

# **Slim-Prim: A Biomedical Informatics Database to Promote Translational Research**

*by Teeradache Viangteeravat, PhD; Ian M. Brooks, PhD; Ebony J. Smith, MS; Nicolas Furlotte, MS; Somchan Vuthipadadon, PhD; Rebecca Reynolds, PhD; and Chanchai Singhanayok McDonald, PhD*

## **Abstract**

With the current national emphasis on translational research, data-exchange systems that can bridge the basic and clinical sciences are vital. To meet this challenge, we have developed Slim-Prim, an integrated data system (IDS) for collecting, processing, archiving, and distributing basic and clinical research data. Slim-Prim is accessed via user-friendly Web-based applications, thus increasing data accessibility and eliminating the security risks inherent with office or laboratory servers. Slim-Prim serves as a laboratory management interface and archival data repository for institutional projects. Importantly, multiple levels of controlled access allow HIPAA-compliant sharing of de-identified information to facilitate data sharing and analysis across research domains; thus Slim-Prim encourages collaboration between researchers and clinicians, an essential factor in the development of translational research. Slim-Prim is an example of utilizing an IDS to improve organizational efficiency and to bridge the gap between laboratory discovery and practice.

**Key Words:** Bioinformatics, health management, clinical trial, basic research, laboratory management, data sharing

## **Introduction**

Advances in information management technology create opportunities for biomedical researchers to more easily share information. Technology allows clinical research and patient care to become more integrated and interactive. In so-called translational science, basic science and clinical researchers work together on interpretation and application of research data in clinical settings. Data sharing is necessary to improve the quality of healthcare and accelerate progress in biomedical sciences from bench to bedside to community. To go from clinical research to community practice, integrated data systems (IDSs) must be created to allow community researchers to easily access secure and confidential research data. These data can then be used to answer questions relevant to specific communities and can be extrapolated to a national level. Furthermore, information can be assimilated for community education to help improve healthcare. To address data integration issues, the Scientific Laboratory Information Management–Patient-care Research Information Management (Slim-Prim) system was developed by the Biomedical Informatics Unit (BMIU) at the University of Tennessee Health Science Center (UTHSC).

The National Institutes of Health (NIH) “Roadmap” identified a lack of communication between basic and clinical scientists as a major roadblock to the development of translational (bench to bedside) technologies (<http://nihroadmap.nih.gov/>). Highlighted in the “Roadmap” was the need for novel bioinformatics solutions to address this issue and to foster a climate of collaboration between basic

scientists and clinicians. The Slim-Prim system was initially developed in response to a need to integrate data from basic science and clinical research at UTHSC. The Slim-Prim system is expanding and increasing numbers of faculty and clinicians at the University of Tennessee are taking advantage of this system for their own research. See Figure 1 for a schematic overview of the Slim-Prim system.

## Literature Review

Interpreting data across multiple systems is challenging, and various integration techniques, with varying levels of complexity, have been proposed to solve this problem.<sup>1-4</sup> Nagarajan et al. introduced data-warehousing-based solutions utilizing relational database management systems (RDBMSs) for assembling and integrating data.<sup>5</sup> A relational database model is composed of classes of data, with each class characterized by a set of attributes. This conventional design is ideal for data sets composed of classes with a limited and fixed number of attributes. When each instance has values for all attributes (or columns) within a class (or table), the database is not filled with numerous null entries, and memory is used efficiently. However, research reveals that this design is not efficient for data sets with large numbers of attributes that vary over time.<sup>6</sup> Because most database engines limit the number of columns per table, they cannot accommodate massive numbers of class attributes. Also, continuously changing the number and type of attributes necessitates frequent modification of the database structure. Inefficient use of memory because of the large number of null entries is also a legitimate concern.

Recent research has proposed a knowledge-based terminology for identifying data dimensions in clinical informatics.<sup>7</sup> Other research has focused on the conceptual development of IDSs using ontology-based systems for the design and integration of clinical trial data.<sup>8</sup> Ontology-based systems allow users to conceptually design the database and integration processes independent of physical designs. However, development of novel ontologies is time consuming and challenging. The inherent variation between databases due to the different demands on each system means that there is no consensus on ontology and metadata descriptions. It might therefore be necessary to define a new ontology for each database. Although this approach gives the database designer freedom at the outset, inexperienced designers can spend excess time researching previous knowledge, seeking an optimum design. Where possible, designers should use preexisting ontologies. These can be modified as necessary to improve accessibility.

