Design and implementation of an institutional case report form library

Meredith Nahm, PhD[a], **John Shepherd**[b], **Ann Buzenberg**[b], **Reza Rostami, MBA**[b], **Andrew Corcoran**[b], **Jonathan McCall**[b], and **Ricardo Pietrobon, MD, PhD, MBA**[a,c]

[a]Duke Translational Medicine Institute, Durham, NC

[b]Duke Clinical Research Institute, Durham, NC

[c]Duke University Medical Center, Durham, NC

Abstract

Background—Case report forms (CRFs) are used to collect data in clinical research. Case report form development represents a significant part of the clinical trial process and can impact study success. Libraries of CRFs can preserve the organizational knowledge and expertise invested in CRF development and expedite the sharing of such knowledge. Although CRF libraries have been advocated, there have been no published accounts reporting institutional experiences with creating and using them.

Purpose—We sought to enhance an existing institutional CRF library by improving information indexing and accessibility. We describe this CRF library and discuss challenges encountered in its development and implementation, as well as future directions for continued work in this area.

Methods—We transformed an existing but underused and poorly accessible CRF library into a resource capable of supporting and expediting clinical and translational investigation at our institution by (1) expanding access to the entire institution; (2) adding more form attributes for improved information retrieval; and (3) creating a formal information curation and maintenance process. An open-source content management system, Plone (Plone.org), served as the platform for our CRF library.

Results—We report results from these three processes. Over the course of this project, the size of the CRF library increased from 160 CRFs comprising an estimated total of 17,000 pages, to 177 CRFs totaling 1.5 gigabytes. Eighty-two of these CRFs are now available to researchers across our institution; 95 CRFs remain within a contractual confidentiality window (usually 5 years from database lock) and are not available to users outside of the Duke Clinical Research Institute (DCRI). Conservative estimates suggest that the library supports an average of 37 investigators per month. The resources needed to curate and maintain the CRF library require less than 10% of the effort of one full-time equivalent employee.

Limitations—Although we succeeded in expanding use of the CRF library, creating awareness of such institutional resources among investigators and research teams remains challenging, and requires additional efforts to overcome. Institutions that have not achieved a critical mass of attractive research resources or effective dissemination mechanisms may encounter persistent difficulty attracting researchers to use institutional resources. Further, a useful CRF library requires both an initial investment of resources for development, as well as ongoing maintenance once it is established.

Correspondence to: Meredith Nahm, PhD, Duke Translational Medicine Institute, Box 3850 Med Ctr Durham, NC 27710, Hock Plaza, Suite 500, 2424 Erwin Rd, Durham, NC 27705, Tel: (919) 668-8339; (919) 668-8478, meredith.nahm@duke.edu.

Conclusions—CRF libraries can be established and made broadly available to institutional researchers. Curation—*i.e.*, indexing newly added forms—is required. Such a resource provides knowledge management capacity for institutions until standards and software are available to support widespread exchange of data and form definitions.

Keywords

Data collection; Clinical research; Clinical trial; Case report form; Knowledge management

---

Introduction

Clinical research has long depended on data collection instruments, generally known as *case report forms* (CRFs), to structure and facilitate collection of data for clinical trials. Most CRFs are customized to collect data specific to a particular clinical study protocol. CRF libraries, which are recommended by both academic texts and industry standards [1,2], can serve as repositories of institutional knowledge that will benefit future research and researchers. We describe our experiences in attempting to update, improve, and increase the institutional use of an existing electronic CRF library at a large academic medical center.

