Scalable Biobanking: A Modular Electronic Honest Broker and Biorepository for Integrated Clinical, Specimen and Genomic Research

A Data Driven and Integrative Approach to Asynchronous, Longitudinal and Collaborative Specimen and Genomic Research

Alex S. Felmeister(1,2)*, Tyler J. Rivera(1), Aaron J. Masino(1), Adam C. Resnick(3), Jeffrey W. Pennington(1)

(1) Dept. of Biomedical and Health Informatics
The Children's Hospital of Phila.
Philadelphia, USA

(2) College of Computing and Informatics
Drexel University
Philadelphia, USA

(3) Department of Pediatrics
Perelman School of Medicine
University of Pennsylvania
Philadelphia, USA

*Contact: felmeistera@email.chop.edu

***Abstract*—Biorepository research has introduced significant challenges to biomedical informatics systems design and implementation. We take a best-of-breed system integrative approach to finding solutions to administer a biobank operationally, use a biobank scientifically and integrate clinical, specimen and genomic data. We introduce an electronic Honest Broker (eHB) and Biorepository Portal (BRP) open source project that, in tandem, allow for integration of data on the input and output while protecting subject privacy, a primary issue in research and healthcare informatics. Research data management systems like REDCap and our proprietary laboratory information management system (LIMS) are used as best-of-breed client systems and integrated with downstream genomic analyses. Taking this honest brokered system-client based, modular approach to biorepository research, we allow for data and specimens to be associated with a biorepository subject at any time point asynchronously, lowering the bar to develop new research projects based on scientific merit and having a future proofed specimen set for collaborative advanced genomic and tissue research yet to be established.**

Keywords— Biomedical Research; Biorepository Research; Translational Bioinformatics; Health Informatics; Precision Medicine; Honest Broker; Cancer Genomics; Data Integration; Data Representation; Information Systems; Open Source; Patient Health Information Protection; Patient Privacy; Computer Security

I. INTRODUCTION

Current research is yielding rapid advances in personalized, precision medicine through targeted therapies based on an individual's genome, genomic biomarkers, and cell biology[1]. This type of research has become increasingly dependent on the collection of large cohorts of high quality human biospecimens that are paired with clinical annotation[2]. While biospecimen-driven research is widely practiced, such efforts are often limited in scope as they require significant, lengthy manual processes of retrospective annotation and cohort identification coupled with recurrent institutional review applications for defined projects. The significant challenge for modern biorepositories we address in this paper is the creation of a robust, prospectively driven platform for the association of clinical diagnoses, patient and genomic data along with research data to a physical biospecimen that has gone through specific rigor in quality assurance practices. Though results of biorepository-driven research is expensive and theoretical, we see a trend towards these resources becoming indispensable in academic medical centers[2], [3].

Biorepository data is typically captured in longitudinal, asynchronous workflows that create software design and data integration challenges[4]. An ideal system must provide de-identified, granular and longitudinal data to researchers while also enabling data collection workflows that require patient identification[5]. The required data is heterogeneous with high dimensionality and often resides in separate systems such as a Laboratory Information Management System (LIMS), Research Data Capture tools, and the Electronic Health Record (EHR). Integrative solutions are necessary at the point of collection and at information and specimen retrieval. The data must be curated to ensure it is presented in an understandable representation for researchers in a specific medical domain[6], [7]. This paper covers the design, implementation and impact of a modular, best-of-breed software component solution developed at the Children's Hospital of Philadelphia, Department of Biomedical Informatics (DBHi) in partnership with the Children's Brain Tumor Tissue Consortium (CBTTC)[8] and the Clinical and Translational Science Award (CTSA). This approach and solution were developed in the

The Children's Hospital of Philadelphia Clinical and Translational Science Award (CTSA) through the Clinical and Translational Research Center (CTRC), U54 KL2RR024132, fund this project. The project is also made possible by the generous philanthropic effort of the Children's Brain Tumor Foundation (CBTF).

specific context of distributed biorespository and biobank studies in biological tissue and genomic research.

II. METHODS

A. Modular Approach

We took a modular, entity-based, best-of-breed integrated systems approach to facilitate this variable set of specimen acquisition and data collection events. The first and primary entity in this case is the subject. The subject entity is moved to the honest broker and stored in a master patient index (MPI). Each external research record in the data management tool or research system, in this case the data management tool (REDCap[9]), can have different entity relationship-type relations with the subject entity[10]. We built a research portal, dubbed the Biorepository Portal (BRP), that can access subject records in the eHB and subsequent external research records through token-based authorization from that client system. The BRP reproduces the REDCap electronic Case Report Forms (eCRFs) based on records stored in the eHB with a custom REDCap client utilizing the REDCap application programing interface (API) in real-time. This produces a complete form for that subject at time of access. It displays the subject information and identifiers at the top of the screen at all times during form data entry and while shifting from form to form. Section D of this paper, "REDCap Client", describes and illustrates this feature. A research coordinator or data manager can enter any temporal and longitudinal research data based on their protocol subject list at anytime or in any order (i.e. asynchronously) while maintaining the continuous de-identification and re-identification of research data automatically.

