2013

'What's in the NIDDK CDR?'—public query tools for the NIDDK central data repository

Huaqin Pan
RTI International

Mary-Anne Ardini
RTI International

Vesselina Bakalov
RTI International

Michael DeLatte
RTI International

Paul Eggers
National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)

See next page for additional authors

DATABASE
The Journal of Biological Databases and Curation

Database, Vol. 2013, Article ID bas058, doi:10.1093/database/bas058

Original Article

'What's in the NIDDK CDR?'—public query tools for the NIDDK central data repository

Huaqin Pan[1,†], Mary-Anne Ardini[1,†], Vesselina Bakalov[1], Michael DeLatte[1], Paul Eggers[2], Laxminarayana Ganapathi[1], Craig R. Hollingsworth[1], Joshua Levy[1], Sheping Li[1], Joseph Pratt[1], Norma Pugh[1], Ying Qin[1], Rebekah Rasooly[2], Helen Ray[1], Jean E. Richardson[1], Amanda Flynn Riley[1], Susan M. Rogers[1], Sylvia Tan[1], Charles F. Turner[1,3], Stacie White[1] and Philip C. Cooley[1,*]

[1]RTI International, Social, Statistical and Environmental Sciences, PO Box 12194, Research Triangle Park, NC 27709, [2]National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), Division Of Kidney, Urologic, & Hematologic Diseases, Bethesda, MD 29892 and [3]City University of New York (Queens College and the Graduate Center), Flushing, NY 11367, USA

***Corresponding author:** Tel: +1 919 541 6509; Fax: +1 919 316 3539; Email: pcc@rtii.org

†These authors contributed equally to this work.

Submitted 20 September 2012; Revised 5 November 2012; Accepted 28 November 2012

Citation details: Pan Huaqin, Ardini Mary-Anne, Bakalov Vesselina, *et al.* 'What's in the NIDDK CDR?'—public query tools for the NIDDK central data repository. *Database* (2012) Vol. 2012: article ID bas058; doi:10.1093/database/bas058.

The National Institute of Diabetes and Digestive Disease (NIDDK) Central Data Repository (CDR) is a web-enabled resource available to researchers and the general public. The CDR warehouses clinical data and study documentation from NIDDK funded research, including such landmark studies as The Diabetes Control and Complications Trial (DCCT, 1983–93) and the Epidemiology of Diabetes Interventions and Complications (EDIC, 1994–present) follow-up study which has been ongoing for more than 20 years. The CDR also houses data from over 7 million biospecimens representing 2 million subjects. To help users explore the vast amount of data stored in the NIDDK CDR, we developed a suite of search mechanisms called the public query tools (PQTs). Five individual tools are available to search data from multiple perspectives: study search, basic search, ontology search, variable summary and sample by condition. PQT enables users to search for information across studies. Users can search for data such as number of subjects, types of biospecimens and disease outcome variables without prior knowledge of the individual studies. This suite of tools will increase the use and maximize the value of the NIDDK data and biospecimen repositories as important resources for the research community.

Database URL: https://www.niddkrepository.org/niddk/home.do

Introduction

Data repositories have become prevalent in the past decade, storing vast banks of specialized data that support a myriad of research fields. When genetics became a key component of all biological research, the number of disease- and health-related databases increased rapidly. Data from clinically annotated biospecimens provide researchers with a precious resource for additional genetic studies without the cost of collecting and testing. The databases also provide the added benefit of increased statistical power by facilitating pooling of data from multiple study populations.

The National Institute of Diabetes and Digestive Disease (NIDDK) Central Data Repository (CDR) is one such data repository. The NIDDK CDR is a web-enabled resource that catalogs clinical research data and supporting materials from NIDDK funded studies (1, 2). It also acts as a warehouse for clinical findings and biospecimen and genetic data from completed research projects. Other comparable data repositories include the Database of Genotypes and Phenotypes (dbGaP) developed at the

Database, Vol. 2013, Article ID bas058, doi:10.1093/database/bas058

National Institutes of Health (NIH) National Center for Biotechnology Information (NCBI) (3), the Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) (https://biolincc.nhlbi.nih.gov/studies/) developed at the National Heart, Lung and Blood Institute (NHLBI), the cancer Human Biobank (caHUB) (4) developed at the National Cancer Institute (NCI) Office of Biorepositories and Biospecimen Research (OBBR) and the Promoting Harmonisation of Epidemiological Biobanks in Europe (PHOEBE) (http://www.phoebe-eu.org). Despite the significance of these NIH agency and European data and biospecimen repositories, their broad-based impact may be limited by the small number of publications available which describe the repositories.

Currently, the NIDDK CDR has data from 40 studies available for request. An additional 54 ongoing studies are being tracked or curated. The CDR contains data for 40 sample types, over 6 million biospecimens representing 2 million subjects from 94 studies/networks. Genotype data are available in the CRD for some studies. Genomic data from the NIDDK Genome-Wide Association Studies (GWAS) are deposited in dbGaP and the NIDDK CDR has links to the studies in dbGaP.

To help users explore the vast amount of data and biospecimens stored in the NIDDK CDR, we developed a suite of search mechanisms called the public query tools (PQTs). In this article, we describe the design and development of PQT and how it assists users to find study information, data and biospecimens in the CDR. Funded by the NIH American Recovery and Reinvestment Act (ARRA) of 2009, the PQT suite gives users new capabilities to perform cross-study searches without prior knowledge of each individual study.