Data collection challenges in academic research

Academic research organizations face particularly acute challenges with regard to data collection. A large academic medical center may have thousands of simultaneously ongoing clinical trials, which together comprise a mix of investigators participating in industry-sponsored research (usually obliged to use sponsor-designed forms), government-sponsored research (often using forms designed by a central trials coordinating center), and single-site or investigator-initiated studies (using forms designed by the local investigator's research team); the lattermost is characterized by teams that independently collect and manage data using systems chosen according to project scope and budget limitations. When faculty, fellows, and staff move to other institutions, knowledge about studies conducted during their tenure often leaves with them. The creation of CRF libraries presents an opportunity to reclaim and preserve historical institutional knowledge locked in CRFs, so that the knowledge can be shared for the benefit of future research and researchers. However, the creation of such knowledge resources presents three informatics challenges: (1) identification of the knowledge sources likely to benefit future research and researchers; (2) identification of information worthy of preservation; and (3) obtaining and indexing each piece of information to facilitate retrieval by future users.

Our institution (Duke University) has more than 3000 ongoing clinical studies, many of them investigator-initiated. As is the case for many academic medical centers, centralized infrastructure for data collection and management was not in place until relatively recently.

Prior work

For the purposes of this manuscript, a CRF may comprise several *forms*. A *form* may be one or more pages in length and may consist of one or more *modules*. A *module* is a group of logically related data elements collocated within a form, e.g., vital signs, demographic characteristics, laboratory test results, or the SF-36 are considered to be *modules*. Wherever CRF libraries exist, standard operating procedures (SOPs) typically stipulate that standard forms or modules from the library should be used when developing CRFs. While this practice has been implemented and shown to be effective by large pharmaceutical companies [3], broadly applied SOPs that dictate the use of specific forms are less feasible in academia, where research is often marked by established patterns of local control, great diversity, and lack of central infrastructure supporting research.

In 2002, the Clinical Data Interchange Standards Consortium (CDISC) created a Web-based CRF repository, the Collaborative Standards Forum, which allowed sample forms to be posted and viewed by members [4]. The Forum has the stated mission of "…augment(ing) and facilitat(ing) CDISC work, specifically by providing and managing an interactive web site to encourage the sharing and development of additional standards that will benefit the biopharmaceutical industry"[5]. Many organizations, however, hesitate to contribute their proprietary forms to this semi-public domain, while others are contractually unable to do so because of existing confidentiality agreements with trial sponsors. Other libraries include one maintained by the Medical University of South Carolina [6], and the OpenClinica (Alkaza Research, LLC) electronic CRF library [7]. In addition, forms from many National Heart, Lung and Blood Institute (NHLBI) trials are available on the NHLBI Web site, and forms from other NIH-funded trials are available through the National Technical Information Service [1].

Some commercially available clinical data management and electronic data capture systems, including Oracle Clinical (Oracle Corp.), InForm (Phase Forward, Inc.), and DataFax (DataFax Systems, Inc.), offer library features that serve as local data element or form repositories for system users. Their use, however, often is limited to database developers and data managers using these specific systems. Metadata registries (also called *data element registries*) such as the ISO 11179-based cancer data standards repository (caDSR) maintained by the National Cancer Institute provide data element viewing at both the data element and form level.

While we could exploit data dictionaries for clinical trials, they would be in formats specific to the source data system (*e.g.,* Clintrial or InForm). Systems available to investigators today cannot directly use data dictionary information from other systems or otherwise use such information to automate the building of data collection screens. There are some systems that can consume the CDISC Operational Data Model (ODM) information in this way, but these are few in number, as are systems that can export a study's data dictionary in ODM format.

Unfortunately, standards governing the exchange of data elements do not yet exist. Once these standards become available, each system will need to be upgraded so that it can use the standards to automate or facilitate creation of data collection screens from the data element definition. At our institution, investigators use a variety of software packages and approaches to manage data, including InForm, Oracle Clinical, hosted systems, REDCap (Vanderbilt University), "homegrown" systems, Microsoft Access, statistical analysis packages, and spreadsheets. The system of choice depends on the size of the project, therapeutic area, available resources, and investigator preference. Thus, storing information about the individual data elements from CRFs, while a best practice, today would not yet increase the usefulness of the forms for our investigators. In the absence of this ideal, images of the data elements as collected on the form meet today's needs to—preserve and share institutional knowledge about what data were obtained for a study, and how they were collected.

