection cost savings. As a result, BMI feels that InfoPath and SharePoint remains the best solution for large-scale data collection projects, particularly when one considers the updated features provided in the newest versions of SharePoint and SSRS.

10 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 system was designed to leverage the state of Ohio’s existing investment in OPQC and SPS. The system needed to be enterprise-scale and be able to handle undefined future extensions. Users needed the ability to customize certain elements while also having the ability to create common reusable templates. Finally, the system needed to be designed so that it could be built out in phases as funding and projects allowed.

11 Project Risks & Mitigations

Taking on a project as large as the BEACON data infrastructure requires careful planning for, and mitigation of, risks. If chosen as the vendor to develop this infrastructure, as an overall risk reduction strategy, BMI will follow defined methods and processes for software development -- specifically, a modified version of the agile development methodology. Experienced project managers and business analysts will be responsible for ensuring proper implementation of this methodology, in which system development occurs in short, iterative cycles. Each cycle includes testing and other quality assurance checks. At the end of each cycle, the development staff and project staff/stakeholders 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.

Over the past three years BMI has successfully used this approach to develop more than three dozen electronic data capture systems including two systems that support phase II clinical trials and two that support existing BEACON initiatives (OPQC, SPS). By leveraging this experience -- particularly in using the technologies and methods described in this document to support quality improvement research -- the BEACON Council will minimize the risk that, in selecting an alternative vendor, its resources will not be used as efficiently and effectively as they will by BMI.
Appendix 1 – Data Infrastructure Development Project: Phase 1 Approach

1. Inventory existing and proposed quality projects; ID key stakeholders
2. Interview 10-20 key stakeholders
3. Prepare for/execute requirements workshop(s) to finalize system requirements
4. Compile user and technical req’s from previous, proposed quality projects
5. Finalize functional data, and records management requirements
6. Develop architecture options
7. Evaluate options
8. Develop high-level architecture
9. Conduct additional analysis and develop detailed design
10. Develop form, data specifications for 2 proposed projects
11. Develop forms for 2 proposed projects
12. Inventory/profile programs and data exchange points, identify system interfaces
13. Test data linkage programs and exchange points
14. Develop detailed plan for Phase 1
15. Develop high-level plans for Phase 2
16. Complete data governance, publication, data collection, and dissemination agreements
17. Provide human-to-human support/help desk functionality for active projects using CCH-MC systems for data collection, analysis, and reporting
18. Manage project:
   - Identify, track, mitigate, and report/escalate (as needed) project risks and issues
   - Track/forecast project financials, allocate resources
   - Develop protocol/obtain and manage IRB approval
   - Create and update timetelines, delegate tasks
   - Monitor and report on project status

Week Ending
13 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?
# Appendix 3 – Interview Schedule

<table>
<thead>
<tr>
<th>Project</th>
<th>Interviewee - Key Stakeholder</th>
<th>Key Stakeholder Organization</th>
<th>Invitation Sent</th>
<th>Interview Date</th>
<th>Interviewee - Operational</th>
<th>Interviewer(s)</th>
<th>Invitation Sent</th>
<th>Scheduled (completed) date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. CHIPRA Medicaid Encounter Data</td>
<td>Gerry Fairbrother</td>
<td>CHMC</td>
<td>1/21/2011</td>
<td>2/16/10, 2pm-3pm</td>
<td></td>
<td></td>
<td></td>
<td>1/27/11</td>
</tr>
<tr>
<td>2. Dept. of Mental Health/ Psych</td>
<td>Beth Ferguson</td>
<td></td>
<td>1/27/11, 2/2/2011</td>
<td>2/24/11, 1-2pm</td>
<td></td>
<td></td>
<td></td>
<td>1/27/11</td>
</tr>
<tr>
<td>4. OPQC - Neo</td>
<td>Ed</td>
<td>CHMC</td>
<td>Completed 2/14/11</td>
<td></td>
<td>Matt Short (PM)</td>
<td>Tim</td>
<td>1/21/11</td>
<td>2/3/11, 3pm-6pm</td>
</tr>
<tr>
<td>5. OPQC - OB</td>
<td>Latasha Powell (QIC)</td>
<td>CHMC</td>
<td>Completed 2/11</td>
<td></td>
<td>Ralph Brueggemann (Analyst)</td>
<td>Brooke</td>
<td>1/27/11</td>
<td>2/3/11, 9am-10am</td>
</tr>
<tr>
<td>6. Medtapp Development (Autism)</td>
<td>Carolie Lannon</td>
<td>CHMC</td>
<td>Completed 2/1</td>
<td>Kevin Stanford</td>
<td>James, Tm</td>
<td>1/27/11</td>
<td>Completed 2/3/11, 4-5pm</td>
<td></td>
</tr>
<tr>
<td>7. Medtapp HKD (Obesity)</td>
<td>Carolie Lannon</td>
<td>CHMC</td>
<td>Completed 3/2/11</td>
<td></td>
<td>Sam</td>
<td>James, Tm</td>
<td>1/27/11</td>
<td>Completed 2/17/11</td>
</tr>
<tr>
<td>8. OCHA HKD OP</td>
<td>Melissa Arnold</td>
<td>AAP</td>
<td>Completed 3/2/11</td>
<td></td>
<td>Stacy Kramer</td>
<td>James, Tm</td>
<td>1/27/11</td>
<td>Completed 1/31/11</td>
</tr>
<tr>
<td></td>
<td>Amy Sternstein</td>
<td>Nationwide</td>
<td>Completed 3/2/11</td>
<td></td>
<td>Bob Murray</td>
<td></td>
<td></td>
<td>Completed 3/23/11, 10am</td>
</tr>
<tr>
<td></td>
<td>Bob Murray</td>
<td>Nationwide</td>
<td>Completed 3/2/11</td>
<td></td>
<td>Samartha Anzelic</td>
<td></td>
<td></td>
<td>Completed 3/23/11, 10am</td>
</tr>
<tr>
<td>9. Anthina</td>
<td>Melissa Arnold</td>
<td>AAP</td>
<td>Completed 3/2/11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10am</td>
</tr>
</tbody>
</table>
## 15 Appendix 4 – Must-Have System Requirements