Design

Researchers can visit the public website (https://www.niddkrepository.org/niddk/home.do) to learn about the CDR contents and identify studies of interest. Website users have access to study protocols, sample information, study metadata, data collection forms and study publications. However, to review actual data values, researchers have had to first submit a formal request and gain NIDDK approval for a proposed investigation. This proposal-specific approach has been time consuming and limited in utility because researchers often need to see data before finalizing a proposed project. Consequently, interested researchers would generally contact CDR staff with questions about the data before going through a formal request process.

To provide quicker access to data for CDR users, we developed the PQT suite to address the question 'what's in the NIDDK CDR?' PQT helps users explore CDR contents with a user friendly interface. The PQT caters to marked differences in user scope by producing results with a range of granularity, from general to very specific. For instance, a user may seek available subjects or samples from studies that have African-American subjects with diabetes who are 50 years of age or older. PQT is able to provide the result quickly and easily.

Although there may be many ways to query the contents of the CDR, for PQT, we chose user perspective as our design starting point. Through our design assessment, we defined the following three user perspectives.

- Design focused user is interested in identifying studies by traits such as purpose, principal findings and main design elements. This user is looking for studies with specific methods and treatment that may provide insight for the design of new studies. Variables needed are study design, treatment (if the design is a clinical trial), duration, principal findings and study outcomes.
- Disease focused user seeks data from a variety of clinical studies to develop a unified view across study subjects. This user searches for commonality of health factors from different but related studies and is interested in viewing studies that follow different protocols about the same disease type. For this perspective, variables from multiple studies are compared for the underlying and undiscovered properties related to a disease and its treatment. These variables include those that examine the extent of the disease (e.g. serum creatinine levels) and confounders (e.g. blood pressure and age) from multiple studies within a disease domain.
- Casual user is interested in identifying the various types and breadth of studies in the CDR that are available to researchers. Users with this perspective are looking for broad level attributes and study descriptions.

After clarifying the perspectives, we identified the kinds of data elements that were needed to yield the desired result. This process revealed that certain data elements were common to all perspectives, while others were only needed for a more in-depth searching of users focusing on disease. Our assessment of data elements resulted in specific groupings of data for each study in the CDR. The assessment process also clarified functionality and user interface requirements and led us to develop the distinct PQT search tools that are now available for public use at the CDR website (Figure 1). The tools include the study search tool, basic search tool, ontology search tool, variable summary tool and samples by condition search tool. Each tool is discussed later in 'PQT suite' section.

Develop data tables

Because actual study data files are only accessible to requestors approved by the NIDDK CDR Reviewer Panel, we must specially prepare the variables that power the search tools.

Figure 1. PQT suite of tools at NIDDK CDR website.

Identifying and importing these data into structured data tables that link to the tools have been a labor intensive process performed by project staff who work closely with the study data and understand the priority of data elements to research objective. However, the benefit is clear—once in place, these data are able to simply and easily provide a user with the response to a query such as, 'what studies have data on retinopathy?'

Identifying and compiling PQT variables

The team developed three categories of data to drive the PQT: study metadata, disease domain variables and common standardized variables.

Study metadata. Specific characteristics are identified to provide a snapshot of each study in the CDR so that users can easily review the most salient aspects at a glance. The Repository staff review study materials when adding a study to the CDR to extract these characteristics, which are used as data elements by the PQT. A complete list of metadata elements is shown in Table 1. The study metadata is most closely aligned with the basic search tool within the PQT suite but is also accessed by the ontology search and study search tools.

Disease domain variables. The ability to compare data across studies is important for researchers whose work focuses on a specific disease. For PQT, we constructed categories called 'disease domains' into which repository studies fall. The disease domain tables are populated with variables whose 'content' is drawn from study datasets, with up to 40 of the most important variables from each study extracted for use in PQT. The disease domains allow a user to compare same or similar variables across studies. PQT employs concept harmonization, not derivation, which means that specific study variables are mapped to disease domain

Database, Vol. 2013, Article ID bas058, doi:10.1093/database/bas058

Table 1. Metadata sources for basic and ontology search tools used for PQT suite

Metadata sources for basic and ontology search tools	
No. study subjects	Image data availability
Condition	Transplant patient
Study design	Dialysis patient
Treatment/intervention	Duration of study
Main outcome measures	Objective
Study outcome	Selection criteria
Stored samples	# Recruitment sites
Non-GWAS data availability	dbGaP GWAS data link

Table 2. Alignment of repository studies[a] to the established PQT disease domains

Disease domain	Studies in the CDR
1. Diabetes	DCCT/EDIC, DPP, DPPOS, DPT1, FIND, GoKinD, HEALTHY, T1DGC andTEDDY
2. Kidney disease	AASK Trial, AASK Cohort, ATN, CDS,CRIC, CRISP, DCCT/EDIC, DAC-FISTULA, DAC-GRAFT, FIND, GoKinD, HEMO, LookAHEAD, MDRD, NANS and PRIDE
3. Liver disease	A2ALL, HALT-C, LTD, LTD2, PEDS-C and VIRAHEP-C

[a]See Table A1 for the list of full study names.

variables while still preserving the variable's study-specific definition. The variable description and name are harmonized, whereas units, statistical formulas used to calculate composite variables, data groupings and distribution (continuous versus categorical) are not. This limited harmonization maintains unique study definitions of variables while providing the user with a method for linking similar measures across studies.

