

Providing Contemporary Access to Historical Biospecimen Collections: Development of the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC)

Carol A. Giffen,¹ Leslie E. Carroll,¹ John T. Adams,¹ Sean P. Brennan,¹
Sean A. Coady,² and Elizabeth L. Wagner³

The National Heart, Lung, and Blood Institute (NHLBI), within the United States' National Institutes of Health (NIH), established a Biorepository in 1976 that initially archived biospecimens from population-based blood product safety surveys. It was later expanded to biospecimens from clinical and epidemiological studies in heart, lung, and blood disorders. The NHLBI also established a Data Repository in 2000 to store and distribute study data from NHLBI-sponsored research. The NHLBI Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) was established in 2008 to develop the infrastructure needed to link the contents of these two related NHLBI repositories, facilitate access to repository resources, and streamline request processes.

Three key program subcomponents were developed simultaneously: 1) the linkage of biospecimen electronic inventory records with their clinical or characterization data; 2) the development and implementation of a website with both public-facing information and private processing workspaces; and 3) the development of processes to maximize efficiency via a web-based system while maintaining workflow control, document tracking, and secure processes.

The BioLINCC website was launched on October 1, 2009 with eight biospecimen collections and data from 72 research studies. By the end of the fourth online year, 38 biospecimen collections were linked and posted, and data from 108 research studies had been made available for request. The number of registered users by the end of the fourth online year approached 2600, and continues to show a trend towards an increasing rate of new users per year. BioLINCC has fulfilled 381 requests comprising 851 data collections, as well as 600 teaching dataset requests and 75 data renewal agreements. 154 biospecimen requests comprising 147,388 biospecimens were fulfilled or actively in process. We conclude that the BioLINCC program has been successful in its goal to increase the visibility and utilization of NHLBI biospecimen and data repository resources.

Introduction

THE NATIONAL HEART, LUNG, AND BLOOD INSTITUTE (NHLBI), one of the 27 Institutes and Centers of the US Government's National Institutes of Health, supports research aimed at preventing and treating conditions related to heart, lung, blood, and sleep disorders. This research includes support for investigator- and Institute-initiated clinical studies. To maximize the scientific use of the data collected on these studies, the NHLBI established a Data Repository in 2000 to store and distribute the study data. The policies used to share clinical study data and the benefits of sharing the data with the scientific community have been previously published.^{1,2}

Many, but not all, of the studies in the Data Repository had biospecimens transferred to the NHLBI Biologic Specimen Repository (Biorepository) for long-term storage and distribution. The Biorepository was initially established in 1976 to support the need for archival research biospecimen collections to address emerging blood-safety concerns, and these collections have made several seminal contributions in human immunodeficiency virus (HIV), and hepatitis B (HBV) and C virus (HCV) research.³ Towards the end of the 1990s, the mission of the Biorepository was expanded to include biospecimen collections from clinical studies in heart, lung, and blood diseases. These studies include unique cohorts and rare diseases such as the Honolulu Heart Program (HHP), Pediatric

¹Information Management Services, Inc., Calverton, Maryland.

²Epidemiology Branch, Prevention and Population Sciences Program, Division of Cardiovascular Sciences, and ³Translational Blood Science and Resources Branch, Division of Blood Diseases and Resources, National Heart, Lung, and Blood Institute, Bethesda, Maryland.

© The Author(s) 2015; Published by Mary Ann Liebert, Inc. This Open Access article is distributed under the terms of the Creative Commons Attribution Noncommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Pulmonary and Cardiovascular Complications in Vertically Transmitted HIV Infection (P2C2), the Hemochromatosis and Iron Overload Study (HEIRS), and the Multicenter Hemophilia Cohort Studies (MHCS), among others.

Although there were numerous clinical and epidemiological studies that had both their biospecimens and data housed within the NHLBI Biorepository and the NHLBI Data Repository, respectively, these repositories were maintained separately. The NHLBI Biorepository has been managed by the NHLBI Division of Blood Diseases and Resources (DBDR), and its physical repository has been funded since 1976 using a contract that is competed every 5 to 7 years. The NHLBI Data Repository was established in 2000 and managed by the Epidemiology Branch in the NHLBI Division of Cardiovascular Sciences (DCVS). Each program had separate paper-based application and approval processes, making it challenging for researchers, particularly non-study investigators, to apply for both specimens and data simultaneously. In addition, searching for biospecimens by clinical data characteristics required requesting the current biospecimen inventory. Due to the then lengthy process for biospecimen searching and then for request approval, use of the Biorepository by non-study investigators was minimal.

As the number and type of biospecimen collections increased, it became clear that a strategy to increase use and manage the growing biospecimen inventory efficiently had to be developed. In 2007 the NHLBI implemented a 5-year business plan to increase utilization and improve efficiencies. Key to the success of the plan was the establishment of the infrastructure that would be needed to centralize the data sets and documentation in the Data Repository, link the study participant data to the biospecimens to enable informative searches to be performed, and make biospecimens and/or data available via online application using standardized procedures and efficient workflows. A Request for Proposals was released for a competitive procurement of coordinating center services to support this infrastructure in September 2007 and the BioLINCC program contract was awarded September 2008.

