

Summary

The National Dialogue on Cancer

The National Dialogue on Cancer (NDC) is a forum that brings together representatives from the private sector, academia, non-profit organizations, and government agencies to accelerate progress against cancer. At a March 2002 meeting of the NDC Research Team, “access to appropriately collected, consented, and annotated tissue” was identified as a critical barrier to developing genomics- and proteomics-based therapies. Following this meeting, the Tissue Access Working Group (TAWG) was formed to address this barrier.

The TAWG first met in August 2002 and concluded that the development of a national tissue resource and data bank was necessary if the nation was to ultimately realize the promise of genomics and proteomics for preventing and curing cancer as well as a range of other diseases. The overall goal of the TAWG is

to establish a national, pre-competitive, regulatory compliant and genetic-privacy protected, standardized, inclusive, highest quality network of biological sample(s) banks; developed in partnerships with and supported by cancer survivors/advocates; shared, readily accessible, and searchable using appropriate informatics systems (e.g., amenable to molecular profiling capability).

Over the last year, a subset of the TAWG, the National Biospecimen Network (NBN) Design Team, has been drafting a strategic plan—the NBN Design and Engineering Blueprint—that identi-

fies the key goals and characteristics of a new NBN. In conjunction with its development of the blueprint, the NBN Design Team recognized that it was important to evaluate existing tissue resources. To assist in its examination of existing tissue resources, the team requested that RAND conduct case studies of existing human tissue resources to evaluate their utility for genomics- and proteomics-based cancer research and that RAND identify “best practices” at these institutions.

This report presents the findings for twelve repositories in the United States that represent a broad spectrum of repository types. Interviews were conducted at each repository with key individuals who were asked questions about repository design; the bioinformatics system; privacy, ethical, and legal issues; and public relations and marketing. The interviews focused on the identification of best practices, including innovative strategies, systems and processes pertaining to specimen and data collection, storage and distribution, bioinformatics systems, and informed consent. This report identifies best practices that can be used by the TAWG in its strategic planning process for the development of a robust resource for genomics- and proteomics-based research that will fulfill the needs of the research community.

Evaluation of Existing Human Tissue Resources

Each of the repositories evaluated for this study was established and designed to meet specific objectives. Thus, each repository’s design is integrally linked to its objectives. Biospecimen collection, processing, and storage techniques vary depending on the purpose of the repository. Likewise, the quality and extent of information collected with the specimens vary depending on the purpose for which the tissue was originally collected. The type of informed consent—whether general surgical consent or specific informed consent for the use of the biospecimen for research purposes—also varies from repository to

repository, sometimes limiting the usefulness of some specimens for certain kinds of research.

While all repositories evaluated for this study have one or more of the features identified by the NBN Design Team as components of a new NBN, none has all of the characteristics identified as necessary for a successful NBN. Some repositories have several of the needed characteristics; others have only a few.

Complicating matters is the fact that there are currently no national standards for tissue repositories that collect and store specimens for research use.¹ Therefore, the way one repository collects, processes, and stores its specimens may be very different from the way another repository does, which may complicate comparisons of research results obtained using biospecimens from different repositories. Furthermore, once samples are distributed to researchers, most repositories do not require those researchers to report their research results back to the repository, and even fewer repositories enter those research results into their bioinformatics systems and make them available to the broader research community.

The NBN Design Team recognized the limitations of existing repositories and decided to design a new kind of repository. The team envisions a network of geographically dispersed tissue repositories to collect, process, store, and distribute appropriately consented diseased and normal tissue and other biological specimens with associated clinical data supported and coordinated by an accessible, user-friendly bioinformatics system networked across the country. The biospecimens would be collected, processed, annotated, stored, and distrib-

¹ Professional societies, such as the International Society for Biological and Environmental Repositories (ISBER) and the National Committee for Clinical Laboratory Standards (NCCLS), have recognized the need for standardization and are developing guidance for establishing and operating biospecimen repositories. ISBER is creating a set of Best Practices for Repositories to provide repository professionals with guidance on repository activities. The NCCLS guidelines will cover all health care institutions and clinics that collect human tissue for research purposes, and will provide standards for addressing all issues associated with the collection of human tissue to support biomedical research, including the ethical, legislative, and legal concerns.

uted in a highly standardized manner to minimize experimental variability and accelerate scientific progress. The NBN would also archive research data submitted by investigators who were using NBN samples and would promote data sharing and meta-analysis.