For example, the domain variable 'hxneuropathy', or history of neuropathy, is a diabetes domain element. The GoKinD and EDIC studies both collected this data. The GoKinD measure is a history of severe symptomatic peripheral neuropathy collected via questionnaire, as 1 = yes, 0 = no and 3 = unknown. The EDIC measure is a clinical diagnosis of neuropathy at the close of the parent study, DCCT, where 1 = definite, 2 = possible and 3 = none. Both GoKinD and EDIC study-specific measures were mapped to the domain variable 'hxneuropathy'; however, the study-specific definitions of each were preserved and may be referenced using the variable summary tool, described later in the 'PQT suite'.

To date, variables pertaining to three disease domains are included in PQT: diabetes, kidney disease and liver disease. Variables relevant to these domains were identified through a literature search and by consulting with disease experts. Because some of the same elements are associated with more than one disease domain, there is overlap between variables of different domains. Table 2 shows how studies fall into multiple disease domains. In the future, we intend to increase the number of disease domains, adding a domain when more than two relevant studies have been submitted to the repository.

Disease domain variables can be grouped into the following categories:

- Adverse events.
- Co-morbidity.
- Demographics.
- Covariates related to outcomes.
- History of disease.
- Outcomes.
- Physical and lab measurements.

Currently, the three disease domains in Table 2 incorporate 119 distinct variables, with 10 demographics variables belonging to all three domains. Please note that some of the demographics variables overlap with other common standardized variables as described in the next section.

Common standardized variables. To identify a common set of broad-based variables across all clinical and genetic studies in the repository, we adopted the NCI Common Biorepository Model (CBM). Although the CBM is designed to describe biospecimens, the variables also supply characteristics of a research subject. By using the CBM standard variables, we were able to describe the subject as well as the samples associated with the subject and, most importantly, illustrate NIDDK's support for a standard method of annotating samples across all research projects.

PQT uses 15 of the 30 established CBM variables. Most are populated by a yes/no field (i.e. those starting with 'Has' in Table 3); however, some contain values abstracted from study data. Some CBM variables can be directly mapped to a study variable. For example, the race of a patient may have been recorded during data collection and is pulled into the variable CBM 'Race'. Others are derived from a review of study data. For example, the 'hasLabdata' variable will be derived from the study dataset called 'Labs', which has information on several measurements on the patient such as cholesterol, creatinine and glucose. The presence of information for any one of the variables in the Labs dataset will imply that: hasLabdata = yes. Table 3 shows the CBM variables used within PQT.

Align data to PQT

Tables containing the required data are integrated into the CDR database to support the search tools, as illustrated in

Figure 2, which maps the data table to the tool. The 12 tables are grouped as follows: study, ontology nodes, disease domains, variables and samples. For PQT to operate, the data are uploaded to these tables within the central database.

Table 3. List of CBM variables used by the PQT suite

	CBM variable
1	Name/acronym of study
2	Diagnosis
3	Sex
4	Race
5	Ethnicity
6	Has additional patient demographics data
7	Has family history data
8	Has histopathologic data
9	Has exposure history data
10	Has lab data
11	Has treatment information
12	Has outcome information
13	Has longitudinal specimens
14	Has matched specimens
15	Year sample first preserved/stored

The data are systematically compiled and passed to the development team, who then initiate the data loading program to make the data available to search. The program parses, analyzes and validates the data against defined database/business rules. After validation, the data are loaded into the core PQT tables shown above and are available to users.

PQT suite

In this section, we describe the final set of query tools that resulted from the design process. Table 4 lists each of the five tools with their distinctive search methods and results.

Study search tool

The study search tool provides key information about each study including metadata and study materials. Users select a study by name in the study search interface. This action brings up a table showing the availability of samples, genotype data and a link to the selected study. Users can follow the link to find a metadata table with key information about the study such as disease condition, study design, objective, selection criteria, outcomes and other items. This page also includes a link to the study page that includes a general description, the study protocol, manual of operations, data forms and a roadmap report describing