At the Children's Hospital of Philadelphia, its Biorepository Core Facility utilizes ThermoFisher Nautilus as its LIMS. As part of our method, we also built a client to this LIMS that allows for association of LIMS records with a subject record. In this way specimens can also be collected longitudinally over time. Data and specimen coordinators have the ability to associate sets of specimens with a subject or event and annotate that specimen on the fly in one front-facing system. For downstream integration, we use the same eHB software service to perform our Extract Transform Load (ETL) processes that are tailored to each project. The result is a fluid, nightly updating non-human subject research database that allows for seamless queries across research and clinical systems. We allow collaborators to access specific sets of data via the data exploration tool, Harvest, customized for each project[11]. Finally, the phenotype and associated specimen records can be integrated with genomic analysis and/or direct clinical data from the EHR with appropriate institutional permissions. Genomic data in the case of the cancer tissue-based biorepository is integration of tumor mutation analysis completed in house at CHOP and pulled from the genomic visualization tool (cBioPortal[12]) via a mix of web endpoints and ETL scripts. What follows is a description of each component of the modular tool kit illustrated at a high level in Figure 1.

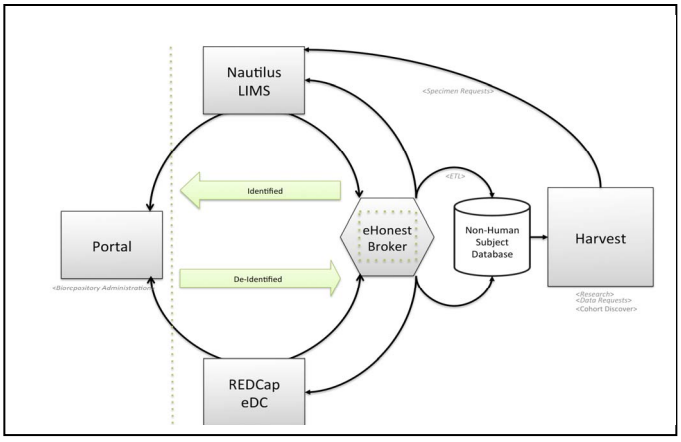

Figure 1. High Level Architecture of systems used in the integration methodology.

B. Electronic Honest Broker

The concept of an honest broker is not a new one and it can be seen implemented in other academic medical center environments to protect specific pieces of information when integrating research data[13], [14, pp. 56–107]. The eHB is a web-based software service with server-side encryption to maintain a MPI and associated research records. This component of the approach is the centerpiece of the biorepository operations. The initial studies/projects targeted with this solution began as a paper process of considerable complexity is now an informatics tool[15]. The MPI in the eHB uniquely identifies each patient through a combination of organizational association (e.g. CHOP) and a unique organization provided identifier (e.g. medical record number). The eHB associates each MPI record with system information and known record identifiers in external systems related to the subject. For annotated biobanking studies, the eHB maintains associations to records specifically in REDCap and the LIMS. The eHB makes data available via a representational state transfer (REST) web service. This technique allows the addition of new application clients to be system and programming language agnostic. To ensure appropriate access, the eHB uses token-based authorization, and encrypts its data both at rest and in transit, relying on appropriate client-side keys to decrypt the payload received from the API. For known applications, the eHB provides subject data with few restrictions. Client applications determine the context of what information is appropriate to display to a user, thereby enabling flexibility to meet different workflow and protocol requirements.

The eHB has a limited web-facing user interface that allows for the administration of access tokens and users, but can be managed through a comprehensive set of create, read, update, delete (CRUD) operations exposed by the REST API. Client applications, explained in subsequent sections, determine the context of the request. The client application requests resources of the eHB service via a URL endpoint secured using transport layer security (TLS) and, with appropriate keys and credentials, can read and write data to and from the eHB service. The eHB REST service handles authorization of the application, encrypts data and formats a response. The actual

database behind the eHB service stores only encrypted information and would be unreadable if accessed. This type of encryption decouples the identifiers and encrypts any and all information going into the eHB database and can be considered "privacy-by-design" by selectively sending and granting access to information based on context and only storing the minimum set of information needed to stitch together a record for data management or data query[16], [17]. The architectural design of the eHB, illustrated in Figure 2, utilizes web request type architecture to be a completely independent component of the tool-kit. The eHB model is similar to prior research in clinical informatics and the EHR. Architecturally, the health record must have the element of being future-proof. There is an assurance of openness and portability through standards, flexibility and scalability, semantic interoperability and acceptance from the domain experts. In Blobel et al., the authors discuss the fundamentals of future-proof health systems describing an "atomic component"[18]. We apply this notion by creating an "atomic component" that must be guaranteed utilizing the eHB to associate the subject on the study (as the atomic component) with their biospecimens, phenotypic and genomic data.