The network of geographically dispersed tissue repositories that the NBN Design Team envisions for collecting, processing, annotating, storing, and distributing tissue is very similar to how some of the repositories evaluated for this study are set up. The Cooperative Human Tissue Network (CHTN), the Early Detection Research Network (EDRN), the Breast and Ovarian Cancer Family Registries (CFRs) (of which the Philadelphia Familial Breast Cancer Registry is a member), and the University of Pittsburgh Health Sciences Tissue Bank (HSTB) are all variations of the model of decentralized resources deployed through a virtual network of geographically dispersed tissue centers coordinated and supported by a centralized bioinformatics and data management system networked across the country. The National Heart, Lung, and Blood Institute (NHLBI) Biological Specimen Repository, the Armed Forces Institute of Pathology (AFIP) National Pathology Repository, Ardaís Corporation, and Genomics Collaborative, Inc. (GCI) have a decentralized collection model but maintain their storage and distribution of specimens and their bioinformatics system at one physical location. The Tissue Array Research Program (TARP) also has a decentralized collection model with the bioinformatics system and storage maintained at one physical location; however, its tissue microarrays are distributed by the CHTN. In contrast, the Duke University Breast Specialized Program of Research Excellence (SPORE), the Mayo Clinic Prostate SPORE, and the University of Alabama at Birmingham (UAB) Breast and Ovarian SPOREs have centralized collection, storage, distribution, and bioinformatics systems and data management. Based on its evaluation of these twelve existing human tissue resources, RAND came up with several best practices that the NBN Design Team and NDC may want to consider as they implement their plan for the NBN.

Biospecimen Collection, Processing, Annotation, Storage, and Distribution: Best Practices

Best practices for biospecimen collection that will increase the quantity and variety of high-quality samples available to researchers, while maintaining appropriate normal controls, include establishing a network of collection sites at academic medical centers and community hospitals, and collecting tissues from a broad range of diseases, non-diseased matching adjacent tissue, normal tissue, and other biological specimens (e.g., whole blood, serum, and plasma). It is also important that tissue be collected from ethnically diverse populations of all ages to ensure that the tissue available for research purposes is diverse and demographically representative of the population, and to expand biomedical research to include understudied/underrepresented populations and the study of health disparities.

The prioritization of patient diagnosis over collection of specimens for research purposes is key to ensuring that patient care is not compromised and that patients continue to donate biospecimens. Pathologists at the collection site play an important role in the initial procurement of the specimen for the repository, and repository pathologists are central to the quality control procedures for verification and evaluation of the specimen. In addition, repository-trained personnel using standard operating procedures and standard collection and processing equipment are important to promoting standardized tissue collection and processing.

Best practices for data collection depend on the mission of the repository. However, no matter what the requirements for the amount of associated data are, certain best practices are applicable. It is important to collect consistent and high-quality data associated with biospecimens and to employ a standardized set of common data elements that are collected with every biospecimen. It is also important to define the data set that is optimal for fulfilling the mission of the repository and the needs of its customers, and to collect the data (such as demographic and pathologic data, family history, medical history, lifestyle and diet history, treatment history, and clinical outcomes) required to meet those needs.

Once data are collected, they must be entered into the repository's bioinformatics system. The use of common data elements and standardized terminology for data collection procedures allows the use of standardized data-entry forms with features that minimize the errors introduced while typing information into forms. In addition, scannable bar codes are used to track biospecimens and associated information throughout their lifetime at the repository. Parsing techniques are used to flag discrepancies and to record errors and their reconciliation. The use of standardized terminology and computer data entry forms, scannable bar codes, and data reconciliation techniques are best practices that ensure data accuracy.

Standards for storage depend on tissue type and preservation condition (e.g., snap frozen, paraffin embedded, tissue microarray). Snap-frozen specimens are commonly stored at -80°C in mechanical freezers or in liquid nitrogen. Paraffin-embedded tissue and tissue microarrays are stored at room temperature or in a climate-controlled environment to protect them from melting or other damage. However, there is no consensus on the optimum storage conditions for specimens.