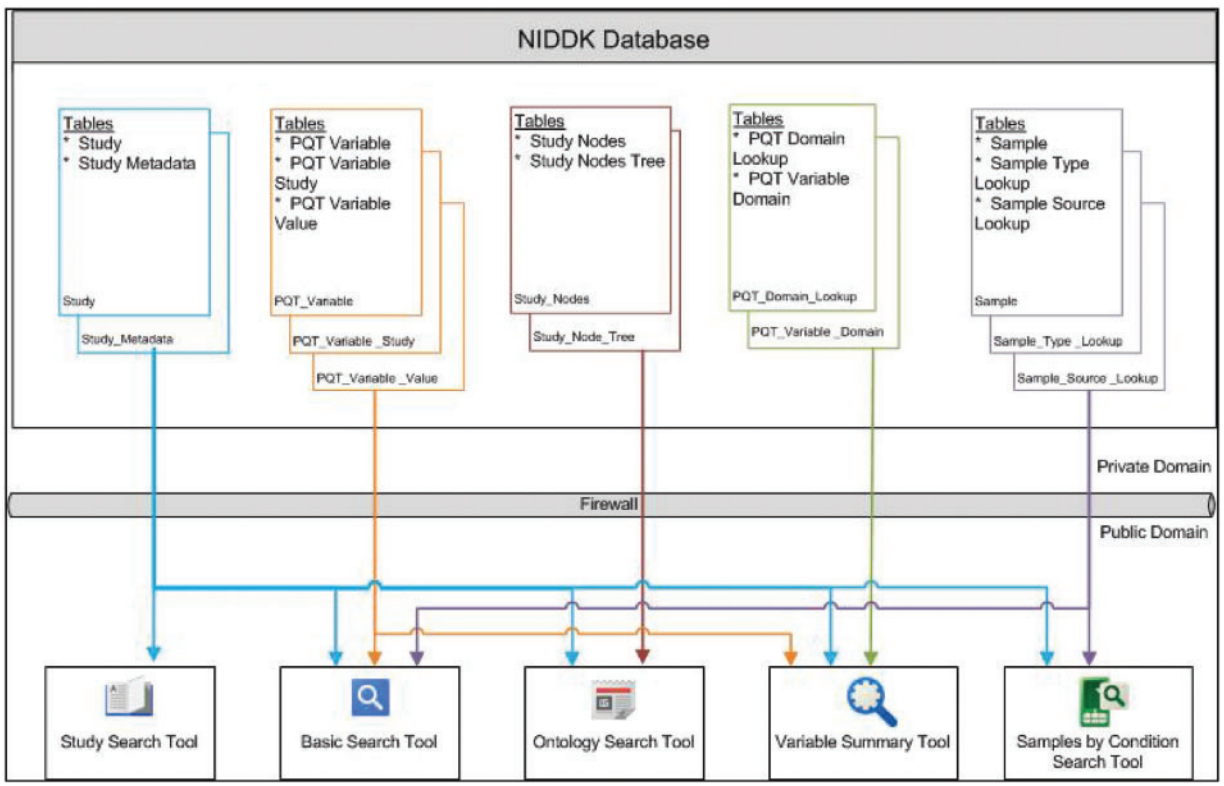

Figure 2. CDR PQT suite tables supporting the query tools.

Database, Vol. 2013, Article ID bas058, doi:10.1093/database/bas058

the data archive. It also identifies a primary publication and displays a replication report, the data dictionary and study samples.

Basic search tool

The basic search tool identifies studies based on user-defined criteria. The user selects from dropdown lists that are a subset of study metadata categories:

- Condition studied—benign prostatic hyperplasia (includes enlarged prostate and prostatitis), chronic pelvic pain, cirrhosis, coronary heart disease (includes cardiovascular disease), Crohn's disease, diabetes Type 1, diabetes Type 2, hepatitis C, hypertension, hypogonadism, impaired glucose tolerance, inflammatory bowel disease, interstitial cystitis, kidney disease, liver disease, neuropathy, retinopathy, sexual dysfunction and urinary incontinence (includes incontinence and stress urinary incontinence and urogynecologic).
- Study design—case-control, clinical trial, cross-sectional, genetic, longitudinal, observational, prospective, retrospective and survey.
- Intervention/treatment—behavioral, dialysis, drug therapy, liver/kidney, transplant, nutrition and surgery.

Table 4. PQT search methods and results

PQT tools	Search methods	Search results
Study search tool	Select predefined study name.	Study name linked to study metadata and documents.
Basic search tool	Select predefined study metadata categories.	Study name linked to study metadata and documents.
Ontology search tool	Enter keyword to search study metadata.	Study name linked to study metadata and documents.
Variable summary tool	Select from predefined variables.	Variable counts/frequency in studies.
Sample by condition search tool	Select predefined disease conditions.	Biospecimen types and counts by study.