### 15.1 Data Collection

<table>
<thead>
<tr>
<th>Requirement ID</th>
<th>Category</th>
<th>Subcategory</th>
<th>Requirement Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC1</td>
<td>Data collection</td>
<td>Functionality</td>
<td>Permits Internet/ web-form entry into network DB</td>
</tr>
<tr>
<td>DC12</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System allows group/ bulk submission of multiple forms (up to 50)</td>
</tr>
<tr>
<td>DC19</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System to allow upload of a standard format file (export from EMR or other enterprise application).</td>
</tr>
<tr>
<td>DC20</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System allows for data to be checked for accuracy, format, and completeness as it is entered; and prompts user to enter the correct data</td>
</tr>
<tr>
<td>DC21</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System provides feedback on data that was just uploaded</td>
</tr>
<tr>
<td>DC24</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System validates data upon upload and prompts user to update incorrect fields</td>
</tr>
<tr>
<td>DC26</td>
<td>Data collection</td>
<td>Functionality</td>
<td>Data extracted from EMR/ enterprise system/ secondary data source is translated to correct format and checked (values within range, unit conversion, completion, data type)</td>
</tr>
<tr>
<td>DC41</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System tracks and documents which designated user inputs data for each entry</td>
</tr>
<tr>
<td>DC56</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System can operate on multiple platform</td>
</tr>
<tr>
<td>DC16</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System to automatically save data upon input at 15 second intervals</td>
</tr>
<tr>
<td>DC30</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System to disallow local modification of network-common forms/ fields</td>
</tr>
<tr>
<td>DC34</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System to create unique identifiers for designated users</td>
</tr>
<tr>
<td>DC45</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System has ability to monitor status of all data fields against validation rules</td>
</tr>
<tr>
<td>DC49</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System allows for data to be checked for accuracy, format, and completeness after it is uploaded to the network database</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------</td>
<td>---------------</td>
<td>------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>DC50</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System to preserve data integrity during upload</td>
</tr>
<tr>
<td>DC51</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System can verify new data against stored data for accuracy</td>
</tr>
<tr>
<td>DC54</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System automatically saves and backs-up all data stored to the network database</td>
</tr>
<tr>
<td>DC58</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System supports automated collection, storage, retrieval, and transfer</td>
</tr>
<tr>
<td>DC67</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System to allow local or network forms to access/make use of standard or common measures definitions from reference materials</td>
</tr>
<tr>
<td>DC68</td>
<td>Data collection</td>
<td>Functionality</td>
<td>Logs and creates reports on errors (validation, upload, etc.)</td>
</tr>
<tr>
<td>DC69</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System to capture/link patient-level data (unique identifier for each patient that links multiple entries to a single patient)</td>
</tr>
<tr>
<td>DC14</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System allow save of partially completed forms (allows return to complete at a later date)</td>
</tr>
<tr>
<td>DC17</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System to allow creation of standard format file for data extraction from secondary data source or EHR</td>
</tr>
<tr>
<td>DC18</td>
<td>Data collection</td>