Once specimens are placed in storage, it is necessary to monitor storage conditions and maintain equipment. Standard operating procedures for freezer maintenance, adequate backup equipment, and redundancy in storage location are best practices for ensuring that specimens are stored and maintained at the necessary temperature and condition and that specimen integrity is not compromised. Periodic auditing, inventories, and certification of the location, identity, and quality of specimens ensure the quality and integrity of samples sent to researchers. Bar coded inventory systems are used to track specimen storage location.

Standardized and carefully monitored shipping procedures track all shipments in and out of a repository. Biospecimens sent to a repository from remote/satellite collection sites and samples sent from that repository to researchers are tracked using electronic technologies, such as bar coded inventory systems or smart chips and radio-frequency identification tags.

Specimen distribution practices clearly depend on the mission of the repository. If the mission is to provide tissue samples to as broad a base of researchers as possible based on the quality of the proposed research, then biospecimen distribution policies should be established to fulfill this mission. If the mission is clearly defined, and if the repository evaluates its ability to meet its goals and changes its policies, procedures, and practices when not meeting those goals, then this is a best practice.

Quality assurance is fundamental to the successful operation of any biospecimen repository. The use of standardized protocols for collection, storage, processing, and distribution of specimens, and the use of common data elements for the annotation of specimens at each of the individual network participant locations make comparative research across participating institutions possible. To ensure that the collection, processing, annotation, storage, and distribution of biospecimens occur at a consistently high level of quality, it is necessary to have a multitiered, fully integrated quality assurance system and standard operating procedures. Quality assurance starts with the training of personnel before biospecimens are ever collected and includes everything up through considering researcher feedback on sample quality.

Bioinformatics: Best Practices

The backbone of any repository is a standardized, scalable, and secure bioinformatics system that is appropriate for repository management, tissue acquisition and management, and data aggregation and analysis. Bioinformatics systems are used for repository management, clinical and pathological data management, collection and analysis of research results, and data mining and advanced statistical analysis to identify patterns and establish relationships. A bioinformatics system that is searchable and minable via varying levels of Web-based access for different individuals—including repository personnel, researchers, patients, and the public—is a best practice. Robust network security

systems and access control are crucial to ensure that the privacy of the tissue source is protected and that the bioinformatics system is secure.

Bioinformatics systems can range from simple databases to proprietary systems developed in house. Close ties between bioinformatics system developers, researchers, data managers, and repository management allow the bioinformatics system to be designed so that it is responsive to the needs of multiple user types.

The use of a standardized language to categorize and describe biospecimens and enter data into the bioinformatics system is essential for comparison of biospecimen characteristics among collection sites. In addition to using a standardized language, it is also important to use either a system that can automatically extract data from medical records or multiple checks of data entry to ensure the accuracy of the data in the bioinformatics system.

Consumer/User Needs: Best Practices

A repository is successful only if it is meeting user needs, and its success can only be determined through continual self-assessment and re-evaluation. Meeting user needs may require different approaches depending on the repository's design, customer profile, and product offerings. Assessing the needs of researchers, tracking the numbers and types of tissue samples distributed, and using this information to determine whether the resource is continuing to meet researchers' needs or whether changes need to be made constitute a best practice.

The review and prioritization system for tissue distribution generally falls into one of four general categories: (1) first come, first served; (2) priority to members of the network, collaborators, and/or contributors to the repository; (3) prioritization based on merit review of research proposals; or (4) prioritization based on a set policy of the repository. Best practices to ensure equitable distribution of tissue to the broadest group of researchers possible include (a) the use of a tissue utilization committee to prioritize tissue distribution based on merit review of researcher proposals, and (b) policies to control

the distribution of rare specimens, to control the last sample of a particular specimen, and to prevent the control of an entire specimen or type of specimen by one researcher. Giving priority to researchers at collecting institutions is also a best practice, one that leads to more support for the resource and higher investment in the quality of the specimens collected.

Committees or review groups in which both providers and consumers are able to provide input on the usefulness of the repository resources are valuable in evaluating how well the repository meets user needs. In addition, solicitation of feedback on sample quality directly from researchers who are using the samples helps to identify systemic problems, inconsistencies, or problems with the specimens in the repository or specimens being collected in a certain way or from a certain collection site. These best practices enable repositories to improve specimen quality and to be responsive to researcher needs.