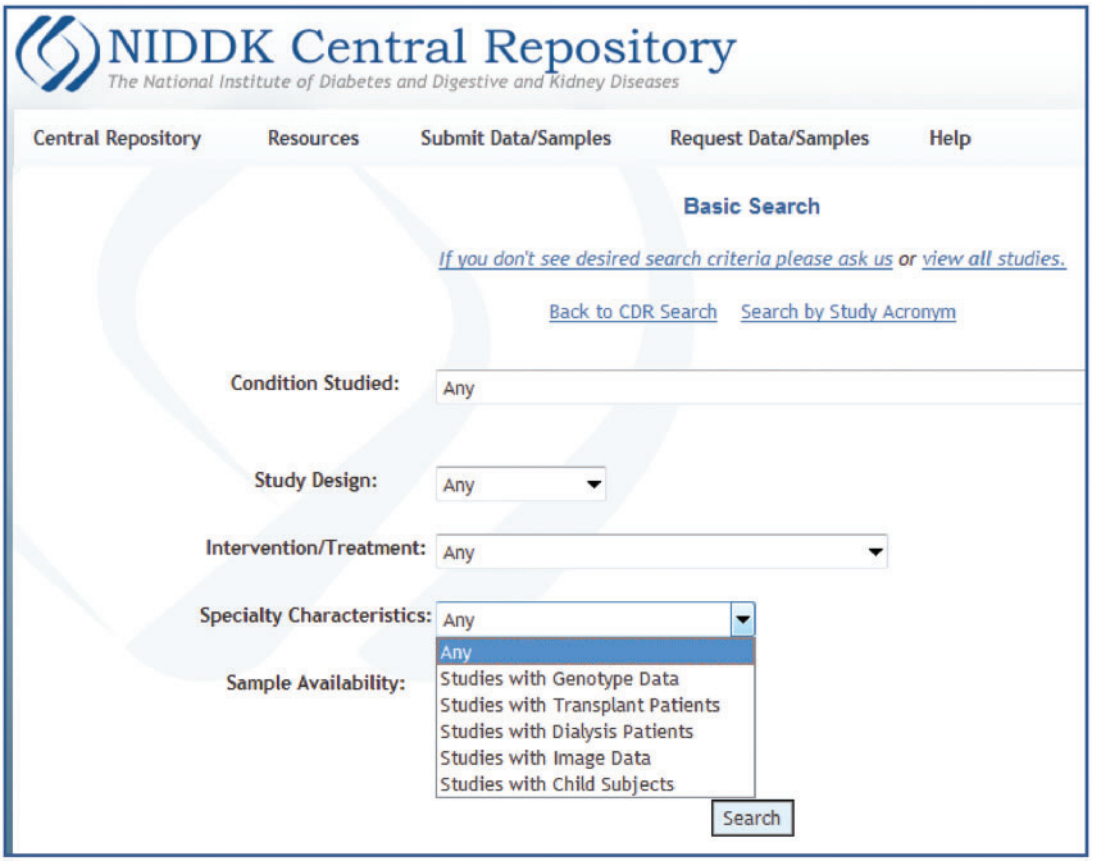

Figure 3. Screenshot showing the predefined categories of 'Specialty Characteristics'.

In addition, the basic search offers predefined queries with the following specialty characteristics:

- Studies with genotype data (GWAS and non-GWAS).
- Studies with subjects who are dialysis patients.
- Studies with subjects who have had a liver or kidney transplant.
- Studies with associated image data (CT, MRI and ultrasound).
- Studies with subjects under 18 years of age.

The tool also allows identifying studies by the types of samples stored in the central bio-repositories:

- Sample availability—DNA, buffy coat, PBMC, plasma, serum, tissue and others.

Researchers set criteria on the main basic search page (Figure 3). Result displays studies matching all criteria in a table that includes stored sample types and availability of genotype data. Study names shown in the table are hyper-linked to a study metadata and another link to all stored study materials, including the protocol and data forms.

Ontology search tool

The ontology search tool identifies studies via keyword search through CBM variables that have been mapped to the NCI Metathesaurus. The ontology supporting this tool was constructed to deliver maximum search results for user specifications. For example, searching 'diabetes mellitus' will find studies that reflect: 'diabetes', 'Type 1 diabetes', 'diabetes, Type 1', 'diabetes Type 2' and 'Type 1 diabetes'. This search strategy increases the sensitivity of returned results with a minimal specificity trade-off.

In addition to the NCI-Metathesaurus-based ontology search, the tool also runs the search in full-text through select study metadata items (outcome measures, condition, outcome, objective, design and criteria). The tool combines results from both searches and presents it to the user. The algorithms aim to provide both high specificity and high sensitivity. The ontology search tool also provides the following two convenient features to the user:

- Search within the search result allows users to narrowq or filter the search results using additional search terms.
- View search path allows users to view the hierarchy of the search path, showing the relationship between the keyword searched and the term found in the study metadata.

Variable summary tool

The variable summary tool allows users to compute summary statistical reports for key variables relating to study