Finally, Wang et al. developed the BioMediator system to provide a theoretical and practical foundation for data integration across diverse biomedical domains via a “knowledge-base-driven centralized federated database” model.<sup>9</sup> However, the efficiency of query processing time and the need to filter out unnecessary query results still are concerns. The data architecture required for clinical data warehousing has been researched in applications such as clinical study data management systems (CDMSs) and clinical patient record systems (CPRSs). They both use an entity-attribute-value (EAV) system (i.e., row modeling) as opposed to conventional database design (column modeling).<sup>10</sup> The EAV system has the advantage of remaining stable as the number of parameters increases when knowledge expands, a common situation in the basic sciences and in clinical trials.<sup>11</sup>

In a system that uses the EAV structure, the referencing tables are composed of rows that consist of one or more facts about an entity. Each row in the referencing table consists of foreign keys to an entity, attribute(s) of the entity, and values for the attribute(s). If more facts about an entity must be entered, subsequent rows can be added with the same entity and different attributes and/or values. The foreign keys link to referenced tables. There are separate referenced tables for attributes, values, and metadata.

The EAV system has demonstrated promising results for assembling data from many different formats and sources into a centralized system via a global data table. However, solutions to the problems of indexing the system to improve the performance of queries across studies and to ensure better integration and navigation have not yet been found. Some database designers question the EAV model's scalability and performance over the conventional database design because each end-user query must unfold many EAV rows back into every record in the result set.<sup>12</sup> This makes the query less efficient and results in reduced performance compared to conventional databases.<sup>13</sup> Geisler et al. recently stated that the EAV system has limited efficiency when performing database management tasks such as indexing,

partitioning, query optimization, ad hoc querying, and data analysis.<sup>14</sup> Systems such as TrialDB (Deshpande et al.) have attempted to address these problems using the EAV system.<sup>15</sup> However, instead of using a single EAV table, TrialDB has a separate EAV table for each data type in the column (e.g., string, integer, real, date, etc.). Deshpande et al. state that with a metadata-driven EAV warehouse, maintenance no longer involves the laborious redesign and reloading of multiple tables required of conventional database designs.<sup>16</sup> All of these considerations were taken into account when developing Slim-Prim. The ultimate choice of a traditional column modeling system with an ontological base is discussed further below.

## **Development of the Slim-Prim System: Metadata**

Clearly, the most important factor driving effective query results, in terms of searching results, grouping data for integration, allowing for correct manipulation, and so forth, is metadata design.<sup>17</sup> The attributes in the metadata are designed to contain all the crucial identification keys, not only to handle data storage and retrieval, searching and sorting, and reporting, but also to handle data in various formats (e.g., numbers, files, text, images, ICD-9 codes). For instance, metadata contain names, descriptions, keywords, control vocabulary, data type, and other fields that are necessary for future mapping, integration, and navigation purposes (Figure 2). The attribute values are stored in the table based on their own data type specified in the metadata table. All the attribute descriptions and values can also be linked together with related identification keys (e.g., an instance object key).

## **Application Programming Interface**

The application programming interface (API) of the Slim-Prim system is based on an object-oriented abstraction of the underlying data structure. A Unified Modeling Language (UML)-based tool is used to create the structure diagrams (e.g., class, object, element, composite structure, package, and deployment diagrams), workflow diagrams (e.g., use case, activity, and state machine diagrams), and interaction diagrams (e.g., sequence, communication, timing, and interaction overview diagrams). To ensure the future capability of federating data with national databases such as the National Institutes of Health (NIH) cancer Biomedical Informatics Grid (caBIG), close attention was paid to caBIG compatibility requirements. In the Slim-Prim maturity model, elements such as programming and messaging interfaces, vocabularies or terminologies, and ontology are tightly constrained; thus the Slim-Prim system is well designed to bridge the gap between laboratory discovery and clinical practice.

## **Patient and Research Subject Screening Tool**

We have found that potential subjects are more likely to use a Web site to obtain information about clinical trials and enrollment criteria than they are to use more traditional methods of obtaining information (e.g., brochures). The patient recruiting and research subject screening tool is designed to inform and motivate potential participants and to prescreen and validate subjects for research. All data entered by potential subjects are stored and are reported to data coordinators for review and validation. The system provides a confidential questionnaire that prequalifies subjects according to criteria predetermined by the clinical investigator. The individual is informed immediately if he or she does not qualify for the study. Because the Web application identifies suitable subjects, recruiters have more time to speak with valid participants. If they do qualify, the system provides a list of trial sites, telephone numbers, and a map to the chosen location.