The graphical form (page)-level representation afforded by image-based CRF libraries is cognitively closer to the researcher's goal of designing a data collection form than is a list of data elements. Thus, according to the proximity compatibility principle, form representations are likely easier for researchers to use when designing data collection tools, suggesting that in addition to storing individual data elements, a form- (page or module) level representation should be preserved [8]. Identifying salient, representational features of forms and models, as well as using them to facilitate automated screen creation, are areas for further research.

Clinical and Translational Science Awards and the Duke CRF library project

The National Institutes of Health's Clinical and Translational Science Awards (CTSA) [9], of which our institution is a recipient, are intended to enhance the availability of infrastructure supporting clinical and translational research. Form design, a process that includes the identification of data to be collected, creation of data definitions, and graphical representation, is a vulnerable point in the research enterprise. Not only is form design subject to the time pressure of the critical path in study start-up; it also is an infrequently-performed task for most investigators. Mistakes in form design can affect the whole trial; for this reason, the opportunity to impart institutional knowledge to individual investigators as part of the form design process can have a favorable impact.

Creation of effective CRF libraries helps ensure that: (1) important data are captured; (2) investigators are aware of applicable standards; (3) field-tested forms are available; (4) exhaustive pick-lists with mutually exclusive topic categories are available; and (5) other data definition and good form design principles are used. The development of a CRF library at our institution represents a natural experiment within the context of a CTSA institution, one aimed at ascertaining the willingness of academic investigators to share and use CRFs within their local academic community and determining the overall acceptance of such a resource within the context of academic research, as evaluated by the number of page views for CRF library pages and the number of forms contributed to the library.

Early attempts at creating an institutional CRF library

In 2001, leaders at our institute for coordination of multicenter trials, the Duke Clinical Research Institute (DCRI), sought to provide additional support to investigators by creating an electronic CRF library, which was made available to faculty and staff through 100 licenses. The electronic library was Web-based, accessed through an institutional intranet, and designed to contain CRFs used in multicenter clinical trials coordinated by the institute. Forms were indexed manually searchable by trial name, therapeutic area, and research sponsor.

The library as initially instantiated was not widely used for several reasons. First, the 100 licenses were allocated to named users according to the software licensing agreement, which meant that users had to be managed and tracked and access to the library had to be controlled. Second, the 100 licenses proved inadequate for the needs of the Institute, which employs more than 900 faculty and staff. Third, the tasks of indexing and adding forms to the library were classified as infrastructure development and hence were not billable to research sponsors, resulting in a shortage of resources to devote to the task. Finally, the library was maintained by the DCRI; investigators and research teams from single-site and investigator-initiated trials across the larger Duke research community were unaware of the resource and would have encountered significant barriers in accessing it. Given these shortcomings, we sought to retool the library for this broader research community, including all clinical and translational researchers and research teams.

Methods

System requirements

In our efforts to redesign the CRF library for the broader Duke research community, we faced two significant challenges: (1) providing access for a significantly expanded user base (more than 5000 users) and (2) creating a sustainable maintenance process. We selected an open-source content management system, Plone (Plone.org), and the associated Web application server Zope to serve as platforms for our CRF library.

Use of open-source software helped overcome cost and access limitations and enabled us to make the resource available to the broad institutional research community. Customization, however, was necessary and involved the following steps: (1) determine and incorporate ways in which the broader group of users (i.e., researchers without knowledge of the trials conducted by the DCRI) would search for CRFs; (2) build work flow in Plone to enable collection of corresponding information about each CRF, the uploading of CRFs, and the management of uploaded content; (3) identify confidentiality requirements or other agreements that would limit availability of CRFs to the broader Duke research community and manage access appropriately; and (4) migrate CRFs from the existing library to the new library, including indexing each form with desired information. Search strategies from the original system were maintained, and search strategies for the broader community of researchers were identified through discussions with subject matter experts. The remaining steps (steps 2–4) were undertaken to develop and implement the system on which we report here.