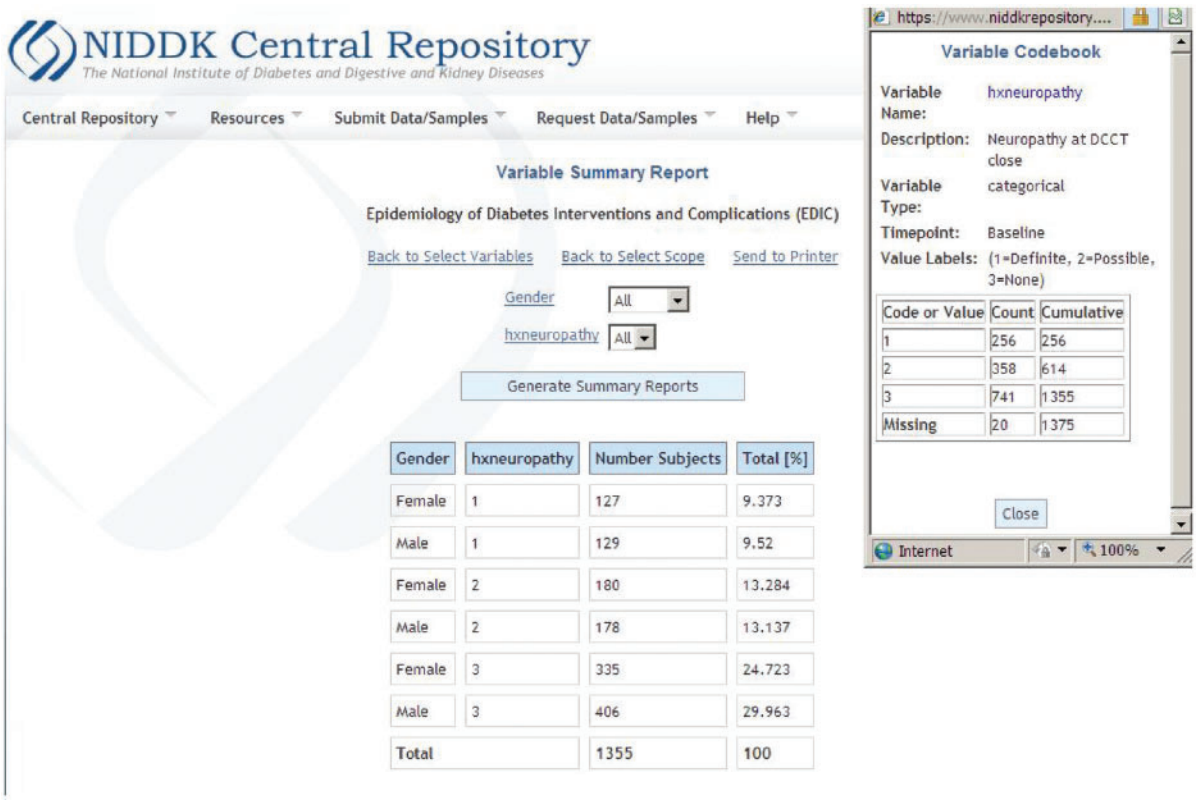

Figure 4. Screenshot showing summary statistics of number of subjects for two variables (gender and history of neuropathy) in the EDIC Study in the single study view.

Database, Vol. 2013, Article ID bas058, doi:10.1093/database/bas058

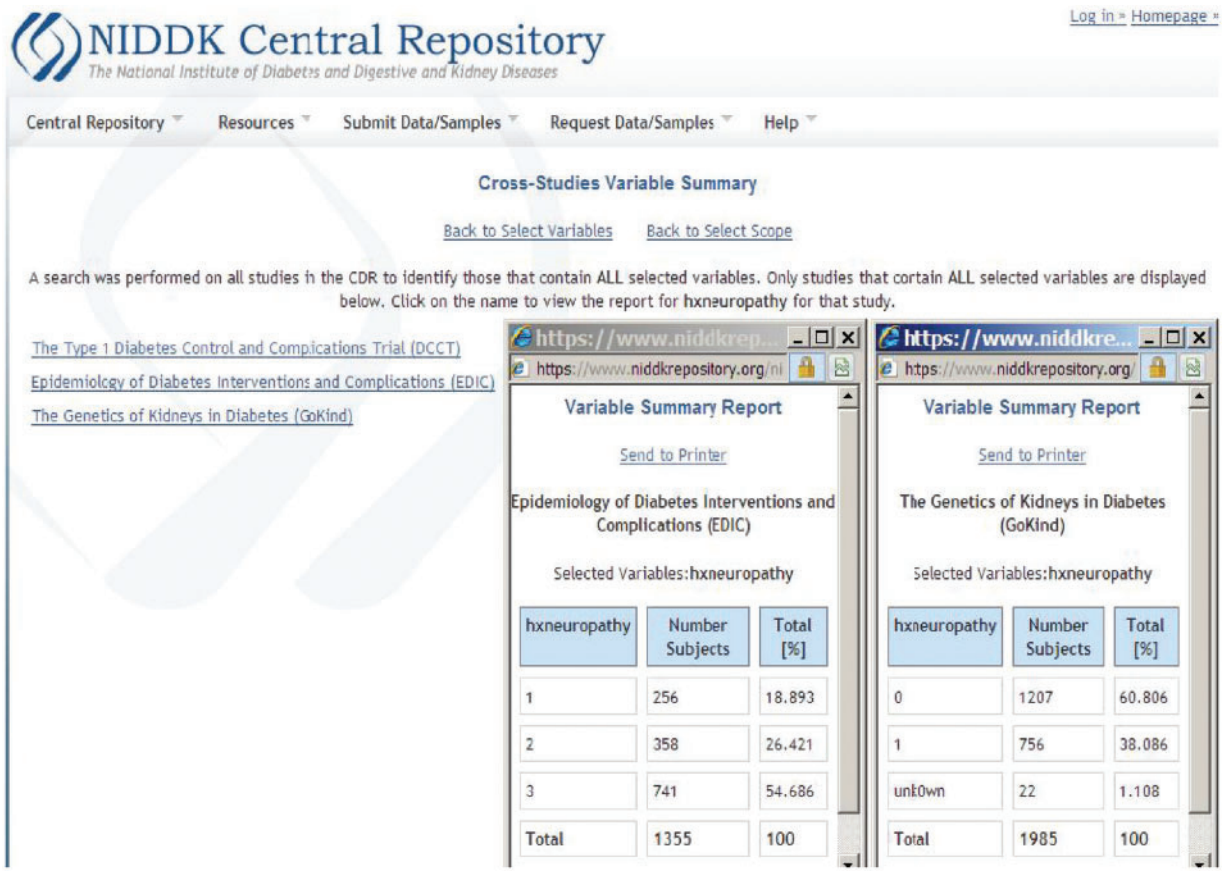

Figure 5. Screenshot showing summary statistics of number of subjects for one variable (history of neuropathy) in two studies (EDIC and GoKinD) in the cross-study view.

subjects. The project team cataloged a core set of variables from each study to support queries. Users select from these core sets and the tool identifies studies that contain the variables. Users can select up to three variables per query to generate summary reports for variable value range and subject counts. The tool generates reports to show continuous and categorical variables for selected demographic and clinical characteristics in archived studies.

This tool has two views.

For the 'single study view', users first select a study from a dropdown list. Then, users can identify up to three measures from that study. Results are summarized as ranges (for continuous variables) or as frequencies (for categorical variables). If more than one variable is chosen, the report will show a cross-tabulation, e.g. if neuropathy diagnosis and gender are chosen, then frequencies of neuropathy diagnosis are shown separately for males and for females. See Figure 4 for an example of this use-case.