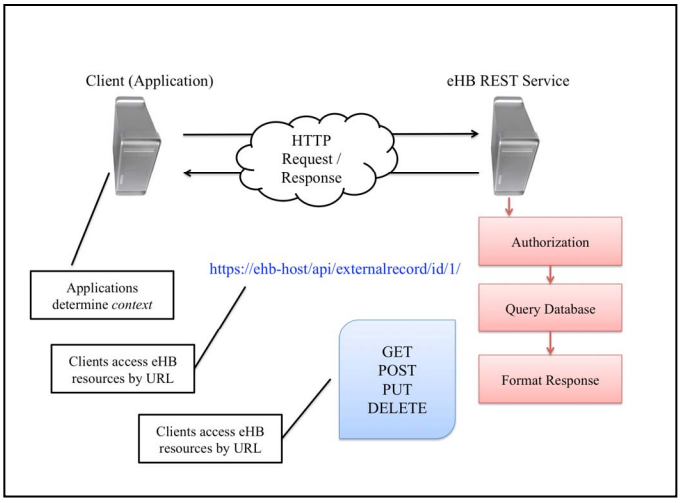

Figure 2. The architecture of the eHB software service and its interaction with client systems.

C. *Biorepository Portal*

The eHB service discussed in the previous section is exposed solely to external client systems through key authorization and does not have a researcher or research staff-facing interface. Therefore, we developed a user-facing component to allow research staff to work with subject identifiers, subject research data and associate specimen records. The BRP organizes the honest brokered data by integrating custom clients using external system APIs to stitch together, in real-time, the research data and patient identifiers. The BRP allows access via authentication utilizing CHOP's institutional identity management system that complies with all network access guidelines that meet the standard of access to hospital clinical systems. Users are granted access to subject records based on protocols. Each protocol is a walled-in list of subjects that only named users can manage. Within the walls of a protocol, there are groups of subjects that allow for users to differentiate between institutional or site-based cohorts within a protocol. Each unique subject enrollment creates a new record for that subject in the eHB MPI. Figure 3 shows a subject data entry point. In this example, the subject record has multiple data entry points and is associated with multiple phenotype and specimen records.

D. *REDCap Client*

The REDCap client in the BRP makes a request for the metadata and data from a specific subject record stored in the eHB and recreates the form requested utilizing the REDCap API. Figure 4 demonstrates the REDCap form client where the patient Medical Record Number (MRN), last name, first name and date of birth are exposed along with the customized research data capture form. Corresponding REDCap forms