## **Content Management System**

Scientific and clinical research databases often contain large amounts of data whose format may not always be known in advance. Furthermore, these databases may be accessed by a variety of users with different viewing and editing privileges. The Slim-Prim system uses an open-source content management system, ContentNOW, that uses a row-modeling structure. Using concept mapping, the row-modeling design assembles data from different formats and sources into a centralized system. A global data table (metadata) is used as a template to create modules to enhance flexibility and functionality. For example,

one study of post-transplantation obesity is currently using Slim-Prim to create complex forms; a repository for genetic, environmental, and lifestyle data; and tools to assist in basic data analysis. The Slim-Prim application for this study includes (1) a screening-report form to collect prescreening data from prospective patients; (2) enrollment report forms that incorporate existing information from screening reports; (3) self-test measures of depressive feelings and behaviors (using the Center for Epidemiologic Studies Depression Scale [CES-D] survey form) for determining depression level; and (4) a report generator system for prescreening data analysis and preparing any nutrition data analysis. These forms and features can be easily adapted for other studies (Figure 3).

## **Biorepository Web Application**

A biorepository application allows Slim-Prim to coordinate the efforts of pathology department tissue core units. Tissue core units usually store paraffin-wax-embedded and frozen tissue sections from biopsies and other surgical interventions. Tissue samples can be linked with patient EMRs to generate an ongoing clinical narrative of disease progression, treatment, and remission. The BMIU is working with researchers at the University of Tennessee Cancer Institute (UTCI) to create a pharmacogenetic repository containing the linked tissue samples and EMRs of cancer patients treated at the UTCI centers. This will provide an organized, centralized, and Web-accessible pharmacogenetic database system, a common goal of translational informatics. The Slim-Prim biorepository's Web-based user interface includes a data entry function, an inventory database, a specimen withdrawal database, a barcode generator database, and a report generator for data analysis.

## **Data Aggregation from Multiple Data Sources**

The Slim-Prim system supports data migration from other database systems and transforms the data into standardized formats. This allows researchers to review data according to their research design. For example, 10 years of triennial data from the Kids' Inpatient Database (KID) was acquired from the Department of Health and Human Services Healthcare Cost and Utilization Project (HCUP), transformed from ASCII into CSV format, and then uploaded into the Slim-Prim system. These data represent almost seven million pediatric hospital discharges from patients with a variety of diagnoses and comorbidities. The Slim-Prim report system also allows complex queries in order to limit the data set. Data are downloadable in Excel spreadsheet format for statistical analysis via the user's chosen software. See Figure 4 for sample screen shots from this site.

## **Slim-Prim Security and IT Infrastructure**

Data security was addressed stringently during the development of Slim-Prim. UTHSC has a risk-based security management system focused on the confidentiality, availability, and integrity of data. UTHSC is a covered entity under HIPAA and conducts research governed by various federal and state laws and regulations, and thus has active programs to insure compliance with these statutes. A key concept of the Slim-Prim system is fostering collaboration under appropriate levels of institutional review board (IRB) and HIPAA compliance. Because Slim-Prim is a Web-based IDS, data can be accessed from anywhere in the world. Access to Slim-Prim is strictly controlled by the principal investigator (PI) of each study: he or she alone chooses how to share data with collaborators. In this way, Slim-Prim makes it easy to provide collaborators with fully de-identified data or data elements.

## **Future Functionality**

The system is being expanded to include a server-mounted geographic information system (ArcGIS, ESRI Inc., Palo Alto, CA). This will significantly expand the user's ability to analyze healthcare data in conjunction with a range of geographic information. Further modular expansion will provide storage and basic analytics for DNA microarray data and confocal microscope images, for example.

## Discussion

The Slim-Prim system was implemented to support translational research by integrating data from different sources. It is clear that the clinical and translational research communities need a Web application allowing easy access to data and rapid data sharing across labs and institutions. Such tools must be designed with strict research compliance and security in mind; thus, robust levels of access have been included as key components of the Slim-Prim system. Slim-Prim facilitates and enhances the effectiveness of scientific laboratory management, allowing researchers to customize their own applications to collect data on laboratory materials, tools, and budgets in different formats. This system also uses patient-care information tools for data collection, storage, retrieval, searching, and reporting, such as online screening forms, online applications, and medical forms as well as patient folders containing demographic, lab, and historical data. Importantly, the expandable and modular nature of the system also allows storage of radiology and image-file data, rehabilitation data, and DNA sequencing files, among other types of data. These scientific and patient-care informational tools are designed to facilitate community research and encourage sharing with other researchers and other communities. These tools also have the capability to de-identify certain sensitive data from the data set prior to transfer between community researchers or coinvestigators.