The primary objective of the program is to maximize the scientific value of historical and contemporary NHLBI biospecimen and data collections by facilitating access by qualified researchers to these research resources, and to further enhance utilization by promoting awareness of these resources to the research community. In this article, we describe methods used in the development of the informatics infrastructure and workflows used by the BioLINCC program and the challenges encountered in making these historical biospecimen collections openly available.

Methods

To accomplish the BioLINCC objective, it was first necessary to develop the infrastructure within each of the following three main task areas to support the intended functionality in a streamlined and sustainable manner.

Task Area 1: Organization of the data and biospecimen inventories

- Dataset and documentation review and preparation for the website
- Biospecimen inventory review
- Linkage of biospecimen with associated clinical data
- Final preparation for sharing

Task Area 2: Development of the BioLINCC website

Task Area 3: Development of standardized workflows for request submission and review

The components of each of these task areas are described below.

Task Area 1: Organization of the data and biospecimen inventories and documentation

Dataset and documentation review and preparation for the website. At BioLINCC program initiation, more than eighty data packages for historical or ongoing clinical trials and epidemiologic studies were available for request from the NHLBI Data Repository. These data packages included a variety of documentation as provided by the original (parent) studies, and generally included the study protocol, manuals of procedures and operations, annotated study forms, and other original study management documents.

The datasets themselves were of two types: private, unredacted study data for NHLBI archival purposes, and public study data that had been redacted to remove personal identifiers (for example, names, specific dates, addresses, clinic locations, social security numbers, etc.) or to collapse or recode data (for example, recoding ages to age grouping, recoding specific outlier values to high or low end groupings), to ensure that study participants could not be readily identified. Also included in most data packages were files, often referred to as README files, which provided data recipients with information regarding the history of the study and how it was conducted, any key events or changes in study methodology during the its course of conduct, and details of the redactions and code-collapses that were applied in the creation of the public datasets.

The data packages transferred from the NHLBI Data Repository to BioLINCC were inventoried and reviewed. Frequencies of the variables received were compared to their publications where possible. For each study, the variables in the redacted files were examined against the documentation to assure completeness or to identify areas for query. Although in some cases there were no remaining contacts among the data coordinating center, study staff or NHLBI program office, in most cases study staff could be found who were in a position to provide either historical information regarding the study conduct or additional documentation which was not included in the original submissions. If supplemental information could not be found, that was documented as an informational note to the file.

As a federally-funded website, of particular concern was compliance with the revised Section 508 of the Rehabilitation Act regarding access to electronic and information technology by those requiring assistive technology. The transferred study documents number in the thousands of pages, and most of them to be posted to the study descriptions on the BioLINCC website were not in compliance with Section 508 requirements. It is important to note that many of the studies were conducted well before these requirements existed; in fact, many of the study documents were in the form of scanned photocopies of materials that were originally produced via typewriter, and/or contained complex tables, manual annotations, or other characteristics which preclude 508-compliance conversion using existing technology.

The historical documents were examined to determine the level of effort that would be required to convert each set.

They were ranked to one of four levels: 1) already compliant or nearly-compliant, 2) medium conversion complexity, 3) high conversion complexity, 4) impossible, the file would need to be re-created. A Section 508 waiver was granted by the Department of Health and Human Services (DHHS) for the posting of historical documents, with the stipulation that 508-compliant versions would be created whenever possible over time. A second stipulation was that all documents that were not yet compliant would be tagged with a screen-readable message providing contact information for assistance, and that BioLINCC staff would respond to such requests for assistance within a business day of the request. In all cases, 508-compliant data dictionary listings of variable names and definitions by dataset were created from the redacted public data via SAS for posting to the website. The purpose of the data dictionaries is to provide a variable-by-variable description of file contents.

Biorepository inventory review. At the time of BioLINCC contract initiation, the NHLBI Biorepository housed approximately 5 million biospecimens from 64 clinical and epidemiologic studies, the oldest dating back to the 1960s. By far the largest collections were those from the transfusion medicine population-based studies conducted within the Retrovirus Epidemiology Donor Study (REDS) during the 1990s, which at program initiation comprised approximately 2 million plasma, serum, and whole blood specimens.

Initially established in 1976, the physical custodianship of the Biorepository changed hands on several occasions. Likewise, the inventory management/tracking methods evolved by both time and support contractor, beginning with paper files that were later migrated to Excel spreadsheets and other electronic record types, which were then migrated to the Biological Specimen Inventory System (BSI). In addition, biospecimens that were deposited into the Biorepository came in over time through different quality assurance/quality control protocols based upon biorepository contractor, then-current capabilities and perspectives on inventory management, varying budgetary considerations and differing delivery methods (i.e., incremental deposits vs. bulk relocations of entire freezers). As part of the preparation for the anticipated increased demand for Biorepository specimens that would result from the BioLINCC program, it was clear that a close inspection of inventory records vs. physical Biorepository contents was needed.