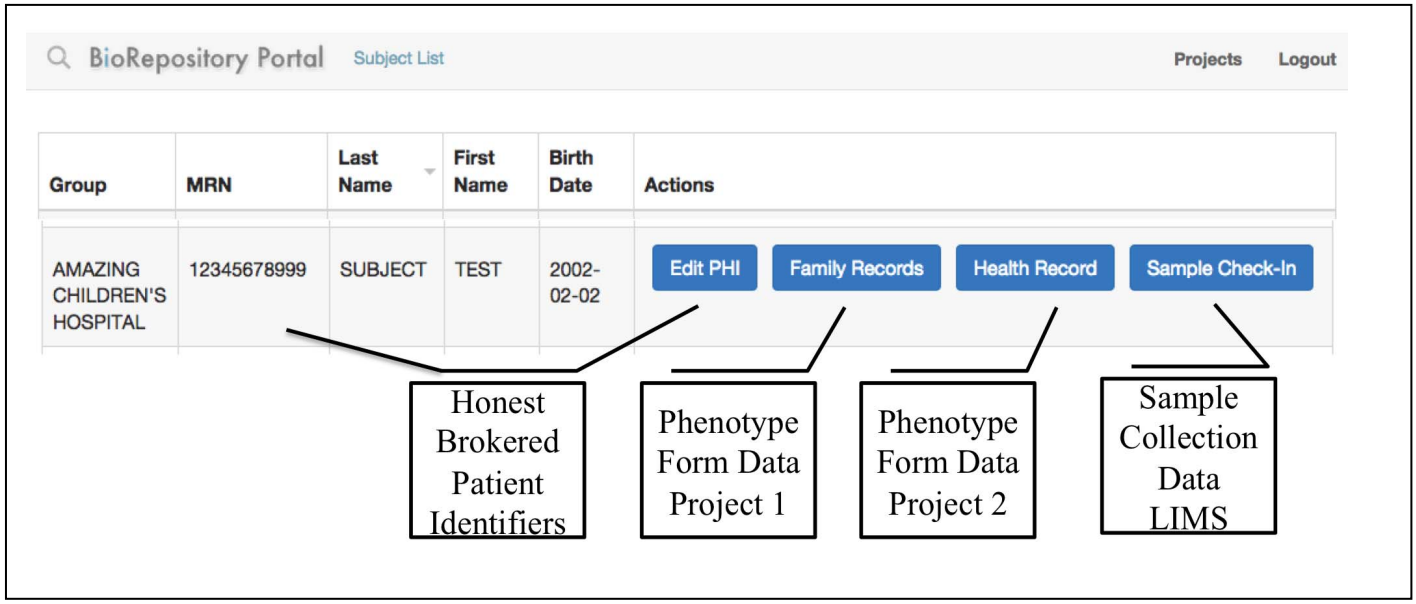

Figure 3. A screenshot from the BRP displaying identifiable data entry along side the de-identified client system.

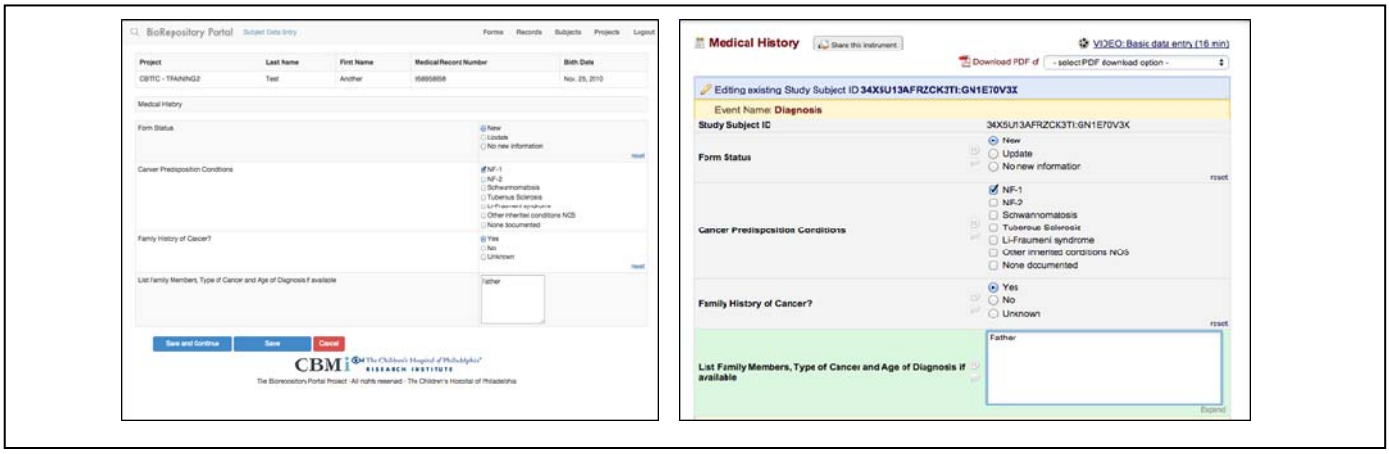

Figure 4. A side-by-side comparison of research electronic case report forms. The BRP produces the dynamic form identified while saving the data, de-identified, to the client system with an honest brokered identifier.

record are displayed by the BRP and assigned identifiers by the eHB. After the form record is saved, the BRP utilizes the eHB software service to either create a new record or modify an existing record in the REDCap project. The REDCap project record identifier is hashed and randomly generated without use of derived patient information. eHB identifiers are generated utilizing the application client key, in combination with a salted hash value which is guaranteed to be unique[19]. Creating a research identifier not derived from a direct patient identifier is required when using patient data for research[20]. Research identifiers are created by the connected research system randomly and are not patient-identifier derived.

Adding a layer that removes the subject entity from the REDCap projects associated with a study allows for REDCap to facilitate user access to projects, form building, data logging, and managing a study data dictionary[9]. The ability to supplement an entire REDCap project(s) or form(s) as specimen annotations is accomplished by associating another project with a project and, in turn, a subject. Our approach includes the ability to have variable numbers of projects and nested project records per patient. There are many variations of studies that use a variable number of REDCap projects/forms and project records depending on the need of the domain. For example, a BRP protocol can capture demographics one time in one REDCap project while collecting many diagnosis-type forms with longitudinal events in another project that allows for multiple records per subject. The eHB mediates and stores the links between the subject entity and their project records. Conversely, we allow for the tools to maintain separate cohorts of identified subjects where the data are stored in the same REDCap project. This is particularly important for studies in which multiple institutions are participating in sending data and specimens to one data and specimen-coordinating center (DCC/SCC). The link to REDCap records depends on the domain and temporal requirements of a biorepository study. The next section describes the LIMS Client.