Business Plan and Operations: Best Practices

Repositories are generally funded by four different sources: the federal government, academia, private industry, and private non-profit. Repositories also follow different business models, including tissue banking versus prospective collection and distribution, networks versus individual sites, and centralized versus decentralized collection, storage, and bioinformatics systems. Establishing a network of collection sites at academic medical centers and community hospitals to perform a combination of banking to collect and maintain a ready supply of tissue and prospective collection to meet researcher needs is a best practice.

When approaching a medical facility about becoming a participating collection site, it is often more productive to start discussions with the pathologists and surgeons rather than with hospital administrators. Once a relationship has been established at a collection site, it is vital to maintain close working relationships with surgeons,

pathologists, nurses, and other relevant staff at that site through good communication between the repository and collection site.

At the repositories evaluated, the cost per specimen to collect, process, store, and distribute was variable (between \$60 and \$150 per specimen at the repositories able to provide estimates) and depended on the amount of clinical information accompanying the biospecimen—the more information, the higher the cost of collecting the tissue and associated data. The cost of samples to researchers ranged from free of charge to \$200 or higher depending on the type of sample obtained, the level of annotation associated with the sample, and whether the researcher was from an academic institution or industry. Accurately determining the actual costs of collecting, processing, storing, and distributing tissue samples, and operating on a cost recovery basis to financially sustain the repository constitute a best practice.

Continually assessing new technologies and taking measures to develop and incorporate new technologies into the repository are necessary for any system to be forward thinking, capable of expansion, and flexible as researchers' needs change. This is usually accomplished through regular meetings with staff to brainstorm about ways to address and incorporate new technologies, in combination with more formal mechanisms, such as committees or workshops established to purposefully scan for improvements and new technologies. Requiring acknowledgment in publications for the use of repository resources, including specific language for such acknowledgment, is a best practice because it raises the visibility of the resource and may encourage future donations and use of the resource. It also allows the tracking of scientific accomplishments made possible by the availability of specimens from the resource.

Privacy, Ethical Concerns, and Consent Issues: Best Practices

Institutional review boards (IRBs) are responsible for the oversight and review of research that involves human participants to ensure that their privacy is protected and confidentiality of data is maintained. Requiring repositories to have IRB approval for the collection, storage, and distribution of biospecimens and associated data, and requiring researchers requesting samples to have IRB approval of research projects that will use the samples are essential for protecting privacy and confidentiality. In addition, convening a bioethics advisory board or other governance and oversight board/committee to oversee privacy and confidentiality procedures provides another layer of review. Limiting access to the codes that link patient identifying information to the sources of the tissue specimens through physical and/or cyber procedures to minimize the chance of identifying information being released is also a best practice to protect privacy and confidentiality.

Obtaining biospecimens from individuals who are fully informed about and have consented to the collection of their tissue by the repository and its use for research purposes is a best practice. Using a tiered consent process that allows individuals to choose the type of specimen(s), if any, they want to donate (e.g., tissue, blood, or urine), the type of research the specimen can be used for (e.g., a specific research project, general research, or genetic research), and whether their medical records and outcomes data can be accessed is also a best practice. Ideally, the consent process should occur separately from the surgical consent. However, since this is not always possible, at a minimum the informed consent for the collection and research use of specimens should be a separate section of the surgical consent form that requires a separate signature.

Intellectual Property and Legal Issues: Best Practices

Individuals who contribute biospecimens must have the right to withdraw their consent and have their tissue removed from the repository. However, once the tissue has been stripped of identifiers so that the link back to the identity of the tissue source has been destroyed, it is not possible to identify the tissue to withdraw it from the repository. Beyond the right to withdraw their consent and their tissue from the repository, tissue sources are given no other rights to their tissues by most repositories. It is a best practice to allow an individual who contributes tissue to a repository to withdraw consent and have the tissue, data, and computer records removed from the repository if the tissue retains identifiers to link it to that individual and it has not been distributed to researchers.

Most repositories do not retain downstream rights to any intellectual property produced through the use of the tissues they distribute. In most cases, institutions that contribute biospecimens to the repository give up their rights to the biospecimens as well. However, some contributing institutions are given priority for tissue requests. In the interest of being clear and to avoid conflicts with tissue sources, researchers using the tissue, or institutions contributing biospecimens, it is a best practice to use a specific published policy on intellectual property. Another best practice is to prioritize tissue distribution based on need while reserving a small percentage of tissue for contributing institutions participating in the repository.