A rigorous quality check effort was undertaken to assess random selections of vials within the various study collections to ascertain the accuracy of the inventory records versus the biospecimens themselves, including freezer/box locations, approximate volumes, possible discrepancies between material types, and other vial observations. Random selections were generated by BioLINCC programmer staff by study collection; the percentage of selected vials in the initial round varied based upon the collection size. Most percentages were in the 1%–5% range. The selections were forwarded to trained Biorepository staff to perform the vial-level inspections. Collections with high levels of discrepancies, particular among critical inventory characteristics such as missing/damaged vials or missing subject identifiers, were either over-sampled or underwent complete re-inventory to ensure that the BSI inventory records accurately reflected the Biorepository collections. Lesser discrepancies were defined as moderate criticality (e.g., hemolysis, location discrepancies across boxes, easily discerned noncritical label typos) or as low criticality (e.g., volume, location discrepancies within boxes,

etc.) were addressed through inventory file corrections or notations.

Linkage of biospecimens with associated clinical data. Concurrent with the development and quality control of clinical data files/documentation and the quality assurance/update activities of the biospecimen QC activities was the development of definitive linking files to associate biospecimen vials with their clinical or characterization data. This step was crucial to be able to search for specific biospecimens which meet research-specific selection criteria. Although a few of the transferred studies had existing linking files, in general this was a data-intensive effort comprised of inspecting across data file variables and the full BSI database for matching fields. Matches might include any number of BSI variable types, including specimen ID, old label contents, BSI ID, subject ID and others.

Typically, once the appropriate matching fields were identified on either a study-specific or study-center-specific level, most of the vials could be readily linked to their data. Particularly among the older collections, or collections that had been transferred in bulk without re-inventory at entry, residual problem vials would be identified that required further investigation, either in the form of contacting original study or NHLBI program office staff, or retrieving historical Biorepository archive files. Ultimately, virtually all (>98%) of the vials in the Biorepository collections were linked with their clinical/characterization data. Those vials which could not be linked to their data were flagged in the BSI inventory system to exclude them from biospecimen searches. These vials will be considered for future discard.

Final preparation of a biospecimen/data collection for sharing. The final steps in the preparation of a collection to be made available as a shared resource are as follows:

- Any associated biospecimens must be flagged in the inventory system for restrictions on sharing based upon the terms of the participant's signed Informed Consent agreement.
- The biospecimens must be linked with associated clinical, longitudinal, or characterization data, and unlinked vials must be flagged for exclusion from requests.
- The requirements for vial re-labeling must be established for the collection. In some cases, the original physical vial label includes information which may not be shared, such as participant name, initials, medical record or social security number or original study ID. These vials must be re-labeled prior to shipment.
- A link between the vial inventory records in the BSI is established.
- A study summary and study background web-page is written and approved by the NHLBI program office.
- Posted materials are either 508-compliant or are historical documents which have been appropriately tagged with screen-readable BioLINCC contact information for assistance.

Task Area 2: Development of the BioLINCC website

The BioLINCC website is accessed at www.biolincc.nhlbi.nih.gov.

This website is the primary interface with the BioLINCC Program. The system is comprised of two levels: public information which is available to all users, and a private workspace which is available to registered users.

The public website provides a wealth of information on NHLBI clinical and epidemiologic studies that have either data or biospecimens in the NHLBI repositories, including study summaries, references, and study operational documents. This information is provided in the form of searchable study web pages. More in-depth search functionality of all material archived on the website is also available. The collections posted on the BioLINCC website are of three types:

- Open collections may include biospecimens and/or clinical data. Open collections are fully within the custodianship of the NHLBI and may be requested online through the BioLINCC program.
- Proprietary collections include biospecimens only. These legacy collections are housed within the NHLBI Biorepository but are under the control of the Parent Study. Release of vials within these collections requires application to the Parent Study and is subject to review according to Parent Study procedures. Fulfillment of approved Proprietary requests is performed by BioLINCC.
- Non-BioLINCC Resources are biospecimen and/or data collections which are managed outside of the BioLINCC program, but which are included on the BioLINCC website to provide additional research options from a wider range of heart-, lung- and blood-related resources. Contact information for these resources is provided for their request, which takes place entirely outside of the BioLINCC program.

The BioLINCC website provides a variety of additional resources and information, including guidelines and recommendations for building new biorepository collections, links to information on how to prepare study datasets which are suitable for use as shared resources, BioLINCC forms, templates and flowcharts, and a news feed of recent additions and updates to the posted collections.