E. Laboratory Information Management System Client

The LIMS, in our particular setup, assigns a random and incremented ID to a specimen collection event, and this identifier is associated with the subject entity in the eHB by user input via the BRP. The BRP has a custom client that allows specimen coordinators with the proper credentials to associate pre-labeled specimen accession kits with the subject entity. Specimen collection kits with proper collection tubes and labels are created prior to subject enrollment in the CHOP Biorepository Core facility. The specimen coordinator then scans the participant label on the kit through a LIMS client in the BRP to associate the kit with the patient. Any downstream laboratory work, for example; receiving, processing, analysis and storage are performed directly in the LIMS. The laboratory technicians processing and receiving specimen kits do not see patient identifiers, only the LIMS assigned identifier. This facilitates the longitudinal capture of multiple specimen collection events associated with one subject. Though we use a proprietary LIMS, the key takeaway is to choose a LIMS with a useful API.

F. Genomic Data Integration

One of the key methods to a useful biobank is closing the loop between phenotypic and genomic information. In the case of the cancer tissue biorepository, allowing a researcher to query specific somatic mutations from comparative analysis of sequencing the germ-line and the tumor and find the physical specimens with that mutation is powerful. For example, a tumor in the biobank may have a somatic BRAF mutation derived from a specific analysis pipeline. We accomplish this by integrating the cBioPortal mutation data loaded by internal and external analyses into our repository. The integrated query tool allows for a back and forth between gene searching, gene visualization in cBioPortal, and the tool for phenotype and specimen requests. We perform this integration in a similar fashion to the other tools in the tool-kit, by using combination of constructing web endpoints and pulling data based on information stored in the eHB and

research data systems. The cBioPortal, originally developed at Memorial Sloan Kettering Cancer Center (MSKCC) is an open access resource for exploration of multi-dimensional cancer genomics data sets[12]. Figure 5 shows the initial data integration performed on this dimensionally disparate data source with clinical and specimen data query resource.

The cancer genomics integration starts with a scripted pull of mutation data via a secure database connection utilizing elements of cBioPortal's relational data structure to store this large set of data. Specimens known to the repository are loaded into the cBioPortal by the bioinformatics team with known specimen identifiers from the LIMS. This creates a natural link between any granular genomic data, the specimen and ultimately the subject. URLs are constructed in the query platform, to be discussed in Section H, that allow for researchers to move directly to cBioPortal to visualize mutation data of interest. This simple integration point opens up a capability unavailable in previous biorepositories wherein one can search from mutations across a genomic data repository to pull a physical specimen for biological research and/or clinical phenotype data for clinical research based on cohorts from genomic results on the physical specimens.

G. Extract Transform and Load

Overall, a custom set of ETL scripts are written to integrate the disparate and de-identified data together for researcher. This ETL process can be described as another application with client access to the eHB. The first part of the ETL script utilizes application key-driven access to obtain a list of subject entities on specific protocols and their respective research identifiers. This ehb-link list is used throughout the ETL process to join together and integrate now disjointed research systems and perform further de-identification where required by a study protocol. The ETL process puts the data together in a simple relational format for researchers to access. It essentially creates one non-human subject database that is used to ask questions of biorepository data.

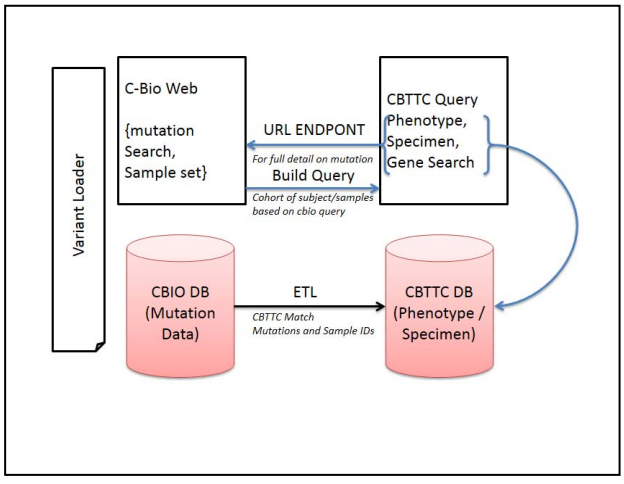

Figure 5. Cancer genomic mutation data integration utilizing the CbioPortal and Harvest.

The ETL process is also where we integrate systems and data that are not part of the data entry in the BRP. If allowed by the protocol, the ETL process can query the eHB for patient records and pull clinical data from the health record and move it to the non-human subject biorepository database.