For the 'cross-study view', again users select up to three variables of interest. The tool produces a list of studies that have all selected variables. Users can then select the studies for which they seek descriptive statistics. Data from each selected study pops up in its own window. This presentation makes it easier to compare and contrast similar variables (such as age range) across multiple studies. See Figure 5 for an example of this use-case.

As a reminder, the actual study data item(s) used to support the presence of a domain variable may vary across studies. Therefore, in Figure 5, the presence of neuropathy in EDIC includes values of '1' (definite) and '2' (possible) but in GoKinD, the only value is '1' (yes). The tool helps identify same or similar variables across studies to support users' research, not to perform data harmonization within PQT. If needed, a user can find further detail on study definition of the variables in the study documentation, which is easily accessible through a link from the summary tool.

Sample search tool

The sample search tool allows users to select a disease condition from a dropdown list, and the tool presents studies with matching biospecimens. The search result provides study name, sample types, sample counts and subject counts for samples stored in NIDDK Biorepositories. As illustrated in Figure 6, the result provides both sample and subject counts for selected sample types by study.

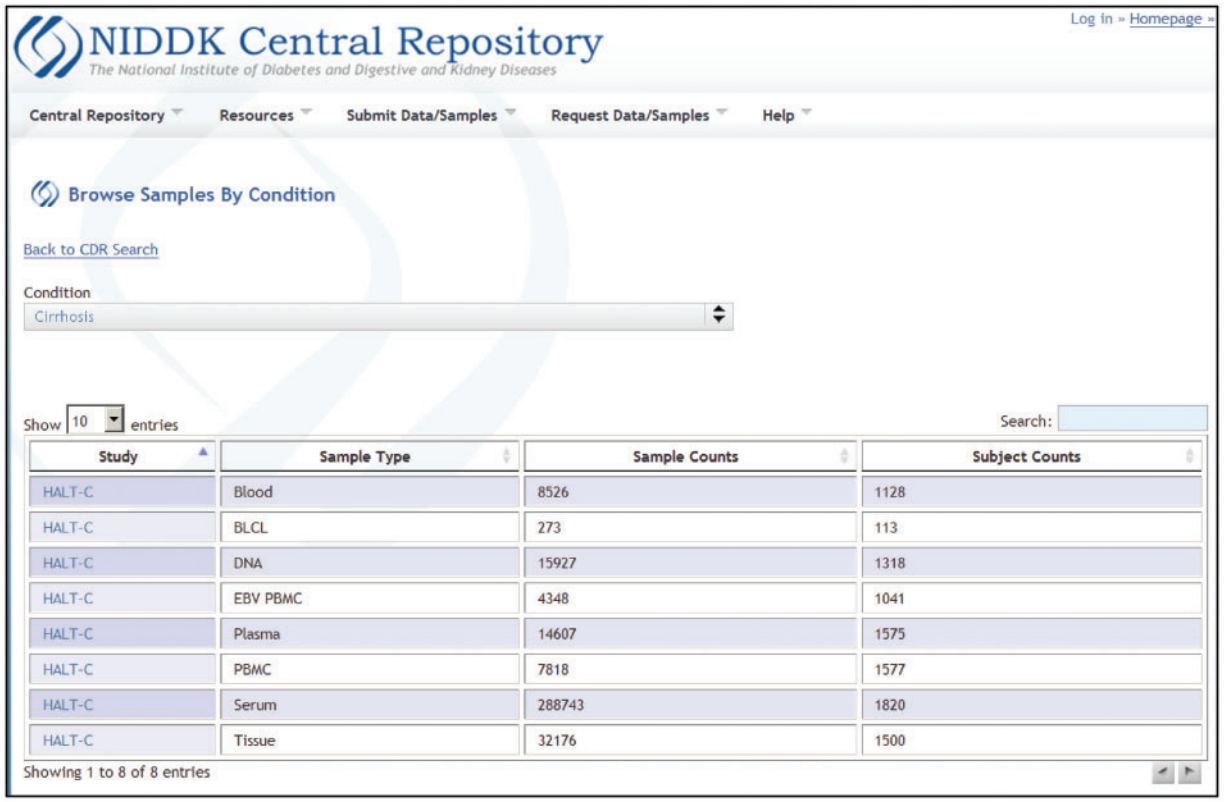

Figure 6. Search result of sample types for condition 'Cirrhosis'.

Challenges, benefits and future enhancements

The PQT suite has been very successful in accomplishing the intended goals. Users can search enormous amounts of data and quickly identify answers to questions. However, the amount of effort to develop and maintain the tools is not insignificant. Here, we discuss our challenges, the benefits to the repository derived from PQT and possible future enhancements.

Maintenance challenges

Tools like the PQT rely on accurate and timely alignment to data in order to provide search results that answer a user's question. Data incorporated into the database through routine registry tasks are not a maintenance concern because the information is updated routinely. This includes the biological sample information that powers the study search, basic search, sample search and ontology search tools.

However, the data elements that feed the variable summary tool are extracted manually from study data and uploaded to the database, a process now performed by data analysts and involving multiple steps before the data are available for search. We will continue to explore more efficient methods for performing this process because these tables drive one of the most unique tools in the PQT suite. A possible alternative we are exploring is a text-based search of variable names in study data files to identify possible matches to the harmonized variable name. Once potential matches are identified, the analyst can more easily review and select within a user interface that copies the selected item into the database. This process will require intensive testing for acceptability but could be a time saving measure that would make maintenance faster and data available to search soon after a study is added to the repository.