The BioLINCC website also provides a private workspace for the processing of requests. Registration is required in order to request resources from the NHLBI Biologic and Data Repositories. Visitors to the BioLINCC website who wish to access the full functions of the site are asked to register. Registration is quick and free. Upon submission of the user's contact information on the registration request, an automated confirmatory email is transmitted to the stated email address. Clicking upon the enclosed link establishes the account and enables access to the private workspace features of the site. Once a request has been submitted, communications and document transmissions between the researcher and BioLINCC are handled electronically through the secure website request private interface.

The BioLINCC website was developed using an open source technology stack to simplify interoperability with existing data systems. The main components of the stack currently in use are the CentOS Linux operating system (version 6.6), an Apache web server (version 2.4.12), a PostgreSQL database (version 9.1), and Python (version 2.7), specifically the Django implementation framework.

Task Area 3: Development of standardized workflows for request submission and review

Prior to the development of the BioLINCC program, procedures had been established within DCVS for the review and approval of data requests by the DCVS program official. With the establishment of the on-line request sys-

tem, the manual processes that had been used, such as the physical mailing of paper requests, required supporting documentation (research plan, CV, Institutional Review Board (IRB) approval), the signed data use agreement and the creation and mailing of the dataset and documentation CDs were replaced with document uploads to the private request workspace and with download transmissions of the requested data. However, the data request review process itself did not require modification.

In contrast, the biospecimen request review process that had been established prior to BioLINCC was found to require significant modification to support the increase in biospecimen requests resulting from the online process. The existing review procedure for all requests from outside researchers had required the assembly of a quorum of at least three members of the standing external Repository Allocation Committee, as well as the identification and recruitment of subject matter experts and at least one representative from the Parent Study. Between identification/recruitment challenges and the scheduling conflicts inherent in a group of senior researchers, review meetings could rarely be scheduled within 6 months of request finalization. Each meeting required a significant level of effort by both NHLBI and BioLINCC staff both administratively and in the preparation of review documents, summaries and background information for the participants.

It became clear that, although the existing process required significant time and effort for all parties, the numbers and types of biospecimens being requested rarely had the potential of significant impact on the collections. We accessed the review processes described by thirteen biorepository resources across eight of the Institutes within the NIH to generate ideas for streamlining our approach. We found that a variety of approaches were in use, mostly involving internal review panels with or without external reviews and/or reviews by original site investigators. A difference between most of the resources surveyed vs. the NHLBI BioLINCC resource collections was the diversity of the NHLBI collections and the resulting requirement for expertise across multiple clinical disciplines. The historical nature of many of the biospecimen collections also impacted the feasibility of identifying original site investigators.

In consultation with NHLBI Leadership and senior staff within each of the extramural Divisions, it was determined that an impact-based review system could be sustainable in the context of an increased request burden. The review approach that was ultimately adopted provides review guidance at three vial impact levels (low, medium, and high). Table 1 provides an overview of the definitions, request restrictions, and review options for each vial impact level.

The final determination of vial impact level is defined by the NHLBI, and following NHLBI guidance, BioLINCC staff assigns the impact level to each requested vial. Whenever possible without affecting the proposed research, low impact vials are selected in favor of those of medium or high impact. However, in some cases, requests may include vials that span multiple impact levels. The review option as described in Table 1 is selected based upon the highest vial impact level. The final decision on vial release may be that all vials may be provided, or it may be that only vials within low- or low-to-medium impact groups can be offered. In the case of partial approval, the decision is conveyed to the requestor to ensure that the vials proposed for request

TABLE 1. BIOSPECIMEN REQUEST “IMPACT”: DEFINITIONS, REQUEST RESTRICTIONS AND REVIEW OPTIONS FOR VIALS WITHIN IMPACT GROUPS

<i>Vial impact group</i>	<i>Definition</i>	<i>Request restrictions</i>	<i>Review options</i>
Low impact	Abundant biospecimens <i>abundant: vials within this group can support more than six new requests based on historical use and/or other collections can support similar research</i>	Can be used for exploratory research Requests to place the biospecimens on hold for six months are generally granted	NHLBI or Biorepository staff with expertise in the scientific area, or the funding group performing the scientific review
Medium impact	Unique and limited biospecimens <i>unique and limited: vials within this group can support four to six new requests based on historical use and no other biospecimens are available</i>	May be used for exploratory research questions but pilot studies may be required Requests to place the biospecimens on hold for six months are occasionally granted	NHLBI staff with expertise in the scientific area and knowledge of the Parent Study, or the funding group performing the scientific review
High impact	Unique and very limited resource <i>unique and very limited: vials within this group can support three or fewer new requests based on historical use, last vials may be included and no other biospecimens are available</i>	Not available for exploratory research Requests to place the biospecimens on hold for six months are granted only in exceptional circumstances	NHLBI staff with expertise in the scientific area and knowledge of the Parent Study (external experts may also be used), or the funding group performing the scientific review

fulfillment are sufficient for the research purpose. The revised tiered-impact approach brings the review process internally within NHLBI or Biorepository staff for low impact requests, within NHLBI or scientific review by funding groups for medium impact requests, and reserves more expanded external review to requests with high impact on the specific biospecimen collection. As a result, biospecimen request reviews can usually be completed within 2–4 weeks rather than the 6–9 months under the previous approach.