## Database Development

Choosing a database design that reflects the real-world representation of data, while providing the flexibility and robustness needed for large studies across multiple platforms, is a crucial step. Most current database systems have a rigid structure that is constructed at the beginning of a study. Once a study has commenced, the structure of the database usually remains static. If during the study there is a need to make changes to the structure of the database, a database specialist must be brought in to implement the changes. Many times this involves rebuilding the entire database structure. After reviewing the EAV approach and the conventional design, we first chose to implement an EAV database design. We based our initial design decision on the likely needs of the users of Slim-Prim and the kinds of data the users will encounter in their work.

Because the system focuses primarily on scientific and patient-care data, the data sets involved are characterized by sparseness, volatility, and duplication. Sparseness refers to entries having only a small subset of nonempty values for the attributes within a class. Volatility refers to the likelihood of the attributes to grow and change over time. Duplication refers to the possibility that an entity may have multiple values for a single attribute. Scientific data collection that occurs during clinical trials conducted over time falls into these categories. Healthcare data is traditionally composed of multiple values for single attributes; for example, a child's height and weight change over time. Healthcare data are also characterized by empty values of specific attributes within a data set. For example, male patients will not have values of attributes related to menstruation or pregnancy. The nature of scientific discovery, especially in regard to patient care, is one of almost constant growth as new procedures and diagnoses emerge. Thus, the attributes within a class of patient-care or scientific data will evolve and grow with technology. All of these characteristics of scientific and patient-related data sets led to the initial choice of EAV design for the Slim-Prim system.

To determine the usefulness of the EAV approach, we performed an analysis of the KID data set, described above. We acquired these data and queried them for a specific pediatric complaint: how many records had an ICD-9 diagnosis code of "7513" (Hirschsprung's Disease, a moderately rare bowel complaint affecting approximately 1 in 5,000 live births). To provide an answer, an EAV row-based design would be required to process approximately 500,000 input/output redirections (I/Os) because a row-based process would need to deal with large amounts of irrelevant data. This would obviously prohibit efficient database management through indexing, partitioning, and ad hoc querying. However, with a column-based design, the answer would require approximately 234 I/Os. This means the data retrieval from a column-based design is much faster and efficient. Thus our nascent database was rebuilt in a conventional column-based design. To overcome problems with adaptability and complex querying of heterogeneous data, we developed an ontology model to underlie the system.

## Use of Metadata and Ontology

Often users do not consider data sharing as a prime motivator in database design and thus do not plan to use standard vocabularies in the first place. While many medical terminology standards exist and are available for use with clinical trials, we have found that most users prefer to use terminology based on personal knowledge. Complex queries can be difficult due to the sparseness and volatility of data elements. We observed the use of the Web Ontology Language (OWL), among others, in development of the semantic web, and decided to model the underlying metadata structure of Slim-Prim using an ontological template. Multiple ontologies can define an overarching structure for a complex database and constrain vocabularies therefore enforcing relationships between data elements. Thus an ontology is used for the definition of the structure and meaning of data stored in a database.<sup>18</sup> To assemble data from different formats and multiple data sources into a centralized system, the Slim-Prim system uses a metadata-driven model that contains built-in knowledge bases (such as SNOMED-CT, ICD-9-CM, CPT-4, LOINC, and other common standards) to control medical terminology. The multiple ontology approach is then used to allow definition of a separate ontology to fit with each trial. By controlling these structures and their content, Slim-Prim helps to enforce controlled data practice across the University of Tennessee Clinical and Translational Science Institute (CTSI).

## Limitations of Slim-Prim

While Slim-Prim is proving invaluable in modernizing the flow of research at the University of Tennessee, we have set reasonable limits on the functionality of Slim-Prim. For example, its report functions allow users to query massive data sets, but the BMIU has no intention of reinventing current advanced statistical software packages, such as SAS. The system allows users to ask complex questions regarding their data and then isolate the resulting limited data sets for external examination via other sources. We provide Slim-Prim as a relational database for storing and querying heterogeneous data. Collaborations are planned to link out from Slim-Prim to other bioinformatics tools for such analyses as literature mining, genomic data analysis, or DNA microarray data analysis.

All of the modules implemented within the Slim-Prim system are active in their own domains, and many have begun to work across boundaries. For example, UTHSC and its partners have collaborated on numerous projects and submitted several NIH grant applications. However, Slim-Prim provides the infrastructure for much greater integration. The creation of the IDS supports data sharing and research design with CTSI resources, creating a comprehensive and interoperable architecture. In addition, the BMIU links these diverse resources with investigators in basic science, clinical research, and public health areas to promote the development and implementation of translational science. Ultimately the development of translational science methodologies depends on the users. It is they who must ask the right questions and tackle the appropriate projects. Slim-Prim is a valuable tool, but it is the imaginations and demands of its users who will guide the future expansion of Slim-Prim and development of clinically relevant research.