H. Researcher Query Tool and Non-Human Subject Data Resource

Tying it all together with an ETL process is not sufficient to catalyze the research potential of a fully integrated biobank. We created an accessible data discovery tool that allows researchers to get quick answers to questions about the resource without involving the Institutional Review Board (IRB) because all data is de-identified in the query tool[6]. This tool is implemented in the open-source application stack, Harvest[11]. This tool gives the informatics team the ability to customize the application where necessary, but also have an out-of-box query tool for each biorepository domain around a set of relational data that can be published securely similar to any webpage for each biorepository study domain. An example of a Harvest instance where users can search the multi-institutional biospecimen and annotation data of the Children's Brain and Tumor Tissue Consortium is shown in Figure 6. This figure shows point-and-click access to multi-dimensional and disparate data in one place.

III. Results

The BRP and eHB integrative software-based approaches to biorepository operations and integrative biomedical science are used on several studies that include multi-dimensional usage of research systems in a longitudinal and asychronous manner. Specifically these methods:

1) Abstract the subject entity from the project entity in research data systems, reducing limiting factors of the one-to-one paradigm of data management tools and patient identifier issues inherent in monolithic software.
2) Utilize each tool in a fashion it was set out to do in a "best tool for the job" approach from electronic case report forms, laboratory management and genomic analysis visualization.
3) Facilitate the longitudinal collection of clinical phenotypic data allowing for data to be collected in over a user-specified time period.
4) Allow for asynchronous and variable collection of specimens. This modular set of tools allows for one subject to have multiple de-identified specimen collection sets of varied type.
5) De-identify and integrate specific research systems involved in a biorepository studies into a non-human-subject open data repository available through a web query tool.

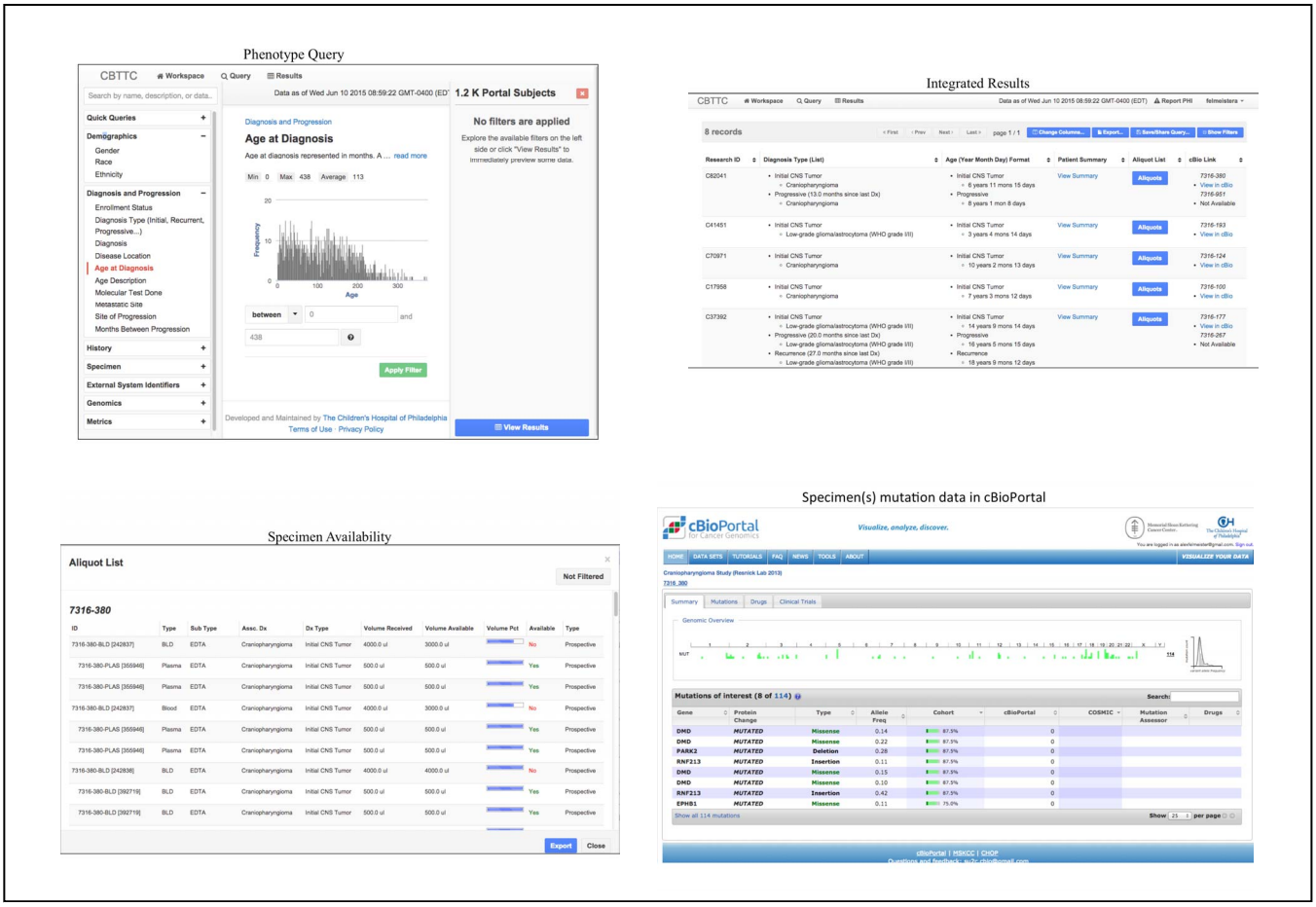

Figure 6. Screen shots of integrated query tool allowing researchers to query across data types; specimen, phenotype, genomics and request physical specimens and quickly gauge new study feasibility based on those results.