<td>Functionality</td>
<td>Ability to add upload modules to import secondary data to registry/database</td>
</tr>
<tr>
<td>DC32</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System to limit changes to network forms to authorized users only</td>
</tr>
<tr>
<td>DC47</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System allows designated resources to revise fields and entries after upload</td>
</tr>
<tr>
<td>DC64</td>
<td>Data collection</td>
<td>Functionality</td>
<td>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)</td>
</tr>
<tr>
<td>DC65</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System to allow users to submit questions/ respond to questions from network manager (e.g., data definition, submission date, etc.)</td>
</tr>
<tr>
<td>------</td>
<td>----------------</td>
<td>---------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>DC71</td>
<td>Data collection</td>
<td>Functionality</td>
<td>System to allow standard forms to be pulled and placed for use by any project</td>
</tr>
<tr>
<td>DC9</td>
<td>Data collection</td>
<td>Performance</td>
<td>System response lag (e.g., refresh on webform) &lt; 5 seconds</td>
</tr>
<tr>
<td>DC10</td>
<td>Data collection</td>
<td>Performance</td>
<td>Form submission &lt; 15 seconds</td>
</tr>
<tr>
<td>DC11</td>
<td>Data collection</td>
<td>Performance</td>
<td>System to indicate delay/lag in time frame to execute</td>
</tr>
<tr>
<td>DC57</td>
<td>Data collection</td>
<td>Performance</td>
<td>System can support multiple networks (up to 100)</td>
</tr>
<tr>
<td>DC66</td>
<td>Data collection</td>
<td>Performance</td>
<td>System availability/ up-time to be &gt; 95%, with outages resolved within 4 hours of report (std business hours)</td>
</tr>
<tr>
<td>DC33</td>
<td>Data collection</td>
<td>Security/compliance</td>
<td>System to limit access to authorized users or groups of users for Functionality and form modification</td>
</tr>
<tr>
<td>DC37</td>
<td>Data collection</td>
<td>Security/compliance</td>
<td>System to disallow simultaneous login/ use of the same account for a single network</td>
</tr>
<tr>
<td>DC43</td>
<td>Data collection</td>
<td>Security/compliance</td>
<td>System tracks and documents all uploads</td>
</tr>
<tr>
<td>DC44</td>
<td>Data collection</td>
<td>Security/compliance</td>
<td>System tracks and documents all validation checks and cleaning activities (entry specific)</td>
</tr>
<tr>
<td>DC46</td>
<td>Data collection</td>
<td>Security/compliance</td>
<td>System tracks and documents all exports (field-level detail)</td>
</tr>
<tr>
<td>DC48</td>
<td>Data collection</td>
<td>Security/compliance</td>
<td>System tracks and documents revisions to data (within database)</td>
</tr>
<tr>
<td>DC60a</td>
<td>Data collection</td>
<td>Security/compliance</td>
<td>System has ability to be in accordance with appropriate regulations for HIPPA</td>
</tr>
<tr>
<td>DC7</td>
<td>Data collection</td>
<td>User interface</td>
<td>System pre-populates fields based on earlier entries (form to form, pre-populates with info from network db)</td>
</tr>
<tr>
<td>DC8</td>
<td>Data collection</td>
<td>User interface</td>
<td>System provides selective questioning, based on earlier field entry (e.g., removes options/ fields that are no longer relevant) (skip patterns)</td>
</tr>
<tr>
<td>DC15</td>
<td>Data collection</td>
<td>User interface</td>
<td>System to notify/ remind authorized users of incomplete forms (saved, not submitted)</td>
</tr>
<tr>
<td>Requirement ID</td>
<td>Category</td>
<td>Subcategory</td>
<td>Requirement Description</td>
</tr>
<tr>
<td>----------------</td>
<td>------------------------</td>
<td>------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>DC61</td>
<td>Data collection</td>
<td>User interface</td>
<td>System provides ability to modify the “look” of the user interfaces: color schemes, fonts, layouts, logos/graphics</td>
</tr>
<tr>
<td>DC70</td>
<td>Data collection</td>
<td>User interface</td>
<td>System provides ability to modify the “look” of the user interfaces: color schemes, fonts, layouts, logos/graphics by an intermediate user</td>
</tr>
</tbody>
</table>