## Conclusions

Development and implementation of translational science technologies demand a close interaction between basic and clinical research. To effect this synergy, the BMIU at the University of Tennessee has developed Slim-Prim, a user-friendly, versatile, and secure Web-based system for data and knowledge management. Slim-Prim is currently active in several basic science and clinical research studies, with research modules demonstrating various aspects of its functionality. Slim-Prim is modeled on a relational database and employs a metadata-driven API. Its use of controlled vocabularies constrains data entry and enforces levels of interoperability, allowing for federation of research data with national databases such as caBIG.

Teeradache Viangteeravat, PhD, is assistant director in the Biomedical Informatics Unit at the University of Tennessee Clinical and Translational Science Institute and assistant professor in the Department of Preventive Medicine at the University of Tennessee Health Science Center in Memphis, TN.

Ian M. Brooks, PhD, is a research associate at the University of Tennessee Clinical and Translational Science Institute in Memphis, TN.

Ebony J. Smith, MS, is a research assistant at the University of Tennessee Clinical and Translational Science Institute in Memphis, TN.

Nicolas Furlotte, MS, is a research assistant at the University of Tennessee Clinical and Translational Science Institute in Memphis, TN.

Somchan Vuthipadadon, PhD, is a postdoctoral fellow at the University of Tennessee Clinical and Translational Science Institute in Memphis, TN.

Rebecca Reynolds, PhD, is an associate professor and director of health informatics and information management at the University of Tennessee Health Science Center in Memphis, TN.

Chanchai Singhanayok McDonald, PhD, is the codirector of the Biomedical Informatics Unit at the University of Tennessee Clinical and Translational Science Institute and associate professor in the Department of Preventive Medicine at the University of Tennessee Health Science Center in Memphis, TN.