6) Future-proof specimens and data to be supplemented with new experimental or relevant information when it becomes available through honest brokering institution specific subject identifiers.

IV. Conclusion and Discussion Points

As July of 2015 there are 23 unique protocols/patient cohorts being managed in the Biorepository Portal (BRP). There are over 3,000 unique subject records in the electronic honest broker (eHB), over 19,000 specimens accessioned and 7 institutions participating in various biobanking activities using this tool kit. Figure 7 is a screenshot from the data and specimen query tool of the Children's Brain and Tumor Tissue Consortium. This ambitious biobanking collaboration between six children's hospitals utilize this set of integrated tools to allow The Children's Hospital of Philadelphia to be the data and specimen-coordinating center (DCC/SCC) for this project. Other projects include a general accrual biobank for the Center for Childhood Cancer Research (CCCR) at CHOP, the Inflammatory Bowel Disease (IBD) Center Tissue and Stool repository, a clinical sequencing project, an Acute Myeloid Leukemia (AML) retrospective biobank for establishing objective AML risk categories for personalized treatments and a neurosurgery tissue collaboration between The University of Pennsylvania and The Children's Hospital of Philadelphia. To date, the toolkit has been the driving process behind a number of successfully funded projects focused on next generation sequencing of biospecimens and the identification of causative mutations underlying tumors. A researcher can look at work established in major childhood brain cancer science and, in short, expand on well-known research similar to references [21], [22] without the need to build a repository before understanding potential study feasibility. Current efforts are focused on the implementations of platforms for integration of the multidimensional genomic data sets generated by such studies, further empowering biorepository-based efforts in ways that have yet to be addressed.

A demonstration of this software is available at http://www.brptoolkit.com. This website also contains documentation, webinars, descriptions and pointers to code repositories in context. Software discussed in this paper is available on the CHOP Department of Biomedical and Health Informatics github repository at https://github.com/chop-dbhi. A specific implementation of the toolkit is available through the Children's Brain Tumor Tissue Consortium at http://www.cbttc.org .

ACKNOWLEDGMENT

The authors would like to thank The Children's Brain Tumor Foundation (CBTF); Tom Curran, PhD; Jena Lilly, MS; Angela Waanders, MD, MPH; David Stokes, PhD; and The Clinical and Translational Research Center at The Children's Hospital of Philadelphia for their tireless effort and contribution to this project and our mission to develop research to improve the care and health of children.

REFERENCES

[1] E. Benz, "Genomics and the Future of Cancer Treatment."

[2] I. Hirtzlin, C. Dubreuil, N. Préaubert, J. Duchier, B. Jansen, J. Simon, P. Lobato De Faria, A. Perez-Lezaun, B. Visser, G. D. Williams, and A. Cambon-Thomsen, "An empirical survey on biobanking of human genetic material and data in six EU countries.," *Eur. J. Hum. Genet.*, vol. 11, no. 6, pp. 475–88, Jun. 2003.

[3] A. Végvári, C. Welinder, H. Lindberg, T. E. Fehniger, and G. Marko-Varga, "Biobank resources for future patient care: developments, principles and concepts.," *J. Clin. Bioinforma.*, vol. 1, no. 1, p. 24, Jan. 2011.

[4] J. Vaught, J. Rogers, T. Carolin, and C. Compton, "Biobankonomics: developing a sustainable business model approach for the formation of a human tissue biobank.," *J. Natl. Cancer Inst. Monogr.*, vol. 2011, no. 42, pp. 24–31, Jan. 2011.

[5] L. Kamm, D. Bogdanov, S. Laur, and J. Vilo, "A new way to protect privacy in large-scale genome-wide association studies.," *Bioinformatics*, vol. 29, no. 7, pp. 886–93, Apr. 2013.

[6] D. Swanson, R. Aamodt, and M. J. Becich, "A multi-disciplinary approach to honest broker services for tissue banks and clinical data: a pragmatic and practical model," vol. 113, no. 7. pp. 1705–1715, 2009.

[7] "HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT OF 1996." 1996.

[8] "The Children's Brain Tumor Tissue Consortium." [Online]. Available: www.cbttc.org.