Additionally, content migration afforded us the opportunity to reconcile library content with trials conducted by the institution and to acquire CRFs from trials conducted before 2001 that were not in the existing library. We also used the migration opportunity to collect a more robust set of metadata about each CRF (Table 1) and to provide a structure allowing storage of other trial documents, such as protocols, analysis plans, and "lessons learned" that also would be helpful to investigators and research teams designing clinical trials.

CRF indexing and retrieval

The information collected and maintained for each CRF is provided in Table 1. Based on discussions with subject matter experts, these attributes were deemed important to users and to information retrieval. Information is retrieved through an online form-based query, shown in Figure 1; a retrieved form showing metadata is provided in Figure 2.

Forms can be obtained from the library by two methods, one of which is to retrieve the entire CRF for a given trial. When designing a CRF, however, researchers often wish to view multiple instances of groups of like data, or *modules* (e.g., comparing the various methods used to collect vital signs data across multiple trials). Because modules are considered an important design aid, CRFs also were indexed and coded by module. For example, one CRF page may contain both ECG interpretation data and vital signs data, thus two modules. Because standard controlled terminology for modules did not exist, we developed a list empirically by categorizing existing modules. Retrieval by module helps investigators and research teams design new forms by providing examples that show how similar data were collected in other studies, including the specific data elements collected, the level of detail, and the data collection format.

CRFs also were coded using Medical Subject Headings, developed by the National Library of Medicine (MeSH; http://www.nlm.nih.gov/mesh/meshhome.html), i.e., terms for the indication under study in the trial in which the CRF was used. ClinicalTrials.gov standard data elements were used for trial, CRF-level metadata. MeSH was selected because it is easily implemented using a type-ahead strategy in the user interface, because it is freely available, and because it contains definitions for terms. To aid association of additional trial documents with CRFs, the library supports storage of the following document types: CRF, Module, Instrument, Instructions, Annotations, Lessons Learned, Protocol, and Statistical Analysis Plan. Trial-level coding and metadata are inherited by documents associated with each CRF.

Approximately half of the CRFs are from cardiology trials, reflecting the DCRI's long history as a coordinating center for such studies. Because the preponderance of cardiology-

related forms could limit the library's applicability to the broader Duke research community, we created a form submittal page that allows investigators and research teams to contribute CRFs and associated documents for curation and inclusion in the library. In addition, other standards-based forms, such as those used to collect data in a manner consistent with CDISC standards, are also accessible through the library.

System description

The Plone open-source content management system (www.plone.org) allows non-technical users to manage Web site content. To meet user requirements for the CRF library, we created seven custom content types: 1) CRF Library; 2) Trial; 3) Trial Document; 4) Instrument; 5) Module; 6) CRF Search; and 7) CRF Configlet. *CRF Library* and *Trial* are "folder-like" objects capable of holding subsidiary objects. *CRF Library* can contain Trials, Instruments, and additional CRF libraries. *Trial* can contain *Trial Documents* and *Modules*.

All content types except for *CRF Search* and *Configlet* have fields that are used to define the object's metadata. Metadata (also called *attributes*) are indexed and searchable; fields are included as search criteria on the Search form view. Certain Trial metadata (e.g., MeSH terms) are inherited by all objects within the Trial and are thus searchable against the parent Trial's metadata.

CRF Search uses a custom screen that causes a custom search script to run; this script returns any object in the system associated with the search criteria. *CRF Configlet* allows authors to manage the list of values available for certain fields (e.g., list of Therapeutic Areas). All system components use our custom workflow (DCRIWorkflows) except for *Configlet*, which has its own custom workflow (DCRIConfigletWorkflows) that allows only users with the designated role of ConfigletAuthor to manage values lists.