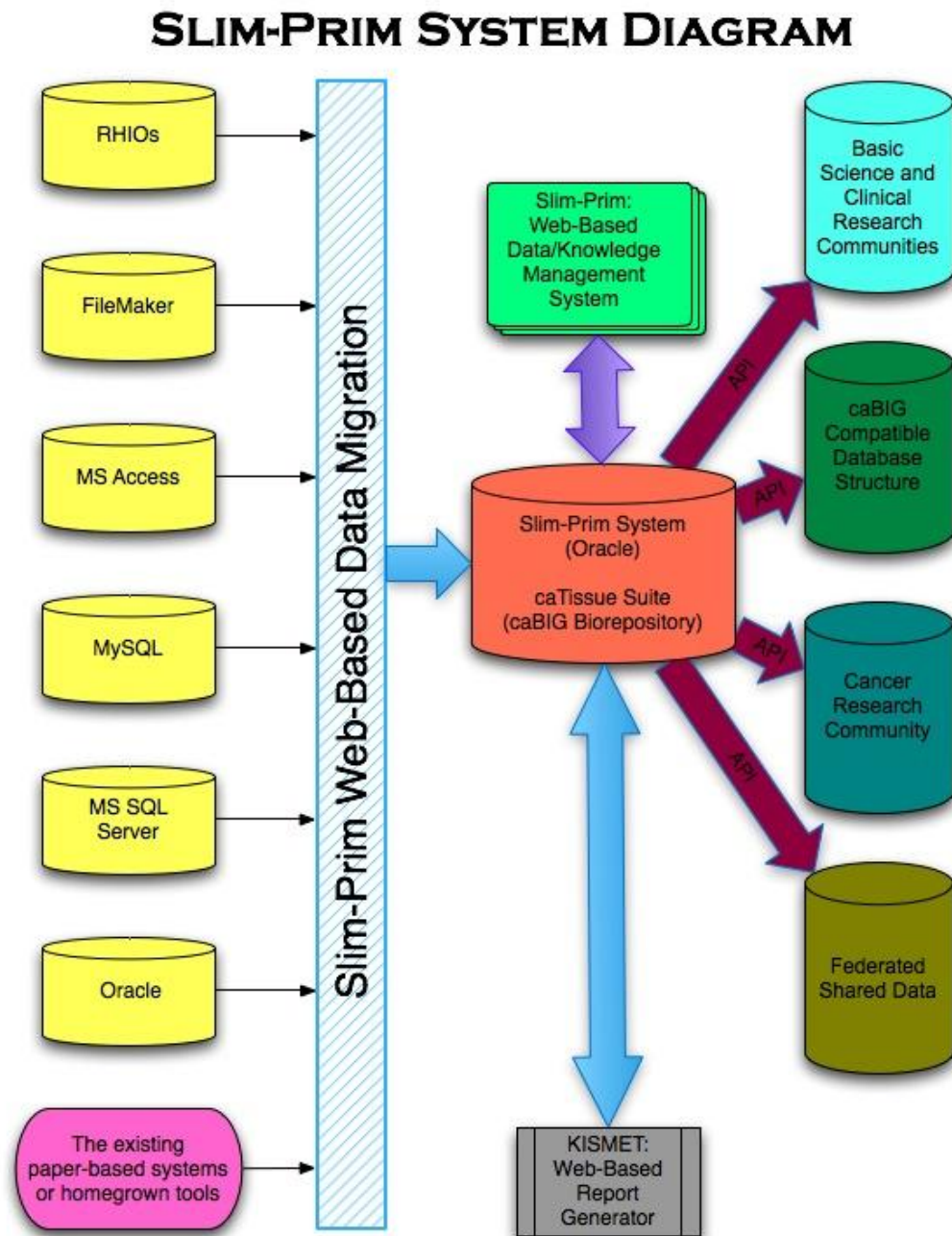
## Notes

1. Brazhnik, O., and J. Jones. "Anatomy of Data Integration." *Journal of Biomedical Informatics* 40, no. 3 (2007): 252–269.
2. Geisler, S., A. Brauers, C. Quix, and A. Schneink. "Ontology-based System for Clinical Trial Data Management." *Proceedings: Annual Symposium of the IEEE/EMBS Benelux Chapter*. Heeze, the Netherlands: IEEE 2007. pp. 53-55. 2007.
3. Wang, K., P. Tarczy-Hornoch, R. Shaker, P. Mork, and J. F. Brinkley. "BioMediator Data Integration: Beyond Genomics to Neuroscience Data." *AMIA Annual Symposium Proceedings* (2005): 779–783.
4. Nagarajan, R., M. Ahmed, and A. Phatak. "Database Challenges in the Integration of Biomedical Data Sets." *Proceedings of the Thirtieth International Conference on Very Large Data Bases*. Toronto: VLDB Endowment, 2004.
5. Ibid.
6. Dinu, V., and P. Nadkarni. "Guidelines for the Effective Use of Entity–Attribute–Value Modeling for Biomedical Databases." *International Journal of Medical Informatics* 76, no. 11 (2007): 769–779.
7. Brazhnik, O., and J. Jones. "Anatomy of Data Integration."
8. Geisler, S., A. Brauers, C. Quix, and A. Schneink. "Ontology-based System for Clinical Trial Data Management."
9. Wang, K., P. Tarczy-Hornoch, R. Shaker, P. Mork, and J. F. Brinkley. "BioMediator Data Integration: Beyond Genomics to Neuroscience Data."
10. Deshpande, A. M., C. Brandt, and P. M. Nadkarni. "Metadata-driven Ad Hoc Query of Patient Data." *Journal of the American Medical Informatics Association* 9 (2002): 369–382.
11. Anhøj, J. "Generic Design of Web-Based Clinical Databases." *Journal of Medical Internet Research* 5, no. 4 (2003): e27.
12. Hughes, R. "Optimal Data Architecture for Clinical Data Warehouses." *Information Management*, vol. 14 (2004): 46.
13. Corwin J., A. Silberschatz, P. L. Miller, and L. Marengo. "Dynamic Tables: An Architecture for Managing Evolving, Heterogeneous Biomedical Data in Relational Database Management Systems." *Journal of the American Medical Informatics Association* 14 (2006): 86–93.
14. Geisler, S., A. Brauers, C. Quix, and A. Schneink. "Ontology-based System for Clinical Trial Data Management."
15. Deshpande, A. M., C. Brandt, and P. M. Nadkarni. "Metadata-driven Ad Hoc Query of Patient Data."
16. Ibid.
17. Fisk, J. M., P. Mutalik, F. W. Levin, J. Erdos, C. Taylor, and P. Nadkarni. "Integrating Query of Relational and Textual Data in Clinical Databases." *Journal of the American Medical Informatics Association* 10, no. 1 (2003): 21–38.
18. Jean, S., Y. Ait-Ameur, and G. Pierra. "Querying Ontology Based Database Using OntoQL (An Ontology Query Language)." *On the Move to Meaningful Internet Systems*, vol. 4275 (2006): 704-721.