### 15.2 User/System Administration

<table>
<thead>
<tr>
<th>Requirement ID</th>
<th>Category</th>
<th>Subcategory</th>
<th>Requirement Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA1</td>
<td>System administration</td>
<td>Functionality</td>
<td>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</td>
</tr>
<tr>
<td>SA3</td>
<td>System administration</td>
<td>Functionality</td>
<td>System to provide information (to designated users) on usage (last time logged in/data entered, volume/frequency, by site/institution)</td>
</tr>
<tr>
<td>SA4</td>
<td>System administration</td>
<td>Security/compliance</td>
<td>System to allow user account permissions to be assigned for a finite or unlimited period of time</td>
</tr>
<tr>
<td>SA5</td>
<td>System administration</td>
<td>Security/compliance</td>
<td>System to capture requests for user/permission changes and forward to authorized resources</td>
</tr>
<tr>
<td>SA6</td>
<td>System administration</td>
<td>Security/compliance</td>
<td>System to automatically notify designated resources about upcoming account expirations</td>
</tr>
<tr>
<td>SA7</td>
<td>System administration</td>
<td>Security/compliance</td>
<td>System shows all current users and their permissions</td>
</tr>
<tr>
<td>SA9</td>
<td>System administration</td>
<td>Security/compliance</td>
<td>System to support multiple layers of permissions (admin, intermediate user (edit content), read only)</td>
</tr>
</tbody>
</table>
## 15.3 Reporting & Analysis

<table>
<thead>
<tr>
<th>Requirement ID</th>
<th>Category</th>
<th>Subcategory</th>
<th>Requirement Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAR1</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>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))</td>
</tr>
<tr>
<td>DAR5</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow creation of clinician-, site-, or patient-specific reports, including patient and aggregate data</td>
</tr>
<tr>
<td>DAR6</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow analysis/manipulation/presentation of qualitative and quantitative data</td>
</tr>
<tr>
<td>DAR12</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>Reports/analysis available in multiple formats (e.g., .pdf and others) that can be read on multiple platforms (computer, smartphone)</td>
</tr>
<tr>
<td>DAR14</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow presentation of graphical and tabular data (to include display of data over time==run charts)</td>
</tr>
<tr>
<td>DAR15</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow comments/narrative to accompany graphical data representations (<em>write-on</em> capabilities) (annotation)</td>
</tr>
<tr>
<td>DAR16</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow creation/modification of custom reports by expert user (graphical and tabular display with commentary/narrative)</td>
</tr>
<tr>
<td>DAR21</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow creation/modification of tabular/graphical reports by an expert user</td>
</tr>
<tr>
<td>DAR22</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow analysis/presentation of current and/or archived data</td>
</tr>
<tr>
<td>DAR3</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>Ability to generate extract to look at root-level data</td>
</tr>
<tr>
<td>DAR7</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>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</td>
</tr>
<tr>
<td>DAR8</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>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</td>
</tr>
<tr>
<td>-------</td>
<td>---------------</td>
<td>---------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>DAR10</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow import/upload of graphical outputs from external systems for inclusion in reports/analysis for read-only inclusion</td>
</tr>
<tr>
<td>DAR13</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow automated publication/push of approved reports to email, web</td>
</tr>
<tr>
<td>DAR17a</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow distribution of reports based on individually set preferences (identified/de-identified) to website</td>
</tr>
<tr>
<td>DAR18</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow graphical presentation of performance vs. benchmarks or standards (predetermined or calculated benchmarks)</td>
</tr>
<tr>
<td>DAR19</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow automated execution of pre-defined analysis/manipulation of a dataset (create 'standard' report upon receipt of clean data)</td>
</tr>
<tr>
<td>DAR23</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow validation/checking of data by authorized users</td>
</tr>
<tr>
<td>DAR24</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow modification of data in database by authorized users with proper training</td>
</tr>
<tr>
<td>DAR25</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>Logs and creates reports on errors (validation, import, export, etc.)</td>
</tr>
<tr>
<td>DAR29</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow for creation standard report templates</td>
</tr>
<tr>
<td>DAR30</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System to allow for design of standard report templates</td>
</tr>
<tr>
<td>DAR31</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>System can identify and store a snapshot of data used for analysis for a predetermined amount of time for a subset for the dataset</td>
</tr>
<tr>
<td>DAR33</td>
<td>Data analysis</td>
<td>Functionality</td>
<td>Reports are printable from system in a standard format</td>
</tr>
<tr>
<td>DAR11</td>
<td>Data analysis</td>
<td>Security/compliance</td>
<td>System tracks and documents revisions to data (within database) - audit trail</td>
</tr>
<tr>
<td>DAR28</td>
<td>Data analysis</td>
<td>Security/compliance</td>
<td>System tracks and documents which designated user inputs/modified data for each entry</td>
</tr>
<tr>
<td>DAR2</td>
<td>Data analysis</td>
<td>User interface</td>
<td>System to allow aggregation or drilldown of data by an intermediate user (e.g., site-specific resource, QIC, etc.)</td>
</tr>
</tbody>
</table>