This application uses Plone 2.1.4, Zope 2.8.2, and Python 2.3.5. The CRF Library code and its dependencies (LiveSearchWidget, AutocompleteWidget, MeSHVocab, DCRIConfiglet-Workflow, DCRIWorkflows) are available from the authors. A publicly available demonstration Web site (http://www.dtmi.duke.edu/crflibrary-demo) has been created for readers. The CRF library described here also is listed on the CTSA Resource Discovery System (http://biositemaps.ncbcs.org/cirwp/); the data model for CRF library customizations can be obtained from the authors.

Results

Content

The original CRF library contained 160 CRFs, ranging in size from a few pages to over a thousand pages; in total, the library comprised an estimated 17,000 pages. The current library now contains 177 CRFs (1080 total Plone objects, including associated trial documents), totaling 1.5 gigabytes. Eighty-two of the forms are now available to the broader research community; 95 presently reside within a contractual confidentiality window (usually 5 years from database lock) and are not available to users outside the Institute. As a coordinating center for multicenter clinical trials, the DCRI conducts a small fraction of the total number of trials at our institution. Thus, the number of forms in the CRF library pales in comparison to the total number of studies conducted at our institution as a whole. While we hope the library infrastructure will entice investigators across our institution to share their forms, thereby vastly increasing the number of forms in the library, only three forms have been volunteered to date.

The CRF library was opened to users from the broader Duke research community in September 2007; its existence and availability have been communicated in an institutional

newsletter. The library is also accessible from two institutional research resource Web sites, as noted above. Since its opening to the wider institutional research community, the library has been used on average 74 times per month. An average of 12.7 hours per month is spent on library maintenance, including completion of module coding and loading. If half of all page views are attributable to library maintenance (a conservative estimate), then the library supports an average of 37 investigator uses per month. In the context of an average of 32 clinical research projects started each month at Duke, access data suggest that the library serves a significant proportion of intended research projects. Monthly CRF library use for the four main screens is shown in Figure 3.

Implementation

The most time-consuming and labor-intensive aspect of building the CRF library was content migration (Table 2). Individual documents were indexed manually with the attributes listed in Table 1, coded with MeSH terminology, and loaded into the library. CRFs also were divided by page and indexed to support retrieval by module. The person-hours required for these activities are provided in Table 2. The project started in March of 2007; migration (minus module coding) was completed in August of 2007.

Ongoing maintenance is an important part of the project. Maintenance tasks include tracking clinical trials as they near completion, communicating with the trial team to acquire the CRF and associated documents, indexing documents and uploading forms into the library, as well as curation of material contributed from the broader institution. We estimate that 10% of a full-time equivalent (FTE) employee's efforts will be required on a continuing basis (actual time spent on maintenance has averaged 12.7 hours/month, or about 7.5% of an FTE). We also anticipate that efforts will be required for minor system changes as library and user needs evolve over time. Because these efforts are devoted to maintaining an institutional research resource, they currently are funded by our CTSA. Institutional commitment will be required for long-term support.

Discussion

Establishing knowledge resources can be costly and time-consuming. Knowledge management presents significant challenges to many organizations and can be especially difficult for academic medical centers with large clinical research programs. Recent experience with our electronic CRF library offers encouragement. Based on anecdotal reports from conversations with research fellows and investigators, many researchers have learned about the library's resources through word-of-mouth communication. These informal communications suggest that reception of the library has been positive across the Duke research community and that the library has yielded direct benefits to researchers. In addition, the willingness of one institution to share a resource across internal boundaries bodes well for CTSA efforts aimed encouraging collaboration across traditional divisions within institutions. The effort is potentially sustainable due to the relatively minimal resources required for ongoing maintenance. Such maintenance efforts include a program of active solicitation of new forms as studies close. This is facilitated by monitoring data in institutional clinical trial management systems, i.e., investigators are contacted when studies are marked as closed in the system.

As the CTSAs provide infrastructure for research management, we anticipate that more information about research projects will be collected and maintained at institutions. The CTSA recipients are collaborating with other national efforts to create standards for research classification and management [10]. These standards will promote consistent description of research and facilitate institutional data stores, and will help investigators identify institutional resources and collaborators, facilitate automation of research-related

administrative processes, and promote cross-institutional benchmarking to expedite and improve clinical and translational research in the United States.