Common to both data and biospecimen requests is the requirement for a Research Materials Distribution Agreement (RMDA). This document was developed jointly by NHLBI DCVS, DBDR, and legal staff with input from BioLINCC and the Biorepository. Designed to replace the various agreement types that had been required prior to the establishment of BioLINCC, the RMDA is composed on-the-fly during the electronic request process. It incorporates relevant clauses based upon the request type, and pre-populates included fields such as Parent Study, Research Study Title, dates, Principal Investigator (PI) name and contact information, Authorized Users, and other required fields. Once generated, the document is signed by the authority at the PI’s institution, and then counter-signed by NHLBI program staff. Because there is a 3-year limit for data (but not biospecimen) use, data users are prompted near the end of the RMDA period either to certify that the data have been destroyed or to submit a dataset renewal request.

Methods to measure program success

Although the development of metrics to gauge program success in meeting its objectives continues to evolve, three primary indicators have been incorporated since its inception. These are basic measures of overall resource utilization, the expansion of utilization beyond original study investigators, and program visibility as measured by website and user counts over time.

The first of these relates to the overall utilization of NHLBI biospecimen and data resources prior to BioLINCC and then subsequent annual utilization. Information on pre-BioLINCC distributions was obtained from NHLBI program office records and the Biorepository’s BSI inventory system. The BioLINCC website maintains tracking information on all requests submitted via that mechanism, and so request types, dates, counts and final dispositions can be readily retrieved.

The second indicator relates to the extent to which the BioLINCC program has helped to expand the utilization of biospecimen resources beyond original study investigators to the wider scientific community. One clear measure of this is the relative extent to which Proprietary Collections are distributed (indicating original study investigators or collaborators) vs. biospecimens from Open Studies. The necessary components to examine this trend over time were identified from pre-BioLINCC DBDR and BSI records, and from information from the website for post-initiation requests.

A third indicator relates to the assessment of program visibility as measured by website access metrics and by the number of unique users who not only view the site but who also become registered users. Website monitoring software is utilized to assess the former, while the registered user database on the site provides information on the latter.

Preliminary program assessments

The BioLINCC home page at www.biolincc.nhlbi.gov provides access to study-level searching, text searching of indexed documents and attachments within the site, links to the major subcomponents of website content, recent news alerts, and log-in/registration functions.

Each collection within the Repositories is described via a BioLINCC study webpage, which includes high-level information on the study conduct; a summary of the study design, population, results and conclusions; a count of available biospecimens by material type (if applicable); and

TABLE 2. ANNUAL DATA AND BIOSPECIMEN COLLECTION AVAILABILITY AND DISTRIBUTION PRIOR TO AND DURING THE FIRST 4 ON-LINE YEARS OF THE NHLBI BioLINCC PROGRAM

	<i>2 years prior to BioLINCC</i>	<i>1 year prior to BioLINCC</i>	<i>BioLINCC transition (pre-Website)</i>	<i>Online year 1 10/2009– 9/2010</i>	<i>Online year 2 10/2010– 9/2011</i>	<i>Online year 3 10/2011– 9/2012</i>	<i>Online year 4 10/2012– 9/2013</i>
Biospecimen collections available on-line	0	0	0	20	29	35	38
Biospecimen requests	12	13	19	23	54	55	22
Data collections available on-line ¹	62	67	72	80	87	95	108
Data collection requests ²	66	66	75	79	111	135	56
Data collections distributed ³	100	106	115	118	221	307	205

¹Study Data Collections were listed on the NHLBI website prior to BioLINCC but the request process was manual.

²Study Data Collections requests are frequently for data collections only and do not include biospecimens.

³Multiple Study Data Collections may be requested in a single request.

study documents such as the protocol, data collection forms, study manuals, etc. Biospecimen counts are generated from a real-time query of the biospecimen inventory at the first viewing each day and held in cache until regeneration at the first viewing of the following day.

When the BioLINCC site was established in 2009, it began with the posting of 8 biospecimen collections and data from 72 research studies. By the end of the fourth online year, the available studies in the Open Period had increased to 38 biospecimen collections and 108 data collections; 31 of these have both biospecimens and associated study datasets. In addition, information on 14 Proprietary biospecimen collections and 3 Other Available Resources had been provided on the website. The full listing of collections currently available for request may be found at <https://biolincc.nhlbi.nih.gov/studies/>.

As previously stated, the primary objective in the establishment of the BioLINCC program was to maximize the scientific value of historical and contemporary NHLBI biospecimen and data collections by facilitating access by qualified researchers to these research resources, and to further enhance utilization by promoting awareness of these resources to the research community. Preliminary observations indicate that BioLINCC has been successful in meeting that objective from several perspectives.