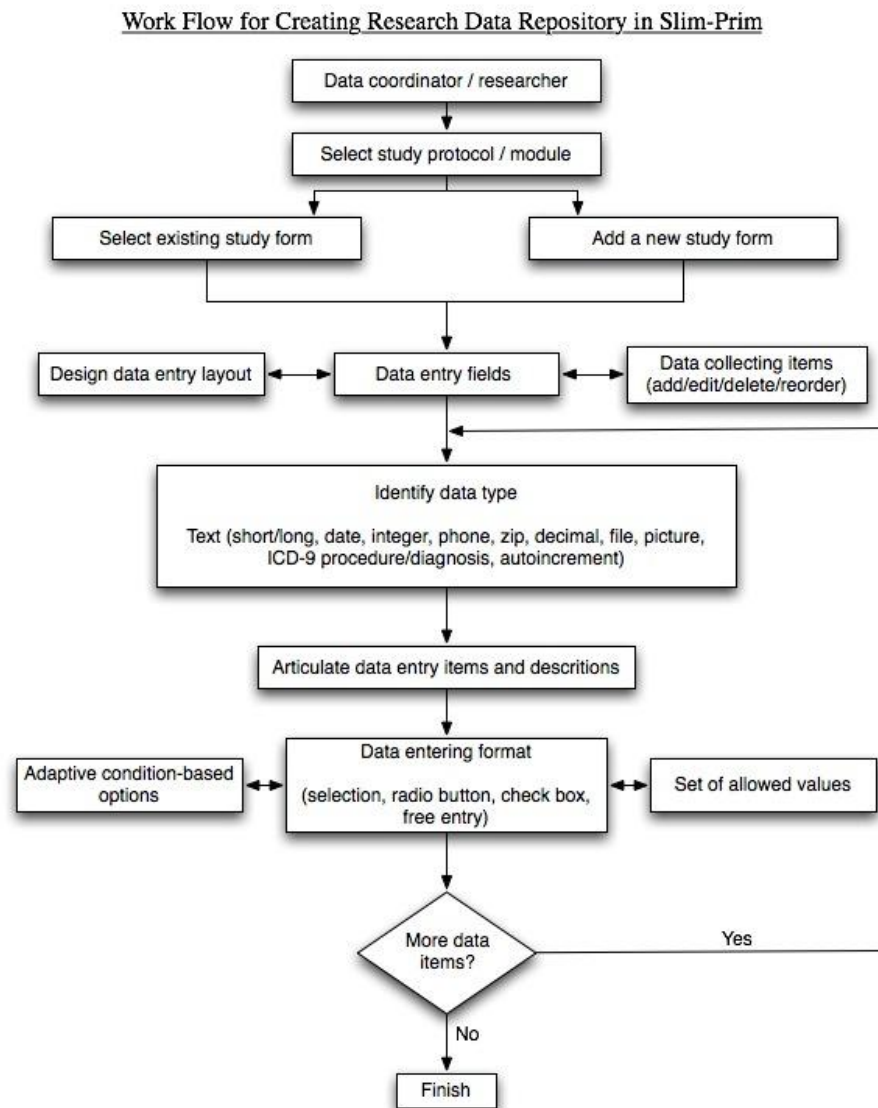


**Figure 1**

Overview of the Slim-Prim System



*Note:* Data may be uploaded to Slim-Prim from a variety of existing sources. Metadata content of Slim-Prim is controlled through use of standardized data elements and vocabularies ensuring upstream compatibility for future data federation.

**Figure 2****Work Flow of Metadata-Based Data Element Design**

*Note:* The Slim-Prim metadata system allows each data coordinator or researcher to design forms appropriate for his or her data type. Once data have been migrated or uploaded to Slim-Prim, they can be linked via unique identifiers allowing data analysis across multiple forms.