As others have suggested, electronic storage and re-use of data elements would increase the efficiency of clinical research [11]. In the absence of software-supported standards to exchange data elements, our future plans are to continue populating the "form/image-based" CRF library until data collection systems can accept electronic data element and screen definitions. We have developed an internal ISO 11179 data element registry, and await such standards for data element transmission and software support. When this occurs, the "form/image-based" CRF library will remain as the archive until the data elements from the CRFs (i.e., knowledge) are curated into the organizational ISO 11179 registry.

Data element registries do not support representational (visual display) aspects (for instance, how prompts and blanks are graphically arranged on a user interface), and current data collection systems are not yet able to consume such information. Thus, modeling and management of such representation information remains an area for further research.

Lessons learned

This undertaking revealed several challenges that deserve discussion. CRFs were originally indexed by therapeutic area using a list specific to organizational divisions (e.g., "cardiovascular device"; "cardiovascular megatrials"). This method did not prove intuitive for users outside the DCRI who searched the library; for instance, users may have wanted to search for all trials of acute coronary syndromes regardless of whether the trial was large or tested a device. Therefore, therapeutic area terminology was converted to a list of standard medical specialties (in addition to the MeSH terms for indication). Any effort aimed at broadening the use of a knowledge resource may encounter similar secondary-use and terminology challenges.

A second challenge arose from the unexpected complexities of providing trial forms and associated documents to users outside of the DCRI. Most industry-sponsored trials contractually require confidentiality regarding trial materials. The DCRI negotiates a confidentiality window in multicenter clinical trial contracts; the window often lasts for 5 years and prohibits disclosure of trial-related documents, with the exception of publication of research results. Not all time windows are for the same duration, however, and the "trial completion" date used to mark the beginning of the confidentiality window is ambiguous. In addition, the DCRI, which holds the confidentiality agreement, is authorized to use the documents within the window. We therefore used Plone workflow and role-based security to allow the curator to make CRFs and trial-related documents either public (viewable within the broader Duke research community) or private (viewable only within the DCRI). We also designated the database lock date as a conservative surrogate for trial completion.

A significant challenge still exists in creating awareness of such institutional resources among investigators and research teams, as well as making the resources easy to locate. For example, e-mailing investigators each time a resource becomes available likely will lead to fatigue and ultimately to communications being ignored, whereas simply making a resource available on a Web site is not sufficient. Additional means of communication, such as discussion at department and staff meetings, presentations, flyers, posters, and announcements on electronic bulletin boards likely will be necessary in the early years of the CTSAs or until "one-stop shops" for research resources become available within institutions. We have not yet fully solved this communication conundrum and anticipate that until a "critical mass" of attractive research resources is reached investigators will be slow to seek resources on institutional CTSA Web sites.

Conclusions

Creation of CRF libraries from historical information requires manual identification and indexing of each piece of information according to a recognized structure such as ClinicalTrials.gov data elements or MeSH controlled terminology. For our institution, the number of visits to the CRF library justifies the effort associated with reclaiming, preserving, and sharing the information about data collected on completed trials, and the ongoing maintenance effort. The middle-of-the-road approach described here, i.e., indexing and coding only a small amount of information and doing so at the module level rather than at the level of individual data elements, allowed us to keep the cost of this effort low and still to provide value to investigators. However, when standards and systems to support re-use of individual data elements become a reality, additional indexing work will be required if historical content is to be migrated forward. In addition, future collection of such information about research studies can be facilitated by indexing this information in required institutional systems at the time of study design and startup. Thus, our approach represents an interim step on the road to rendering our institutional knowledge about data elements collected on historical trials broadly reusable. Libraries such as the one we describe here have proven useful to investigators and research teams, and are a small but necessary step toward the development of a mature system of institutional knowledge management for clinical and translational research.