There has been a striking increase in the utilization of the historical NHLBI biospecimen and data resources since the establishment of the BioLINCC website in October 2009.

There are four principal types of requests that may be fulfilled through BioLINCC. In decreasing order of frequency, these are teaching datasets only; study data collections only; specimens with or without full study datasets; and study dataset use renewal agreements. Researchers requesting study data collections only often request multiple studies within the same data request. Study data collections are released only in their entirety; data subsets are not released. A compilation of the resource distribution records for the 2 years prior to the establishment of BioLINCC, distributions during the first year of BioLINCC but prior to website launch (transition period), and distributions over the first 4 full years since the website was opened illustrates the effectiveness that the online access has provided.

Table 2 provides information on the counts of the biospecimen collections and study data sets available online at the end of each index year, the number of requests for each resource type, and the number of data or specimen collections which were distributed to the scientific community. During its tenure over the reporting period, BioLINCC has fulfilled 381 requests comprising 851 data collections, as well as 600 teaching dataset requests and 75 data renewal agreements. 154 biospecimen requests comprising 147,388 biospecimens were fulfilled or actively in process. This table demonstrates both the increase in the number of resources which were made available for request as well as the increase in distributions.

Table 2 also illustrates the impact of funding on research material distribution. By the second online year, the BioLINCC

TABLE 3. STUDY PERIOD REQUEST SUMMARY—SUCCESSFUL BIOSPECIMEN REQUESTS BY OPEN VS. PROPRIETARY COLLECTION STATUS

	<i>2 Years Prior to BioLINCC</i>	<i>1 Year Prior to BioLINCC</i>	<i>BioLINCC Transition (pre-Website)</i>	<i>Online Year 1 10/2009– 9/2010</i>	<i>Online Year 2 10/2010– 9/2011</i>	<i>Online Year 3 10/2011– 9/2012</i>	<i>Online Year 4 10/2012– 9/2013</i>
Open requests / (aliquots identified)	5 / (226)	5 / (1,975)	9 / (4,646)	14 / (14,351)	50 / (55,682)	54 / (66,481)	18 / (18,159)
Proprietary requests / (aliquots identified)	7 / (20,348)	8 / (10,421)	10 / (30,595)	9 / (6,990)	4 / (1,815)	1 / (1,400)	4 / (2,654)
Total requests with successful searches	12	13	19	23	54	55	22
Total aliquots identified	20,574	12,396	35,241	21,341	57,497	67,881	20,813
% requests from open studies	42%	38%	47%	61%	93%	98%	82%
% aliquots from open studies	1%	16%	13%	67%	97%	98%	87%

infrastructure, number of linked resources, and workflow processes had sufficiently matured to the level that promotional activities could be undertaken and supported. Therefore, during the second and third online years, NHLBI grant funding was made available specifically for biospecimen research using BioLINCC resources through a competitive R21 process (RFA HL 12-004).⁴ Peaks of request activity may be seen for both of those online years for all resources. This funding was not available in the fourth online year, and we speculate that the relative decrease from year 3 to year 4 reflects that fact.

A second indicator of the success of the program in reaching its goal to expand access to the wider scientific community is the study period of the biospecimens being requested. The biospecimens in the NHLBI Biorepository include both those that are still in the Proprietary Period, meaning that their access is partly controlled by the Parent Study, and those which are designated as Open Period, meaning that their access is under the custodianship of the NHLBI. Proprietary Period biospecimen utilization has historically been by the Parent Study researchers themselves, or by outside researchers with both a collaborative relationship with the Parent Study and access to its guidance. However, this type of support is necessarily limited by Parent Study funding which inevitably declines over time.

In the context of creating a vibrant and sustainable biorepository resource for the future, the key is to bring in new outside researchers, without ties to the original Parent Studies. To do this, the Biorepository must be able to provide sufficient information on the available biospecimen and associated data so that research can be conducted independently of Parent Study support. A measure of the breadth of scientific community access to the NHLBI Biorepository is the relative distribution of successful requests within the Proprietary vs. the Open Periods over time. Mirroring the same time windows as displayed in Table 2, above, Table 3 displays the numbers of requests and the number of requested biospecimen aliquots during the reporting periods. As reflected in both the percentage of requests for Open Study resources and the percentage of aliquots from those resources, there has been a considerable change in the relative numbers within Proprietary vs. Open collection categories. Once a minority, the vast majority of requests are now from the wider scientific community for Open resources.

An additional measure of program success is the assessment of program visibility. Although awareness of the program is reflected in the requests submitted and fulfilled, another approach is through the use of website statistics to examine the degree to which users are accessing the website, and whether the audience appears to be expanding over time due to increased awareness or if site use is leveling off. In the BioLINCC program, users must become registered to request resources, although registration may be done without a request. However, most of the website functions are available to all users regardless of registration status.