Requiring researchers to sign an agreement that covers the legal issues associated with the use of biospecimens is a best practice. The tissue use agreement should contain language to the effect that the specimens will be used only for the purposes cited in the application, no attempt to obtain identifying information will be made, no specimens will be sold or shared with a third party without the prior written permission of the repository, all specimens will be treated as potentially infectious, all personnel who will be handling the specimens will be properly trained, there is no implied warranty on the specimens, any publications resulting from the use of repository specimens will acknowledge the repository, and the researcher/institution using

the tissue assumes responsibility for all risks associated with the receipt, handling, storage, and use of the tissue.

The matter of liability with respect to safety issues associated with the use of the specimens, loss of privacy or breach of confidentiality of tissue sources, claims by tissue sources of physical/psychosocial harms, or claims of tissue sources to property rights for discoveries made using their tissues is a major concern to repositories. Therefore, it is important to explicitly specify the responsibility for assuming risks in connection with use of biospecimens, to fully inform tissue sources about risks to their rights and welfare, and to clarify ownership issues in tissue use agreements and during the informed consent process. Similarly, it is a best practice to carefully review researchers' submissions and credentials to ensure that tissue is being used by legitimate researchers for legitimate purposes. This review should include the inspection of IRB documentation, review of the study design for which the samples will be used, and verification that the researcher requesting samples is associated with a legitimate research institution.

Public Relations, Marketing, and Education: Best Practices

Public relations, marketing, and education are critical to the success of any tissue repository. Utilizing a combination of approaches to publicize the resources available at the repository—including exhibits at scientific meetings, advertising in scientific journals, newsletters, Web sites, direct mailings, and word of mouth—is a successful way to increase the visibility of the repository, its resources, and its mission. Although not widely done today, a best practice for repositories is to provide feedback to tissue sources, physicians, and researchers through scientific and patient workshops that report generalized research findings, by disseminating research news and patient education information on a Web site, by sending newsletters to tissue sources and researchers summarizing research with repository resources, or through other outreach venues.

Conclusions

Each of the repositories evaluated in this study was designed according to a specific vision, which was not necessarily the same as the vision of the NBN Design Team. Due to these different visions, none of the repositories in this report exhibits *all* of the elements identified as important by the NBN Design Team for the proposed NBN. However, in most cases the repositories are flexible and, with appropriate funding and guidelines, have the potential to be an integral part of the NBN. In fact, this study revealed that most of the repositories have undergone a significant learning curve and that their current successes are based on years of experience and learning from early operations. This wealth of experience should not be overlooked as NDC goes forward with its plan to establish a new NBN.

All of the repositories evaluated exhibit some characteristics that would be useful for an NBN, but some of the repositories incorporate more of the NBN Design Team requirements than others do. CHTN, University of Pittsburgh HSTB, Ardaís, and GCI have several of the characteristics identified by the NBN Design Team as necessary for a successful NBN. CHTN is a virtual network with the proven ability to distribute tens of thousands of biospecimens in a variety of forms (e.g., fresh, snap frozen, and paraffin embedded) to meet researchers' needs. University of Pittsburgh HSTB has developed a Web-based bioinformatics system that includes proteomics and genomics information and is already being used in Pennsylvania to create a virtual network of repositories. Ardaís and GCI, the two private companies in this study, have streamlined specimen collection, processing, storage, and distribution through specific standard operating procedures, and they both minimize operator and data entry errors through the use of bar-code systems.

Other repositories only have a few of the key components of the proposed NBN. For example, TARP develops and disseminates tissue microarrays for high-throughput screening of multiple tumor tissues (300 to 500 tissues per array). EDRN requires that specimens be collected, processed, and annotated in a standardized manner and that a set of common data elements be collected with each specimen. Phila-

delphia Familial Breast Cancer Registry also uses common data elements, and it routinely collects longitudinal data. The SPOREs at Duke University, the Mayo Clinic, and UAB routinely collect detailed clinical information and longitudinal data.

Whether NDC decides to fulfill the NBN goal by building a brand new repository or by using existing repositories in the development of a national network, learning from the existing repositories will be an important step. This report identifies the best practices at twelve biospecimen repositories in the United States. As the NBN gets under way, more detailed analyses of existing biospecimen repositories and the inclusion of key personnel from existing repositories will be warranted.