Acknowledgments

References

1. Spilker, B.; Schoenfelder, J. Data Collection Forms in Clinical Trials. New York: Raven Press; 1991.

2. Society for Clinical Data Management Web site. [accessed on May 10, 2008] Good Clinical Data Management Practices Document (version 4.0) 2007. Available at: www.scdm.org

3. Young, S. Implementation of global clinical data standards at Eli Lilly and Company. Presented at Society for Clinical Data Management Fall Conference; 2004; Toronto, Canada. Available at www.scdm.org

4. Clinical Data Interchange Standards Consortium Web site. [accessed May 11, 2008] CDISC Press Release, CDISC Collaborative Standards Forum (CSF). May 9. 2002 Available at: http://www.cdisc.org/news/

5. Ruberg, S. No one is as smart as everyone: CDISC and the Collaborative Standards Forum. Presented at the DIA eBusiness Conference; October 22, 2002; Available at http://www.cdisc.org/publications/past_presentations.html

6. Medical University of South Carolina Web site. [accessed May 11, 2008] Data Coordination Unit Case Report Form Library. Available at https://dcu.musc.edu/tools/crf_lib.asp

7. OpenClinica Enterprise Web site. [accessed May 11, 2008] OpenClinica press release. Hundreds of standardized CRFs now available in the OpenClinica Enterprise Case Report Form Library. Jan 23. 2008 Available at: http://www.openclinica.org/crflibrary/

8. Wickens, CD.; Hollands, JG. Engineering Psychology and Human Performance. 3. Upper Saddle River, NJ: Prentice Hall, Inc; 1999.

9. [accessed September 12, 2008] Clinical and Translational Science Awards Web site. Available at: www.ctsaweb.org

10. Carini S, Pollock BH, Lehmann HP, et al. Development and evaluation of a study design typology for human research. AMIA Ann Symp Proc. Nov 14.2009 :81–85.