By their very nature, unregistered users are not well defined, but most likely represent a mix of users browsing the site for basic information on NHLBI clinical and epidemiologic studies and results; more focused reviews of protocols, study documents and available biospecimen types prior to the submission of resources requests; and casual website browsers. Although not perfect, a measure of the number of different users who have ever accessed the BioLINCC website is through counts of unique IP

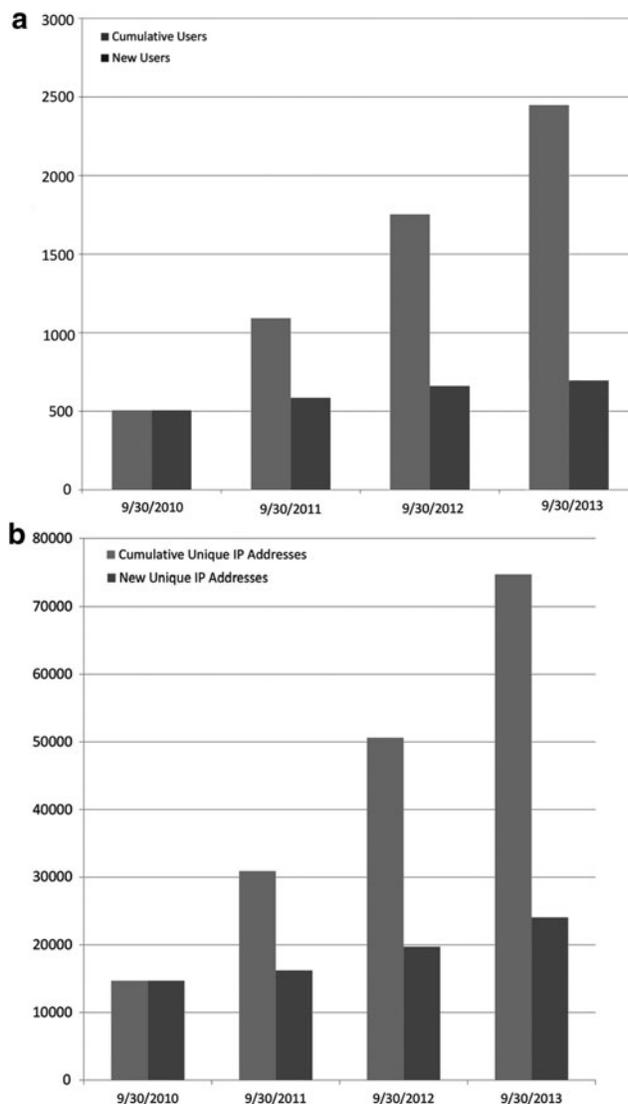


FIG. 1. (a) Cumulative and annual incremental increases in registered BioLINCC users over program years; (b) Cumulative and annual incremental increases in unique IP address access to BioLINCC over program years.

addresses. The website monitoring software excludes site access by bots and other automatic means, and only includes site access by individuals.

We examined the counts of registered users and the counts of unique IP address hits to assess access by both dedicated and more casual users over time. Figure 1a presents the cumulative number of registered users at the end of each of the first 4 website years, as well as the increase in users from the previous year's count. Figure 1b presents similar information regarding unique IP address access to the BioLINCC site over the first four years. Usage has continued to increase throughout the reporting period, and as of the end of the first four years there were 2419 registered users and 74,667 unique IP address site visitors.

Discussion

The NHLBI has actively pursued the acquisition of research study biospecimen and data collections for decades

with the intent of making those resources available to the broader scientific community for future research. The acquisition and maintenance of these stored collections, particularly the biospecimen collections, has required a considerable financial commitment by the Institute. Despite the increase in the number of biospecimen collections across heart, lung, blood, and other disorders of interest to the Institute, it became apparent that the collections were underutilized, and a 5-year plan to increase utilization was established. A large part of this plan was to increase utilization not only through increasing visibility via a web-based mechanism, but also but establishing the infrastructure that was necessary to be able to link each vial to its full clinical data as well informed consent restrictions on research use, and including collection dates relative to its respective study timelines. By enriching the information available on each vial and providing an accessible mechanism for the request process, the goal of the plan was to maximize the contemporary usefulness and utilization of historical repository resources.

The BioLINCC program was initiated under a contract mechanism in September 2008, and was funded for approximately 5.5 annual full-time equivalent staff comprised of project management, web developers, analytic programmers, biomedical analysts, and support staff, at a cost of \$800,000 per year. Each of the tasks required about a year's worth of development, with the initial release of the studies prepared under Task 1 and the public release of the website developed under Task 2 having been accomplished by the end of Year 1. The amount of time required to run a new collection through the processes described in Task 1 varies widely, depending on the state of the data and documentation and the accuracy of the inventory and linking file, but probably averages about 100 hours of staff comprised of analytic programmer, biomedical editor, and Section 508 remediation.