[9] P. a Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. G. Conde, "Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support.," *J. Biomed. Inform.*, vol. 42, no. 2, pp. 377–81, Apr. 2009.

[10] P. P.-S. Chen, "The entity-relationship model---toward a unified view of data," *ACM Trans. Database Syst.*, vol. 1, no. 1, pp. 9–36, 1976.

[11] J. W. Pennington, B. Ruth, M. J. Italia, J. Miller, S. Wrazien, J. G. Loutrel, E. B. Crenshaw, and P. S. White, "Harvest: an open platform for developing web-based biomedical data discovery and reporting applications.," *J. Am. Med. Inform. Assoc.*, vol. 21, no. 2, pp. 379–83, 2013.

[12] E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander, and N. Schultz, "The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data," *Cancer Discov.*, vol. 2, no. 5, pp. 401–404, 2012.

[13] R. Dhir, A. a. Patel, S. Winters, M. Bisceglia, D. Swanson, R. Aamodt, and M. J. Becich, "A multidisciplinary approach to honest broker services for tissue banks and clinical data: A pragmatic and practical model," *Cancer*, vol. 113, no. 7, pp. 1705–1715, 2008.

[14] J. J. Hutton, *Pediatric Biomedical Informatics*, vol. 2. Dordrecht: Springer Netherlands, 2012.

[15] National Cancer Institute, "NCI Best Practices for Biospecimen Resources," 2010.

[16] A. D. Boyd, P. R. Saxman, D. a Hunscher, K. a Smith, T. D. Morris, M. Kaston, F. Bayoff, B. Rogers, P. Hayes, N. Rajeev, E. Kline-Rogers, K. Eagle, D. Clauw, J. F. Greden, L. a Green, and B. D. Athey, "The University of Michigan Honest Broker: a Web-based service for clinical and translational research and practice.," *J. Am. Med. Inform. Assoc.*, vol. 16, no. 6, pp. 784–91.

[17] A. D. Boyd, D. A. Hunscher, A. J. Kramer, C. Hosner, P. Saxman, B. D. Athey, J. F. Greden, and D. C. Clauw, "The 'Honest Broker' method of integrating interdisciplinary research data.," *AMIA Annu. Symp. Proc.*, p. 902, 2005.

[18] B. Blobel, R. Nordberg, J. M. Davis, and P. Pharow, "Modelling privilege management and access control," *Int. J. Med. Inform.*, vol. 75, pp. 597–623, 2006.

[19] R. Morris and K. Thompson, "Password security: a case history," *Commun. ACM*, vol. 22, no. 11, pp. 594–597, Nov. 1979.

[20] The Office for Civil Rights (OCR) and B. Malin, "Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule," *Heal. Inf. Priv.*, pp. 1–32, 2012.

[21] P. K. Brastianos, A. Taylor-Weiner, P. E. Manley, R. T. Jones, D. Dias-Santagata, A. R. Thorner, M. S. Lawrence, F. J. Rodriguez, L. a Bernardo, L. Schubert, A. Sunkavalli, N. Shillingford, M. L. Calicchio, H. G. W. Lidov, H. Taha, M. Martinez-Lage, M. Santi, P. B. Storm, J. Y. K. Lee, J. N. Palmer, N. D. Adappa, R. M. Scott, I. F. Dunn, E. R. Laws, C. Stewart, K. L. Ligon, M. P. Hoang, P. Van Hummelen, W. C. Hahn, D. N. Louis, A. C. Resnick, M. W. Kieran, G. Getz, and S. Santagata, "Exome sequencing identifies BRAF mutations in papillary craniopharyngiomas.," *Nat. Genet.*, vol. 46, no. 2, pp. 161–5, 2014.

[22] D. W. Parsons, M. Li, X. Zhang, S. Jones, R. J. Leary, J. C.-H. Lin, S. M. Boca, H. Carter, J. Samayoa, C. Bettegowda, G. L. Gallia, G. I. Jallo, Z. a Binder, Y. Nikolsky, J. Hartigan, D. R. Smith, D. S. Gerhard, D. W. Fults, S. VandenBerg, M. S. Berger, S. K. N. Marie, S. M. O. Shinjo, C. Clara, P. C. Phillips, J. E. Minturn, J. a Biegel, A. R. Judkins, A. C. Resnick, P. B. Storm, T. Curran, Y. He, B. A. Rasheed, H. S. Friedman, S. T. Keir, R. McLendon, P. a Northcott, M. D. Taylor, P. C. Burger, G. J. Riggins, R. Karchin, G. Parmigiani, D. D. Bigner, H. Yan, N. Papadopoulos, B. Vogelstein, K. W. Kinzler, and V. E. Velculescu, "The genetic landscape of the childhood cancer medulloblastoma.," *Science*, vol. 331, pp. 435–439, 2011.