11. Singer SW, Meinert CL. Format-independent data collection forms. Controlled Clin Trials. 1995; 16:363–376. [PubMed: 8720015]
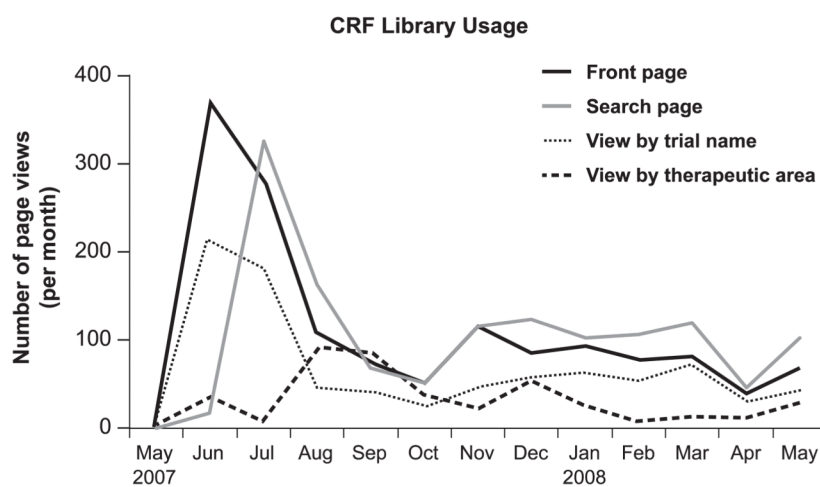
Figure 1.
Form-based Query Supports Information Retrieval from the CRF Library.
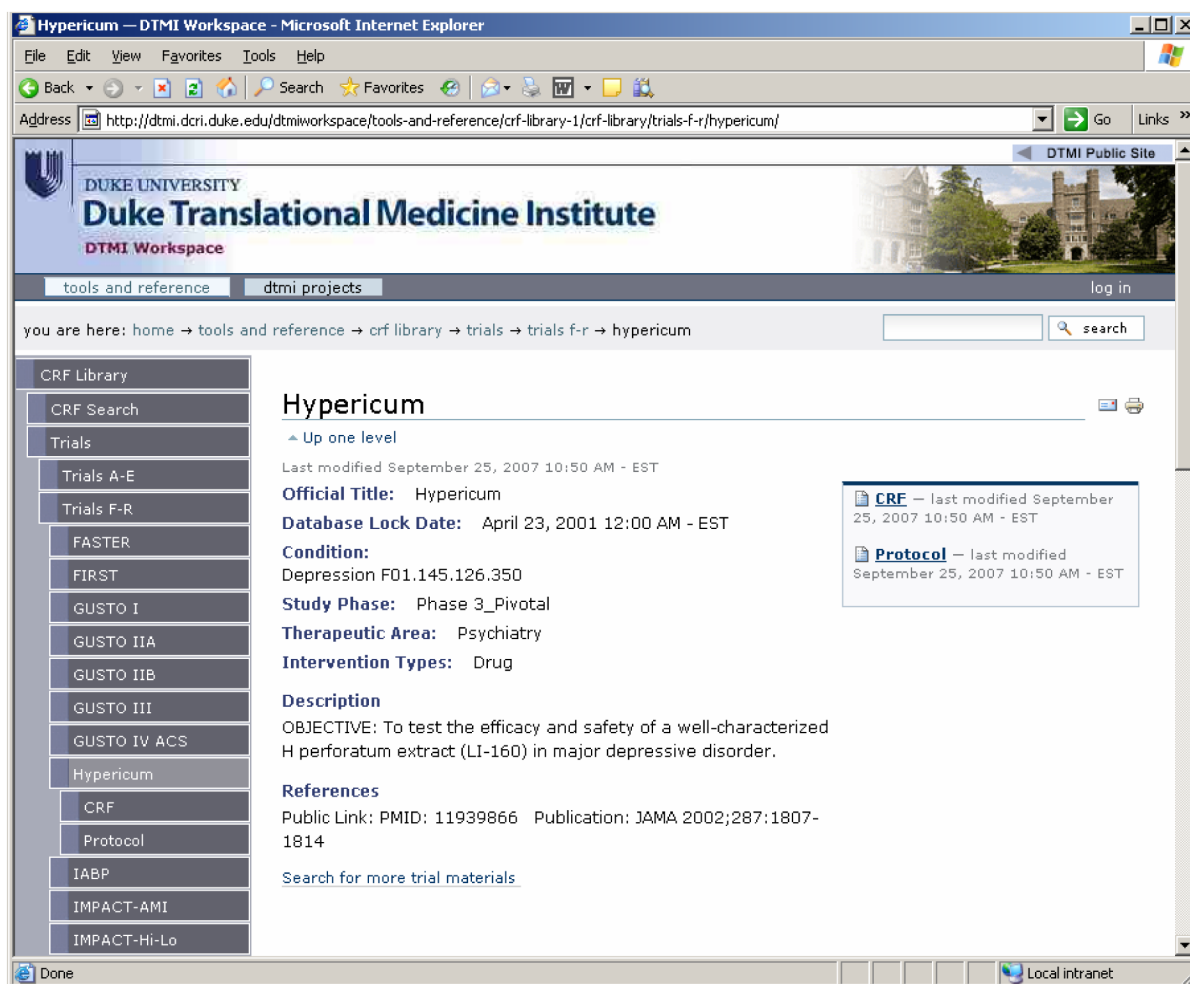
Figure 2.
Information Retrieved from the Library.

Figure 3.
CRF Library Usage, May 2007 through May 2008.

Table 1

CRF Library Metadata Categories

Trial name
Duke project ID
Study phase
Sponsor
Therapeutic area
Intervention type
Condition under study
Description of trial
Database lock date
Reference link to primary manuscript

Table 2

Personnel resources required for CRF library project

Task	Hours	Number of individuals
Project management	67	1
Information design	40	1
Analysis and IT project management	45	1
Development and testing	465	3
Communications and editorial activities	121	2
Content migration	763	5
Total	**1501**	**10**