The review methods development described in Task 3 also required about a year of development in calendar time, including research and implementation but also including Institute approval processes. The prototype BioLINCC website was presented to NHBLI and Institute consultants three months after contract initiation and was released to the public on October 1, 2009, or about a year after contract initiation. The first release included all of the study data that were then available, and eight of the biospecimen collections that had been fully linked by that time and considered to be fully ready for distribution. The focus for that first release was on the largest biospecimen collections; for the most part these were the large population-based transfusion medicine studies (GSR/GLPR, HHV8, RADAR, TTVS), the COBLT cord blood collection, two clinical studies in heart disease (PEACE and HEIRS), and a clinical study focused on lung sarcoidosis (ACCESS). More than half of the total biospecimens were made available at the initial release. Additional new or updated data collections have been added throughout the period, as well as newly-linked or newly-available biospecimen collections. Through this process we have demonstrated that historic biospecimen and data collections can be linked, given the appropriate source data and inventory records, and made available for use in contemporary research.

The usefulness of these resources is anchored in the strength of the linkage and the richness of the clinical and inventory data. The original programmatic expectations related to the level of detail that could be explored in searching for biospecimens which were fit for the proposed research

were modest; specimen type, date, basic subject demographics, treatment group, or case/control status, etc. However, it quickly became clear from the submitted requests that most searches would be far more complex, involving numerous variables including biochemical/assay/genotype data, co-morbidities, outcome measures, and others. Furthermore, the production of listings of available specimens had to also include measurements of impact on these non-renewable resources, including volume restrictions, last-available vial considerations and the need to either combine low-volume aliquots within a specimen draw to create vials of the requested volume or to formulate cost-effective aliquot schemes for high-volume source vials. The latter considerations not only provide information needed by the Institute to decide which vials to distribute, but are also applied outside of the request process for overall Biorepository collection inventory maintenance or possible inventory reductions, and to guide approaches to freezer management.

During the process of making biospecimens fully ready for sharing, it has become apparent that collections that were expressly built as shared resources differed substantially from those with less well-defined or frankly vague intent to merely store some samples for possible future research. The value of prospective biorepository collections is maximized through meticulous planning for a reasonable number of specimens to be stored, standardized collection procedures and aliquot sizes, and attention to detail in inventories and linked data. Studies with a biorepository goal must include adequate, designated funding and staff to sustain the biospecimen and linked data from its collection through its transfer to the central repository.

The value of historical collections can be substantial but planning for such shared resources must include a commitment to not only build the data infrastructure but also to sustain the analytic and support staff required to maintain it and to guide access.

Conclusions

By currently available metrics, the BioLINCC program has been successful in its goal to increase the visibility and utilization of NHLBI biospecimen and data repository resources by the wider scientific community. Due to the lag time between the distribution of research resources, their analysis and their publication, there are currently insufficient data within the program metrics related to the capture and assessment of the scientific impact of the results obtained using distributed resources. Program activities will continue to obtain information on publications resulting from requested resources, as well as to characterize requests by investigator (e.g., early stage versus established), institution type (e.g., non-profit, academic, commercial), funding (e.g., grant type, institutional, etc.), and research aim (e.g., exploratory, pilot, definitive, etc.).

Acknowledgments

We would like to acknowledge and thank Lisbeth Welniak (Translational Blood Science and Resources Branch, Division of Blood Diseases and Resources, NHLBI) and Phyliss Sholinsky (Prevention and Population Sciences Program, Division of Cardiovascular Sciences, NHLBI) for their significant scientific and program development contributions to the BioLINCC program, as well as for their

critical reading of this manuscript and extremely helpful suggestions. We would also like to acknowledge and express our appreciation to Kathi Shea (NHLBI Biorepository at Precision Bioservices) for her guidance and advice in program development and workflow logistics.

BioLINCC is funded under NIH/NHLBI contract HHSN268201400014C. The NHLBI Biorepository is funded under NIH/NHLBI contract HHSN268201400035C.

Author Disclosure Statement

No conflicting financial interests exist.

References

1. Geller NL, Sorlie P, Coady S, et al. Limited access data sets from studies funded by the National Heart, Lung, and Blood Institute. *Clin Trials* 2004;1:517–552.
2. Coady SA, Wagner E. Sharing individual level data from observational studies and clinical trials: A perspective from NHLBI. *Trials* 2013;.14:201
3. Allain JP, Busch MP. Donation archives and prospective donor-recipient repositories: Indispensable tools for monitoring blood safety. *Transfusion* 2007;47:1110–1114.
4. The full text of R21 RFA-HL-12-004 may be found at <http://grants.nih.gov/grants/guide/rfa-files/RFA-HL-12-004.html>. Last accessed June 26, 2015.

Address correspondence to:

Dr. Carol A. Giffen

Information Management Services, Inc.

3901 Calverton Blvd, Suite 200

Calverton, MD 20705

E-mail: giffenc@imsweb.com