Performance and security

The PQT is accessed through the NIDDK repository website. Data to support a query are stored in the central database that resides behind the RTI's secure firewall. When a user submits a query, the data are parsed from the database to produce the result on the webpage. Generally, results are displayed in real time with no noticeable delay for the user. However, if a large amount of data are required for a cross tab search, the results require more time. The project team is investigating speed time improvements to reduce the delay for displaying these types of results.

Quality control support

PQT has become a valuable tool for casual users and professional researchers and has provided unexpected additional support to the CDR for monitoring registry progress and conducting quality control spot checks on the registry database. We now routinely run queries to check study details such as sample tallies for material types and subject totals. We also conduct queries to gain a better understanding of the topic concentration of the research being funded by NIDDK. These trend data are interesting to researchers and to NIDDK because they can easily compare actual trends against their own strategic plans.

Expansion of variables available to PQT

As discussed in the context of maintenance challenge, the functionality of a query tool that is driven by actual study data depends wholly on the variables included in the data tables. While there are some follow-up measures such as adverse events available through PQT, the tools primarily access subject baseline data. Because many studies in the repository are longitudinal in nature, an important future development would be preparation and inclusion of additional time points as well as specific variables. Furthermore, we aim to expand the number of disease domains so that study data from a wider variety of diseases is available for review without submission of an official request.

Conclusion

The NIDDK Central Repository was established to increase the impact and extend the utility of valuable data and samples by making these materials available to the broader scientific community. The new suite of PQT provides effective and flexible methods to search repository study data and samples, allowing researchers to find content of interest without prior knowledge of the studies. PQT opens up the repository to a much larger user group to stimulate research ideas and greater use of NIDDK data and biospecimens which are an important resource to the disease research community.

Funding

National Institute of Diabetes and Digestive and Kidney Diseases; National Institutes of Health, NIH American Recovery and Reinvestment Act (ARRA) of 2009, Department of Health and Human Services, under Contracts (HHSN: 267200800015C, 267200800016C and 267200800018C). Funding for open access charge: HHSN267200800016C.

Conflict of interest. None declared.

References

1. Cuticchia,A.J., Cooley,P.C., Hall,R.D. *et al*. (2006) NIDDK data repository: a central collection of clinical trial data. BMC Med. Inform. Decis Mak., **6**, 19.
2. Turner,C.F., Pan,H., Silk,G.W. *et al*. (2011) The NIDDK Central Repository at 8 years—ambition, revision, use and impact. *Database*, **2011**, article ID bar043; doi:10.1093/database/bar043.
3. Mailman,M.D., Feolo,M., Jin,Y. *et al*. (2007) The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.*, **39**, 1181–1186.
4. BioLINCC in ISBER News: Shea,K. and Wagner,E. (2010) BioLINCC: facilitating web access to NHLBI data and biospecimen repositories. **10**, 5–7. http://www.isber.org/newsletters/documents/July2010.pdf, accessed January 14, 2013.
5. Massett,H.A., Atkinson,N.L., Weber,D. *et al*. (2011) Assessing the need for a standardized cancer HUman Biobank (caHUB): findings from a national survey with cancer researchers. *J. Natl Cancer Inst. Monogr.*, 2011, 8–15.

Appendix

Table A1. Study acronyms and full names

A2ALL	The Adult-to-Adult Living Donor Liver Transplantation Cohort Study
AASK Trial	The African American Study of Kidney Disease and Hypertension Study
AASK Cohort	The African American Study of Kidney Disease and Hypertension Cohort Study
ATN	The Acute Renal Failure Trial Network Study
CDS	The Comprehensive Dialysis Study
CRIC	Chronic Renal Insufficiency Cohort Study
CRISP	The Consortium for Radiological Imaging Studies of Polycystic Kidney Disease
DAC—Fistula	The Clopidogrel Prevention of Early AV Fistula Thrombosis Study
DAC—Graft	The Aggrenox Prevention of Access Stenosis Study
DCCT/EDIC	The Type 1 Diabetes Control and Complications Trial/ Epidemiology of Diabetes Interventions and Complications
DPP	Diabetes Prevention Program
DPPOS	Diabetes Prevention Program Outcome Study
DPT-1	The Diabetes Prevention Type 1
FIND	The Family Investigation of Nephropathy and Diabetes
GoKinD	The Genetics of Kidneys in Diabetes
HALT-C	The Hepatitis C Antiviral Long-term Treatment against Cirrhosis
HEALTHY	Middle-School-Based Primary Prevention Trial of Type 2 Diabetes
HEMO	The Hemodialysis Study
LookAHEAD	Action for Health in Diabetes
LTD	Liver Transplantation Database
LTD2	Liver Transplantation Database Followup
MDRD	The Modification of Diet in Renal Disease
NANS	National Analgesic Nephropathy Study
PEDS-C	Pegylated Interferon Ribavirin for Children With HCV
PRIDE	The Program to Reduce Incontinence by Diet and Exercise
T1DGC	The Type 1 Diabetes Genetics Consortium
TEDDY	The Environmental Determinants of Diabetes in the Young
VIRAHEP-C	The Study of Viral Resistance to Antiviral Therapy of Chronic Hepatitis C