### Figure 3

#### Screen Shots from a Sample Patient Recruitment Form

|                                |   |
|--------------------------------|---|
| 1. Date collected *            |   |
| 2. Visit Code *                | Random Sample Diet History  |
| 3. Patient's Last Name *       | 17. Relationship to patient   |
| 4. Patient's First Name *      | 18. Has this patient been evaluated for and positively satisfied the absolute and relative contraindications and risk factors for transplantation used at this Institute? |
| 5. Patient's MI                | <input type="radio"/> yes   |
| 6. Social Security Number      | <input type="radio"/> no  |
| 7. Patient's DOB *             | 19. If yes, are they approved to receive a living related, living unrelated, or cadaveric donor renal transplant? (check all that apply)                                  |
| 8. Age                         | <input type="checkbox"/> Living related   |
| 9. Weight                      | <input type="checkbox"/> Living unrelated   |
| 10. Weight measured in         | <input type="checkbox"/> Cadaveric  |
| 11. Height (ex: #' #", # feet) | 20. Is this the first transplant for this patient?  |
| 12. Height measured in         | <input type="radio"/> no  |
| 13. BMI                        | 22. Does this patient meet the inclusion criteria for this study?   |
| 14. BMI status                 | 23. Outcome: (Check all that apply)   |
| 15. Does this patient posse    | <input type="checkbox"/> Appointment previously rescheduled   |
| 24. If Enrolled, as            | <input type="checkbox"/> No-Show  |
| 25. If Enrolled, as            | <input type="checkbox"/> Patient unavailable  |
|                                | <input type="checkbox"/> Refused  |
|                                | <input type="checkbox"/> I will decide later. Please call me  |
|                                | <input type="checkbox"/> I will decide later. Don't call me. I'll call you  |
|                                | <input type="checkbox"/> Enrolled - ICF signed & witnessed  |
|                                | 24. If Enrolled, assign study site ID#:   |
|                                | 25. If Enrolled, assign Patient ID  |
|                                | 26. If Refused, Reason, if offered  |
|                                | 27. If decide later, please call me. What is the Phone number?  |
|                                | 28. If decide later, don't call me. I'll call you   |
|                                | <input type="checkbox"/> Took business card   |
|                                | <input type="checkbox"/> Took consent forms   |

*Note:* These sample forms are from a study on transplants and obesity; however, forms are customizable for any study. The study administrator can make any question compulsory, forcing patients or research coordinators to enter certain data. Adaptive conditions can be added as necessary to prevent, for example, male enrollees' being asked questions regarding pregnancy.

**Figure 4**

Screen Shots from the KID Database Module

The screenshot displays the 'TAB content' configuration interface for a 'Diagnosis' table. On the left, a sidebar shows a tree view with 'Diagnosis' selected. The main configuration area includes fields for 'db Table Name' (Diagnosis), 'Content Type' (Dynamic (Import From)), and 'Startdate' (29.10.2008 15:28). Below these are 'Save' and 'Save/Close' buttons, and options to 'Add New Question' or 'Configure Form'. A note states: 'Note: Drag and drop questions to reorder. Move'. A table lists 11 questions with columns for 'Question Number', 'Name', 'Description', and 'Description'. The preview on the right shows the user-facing form with fields for 'Sequential record number', 'Clinical Classifications Software, formerly known as Clinical Classifications for Health Policy Research: Principal diagnosis', 'Number of diagnoses for this discharge', and three 'Diagnosis 1 (ICD-9)', 'Diagnosis 2 (ICD-9)', and 'Diagnosis 3 (ICD-9)' fields, each with a 'valid code' and 'invalid code' radio button. At the bottom, a table lists the questions and their properties:

| Question Number | Name   | Description   | ICD9 Diagnoses | free entry | no | edit delete |
|-----------------|--------|---|----------------|------------|----|-------------|
| 1               | recnum | Sequential reco   |                |            |    |             |
| 2               | dchpr1 | Clinical Classif<br>Clinical Classif<br>Principal diagn |                |            |    |             |
| 3               | ndx    | Number of diag  |                |            |    |             |
| 4               | dx1    | Diagnosis 1 (IC   |                |            |    |             |
| 5               | dxv1   | Diagnosis 1 val   |                |            |    |             |
| 6               | dx2    | Diagnosis 2 (IC   |                |            |    |             |
| 7               | dxv2   | Diagnosis 2 val   |                |            |    |             |
| 8               | dx3    | Diagnosis 3 (ICD-9)                                     | ICD9 Diagnoses |            | no | edit delete |
| 9               | dxv3   | Diagnosis 3 validity flag                               | text (short)   | radio      | no | edit delete |
| 10              | dx4    | Diagnosis 4 (ICD-9)                                     | ICD9 Diagnoses | free entry | no | edit delete |
| 11              | dxv4   | Diagnosis 4 validity flag                               | text (short)   | radio      | no | edit delete |
|                 | dx5    | Diagnosis 5 (ICD-9)                                     | ICD9 Diagnoses | free entry | no | edit        |

*Note:* The screen shot in the background shows the diagnosis form as viewed from the administrator/PI page. These forms can be created by PIs with little experience in database management because of Slim-Prim’s user-friendly interface. The inset shows the page rendered for the user site, as seen by users wishing to query the data.