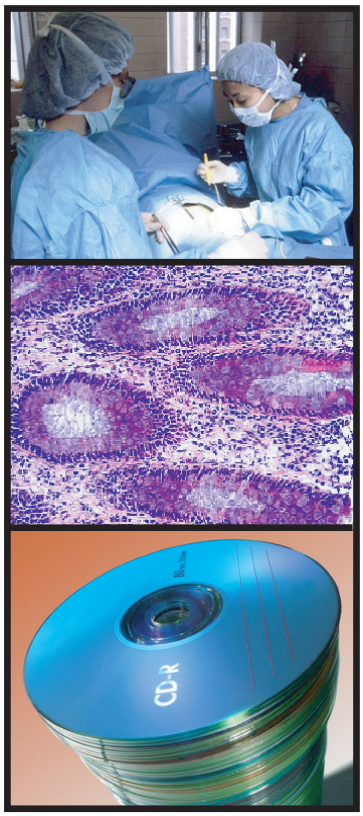

National Biospecimen Network Blueprint

September 2003

The research described in this blueprint was conducted by Constella Group, Inc., design team, and advisors under Contract 263-FQ-310590.

In order to facilitate access to the original publication source, the following form for the citation is suggested:

> National Biospecimen Network Blueprint, Andrew Friede, Ruth Grossman, Rachel Hunt, Rose Maria Li, and Susan Stern, eds. (Constella Group, Inc., Durham, NC, 2003)

Constella Health Sciences serves the public by creating, managing, and sharing knowledge that helps protect and enhance human health. Constella Group is a registered service mark. Support for this blueprint was provided by The National Cancer Institute and the non-profit institution, National Dialogue on Cancer. The views presented in this report are those of the design team and not necessarily those of the funding institutions.

Published 2003 by Constella Group, Inc.
2605 Meridian Parkway, Durham, NC 27713
10401 Fernwood Road, Bethesda, MD 20817
Constella Group URL: http://www.constellagroup.com
To order Constella Group, Inc. documents or to obtain information, contact
Science Management and Research Administration:
Telephone: (301) 571-0240; Fax (301) 571-0264

Foreword

This report represents the collaborative efforts of scientists, clinicians, industry representatives, and patient advocates who participated in the design of a National Biospecimen Network (NBN) Blueprint with the ultimate purpose of accelerating scientific discovery in the battle against cancer. The development of the NBN Blueprint was spearheaded by members of the National Dialogue on Cancer (NDC), which, in addition to researchers, includes representation from the National Cancer Institute (NCI), patient advocacy groups, and the pharmaceutical and medical diagnostics industries. NDC is a nonprofit organization that brings together the principal leaders of key organizations, institutions, and other constituencies from the public, private, and not-for-profit sectors as coequal partners united to eliminate cancer as a major public health problem. The NDC provides a unique platform for defining a common vision to combat cancer.

This document was written by Constella Health Sciences staff under contract to the NDC and NCI, with substantial input provided by the NBN Design Team members, a group of experts representing the NDC Tissue Access Working Group (TAWG). The Design Team was supported by numerous *ad hoc* consultants, including representatives from the United Kingdom National Cancer Tissue Resource, who have recently inaugurated a similar effort and whose strategic plan provided insights that helped to accelerate the development process for the NBN Blueprint. The deliberations were enhanced by site visits to biospecimen resource operations considered to represent state-of-the-art facilities with optimal data access processes. The sponsoring agencies and Constella Health Sciences staff gratefully acknowledge the contributions of these outside experts who have volunteered countless hours reviewing earlier drafts and participating in numerous consultative meetings that have been essential for the development of this Blueprint.

To further inform the Blueprint development process, a concurrent analysis of existing human tissue resources was conducted by the RAND Corporation to identify a set of "best practices" for the collection, processing, storage, annotation, and distribution of biospecimens in the postgenomic era. Through site visits and interviews, the RAND team examined selected repositories from government institutions, academic medical centers, and the private sector to identify innovative strategies in terms of repository design, bioinformatics infrastructure, and policies on informed consent and intellectual property and to determine aspects compatible with the proposed NDC TAWG model. The results of these findings will be available in a separate report that features options and recommendations relevant to the establishment of the NBN.

The NBN Blueprint also has been informed by the results of a voluntary NDC- and NCI-sponsored questionnaire administered at the July 2003 American Association for Cancer Research international conference. Results from this questionnaire provide one snapshot of the composite "portrait" of potential NBN customers. Meeting participants responded to a series of questions about their affiliations, their specific specimen and data requirements, and the value of well-characterized biospecimens such as those proposed by the NBN.

Based on these multiple information-gathering efforts, it appears that the time has come for a centralized and accessible national repository of biospecimens focused specifically on supporting genetic and proteomic research. If it is to be, such a resource will be realized only through the concerted effort of individuals across the spectrum of cancer research, drug and diagnostic development, and patient advocacy. This Blueprint is intended as the first step in an exciting and bold endeavor to develop a large-scale national system of biospecimen and associated data collection and dissemination. The system will support the application of genomic and proteomic technologies for cancer research, to ultimately affect the way cancer and other diseases are treated, thereby improving the quality of life for countless numbers of people.

This Blueprint presents a strategy intended to serve multiple constituents in the cancer research enterprise, and it is being disseminated for broader public comment. It features six modules that focus on various aspects of the NBN. The first two modules, "Why the National Biospecimen Network?" and "Management of Ethical and Legal Issues," articulate the rationale, requirements, and ethical framework for developing a national biospecimen network. The following three modules, "Biospecimen and Data Collection and Distribution," "Bioinformatics and Data Management," and "Communications," describe key operational aspects needed for system implementation. Finally, the "Governance and Business Models" module outlines a recommended management structure for the NBN to promote the participation of a broad range of institutions and existing biospecimen resources. The Blueprint also includes a section entitled, "National Biospecimen Network and Public Health," which describes the potential value of linking a national biospecimen resource to epidemiological investigations. It concludes with a proposal for moving forward, in the form of a "Demonstration Project."

This report and additional related information are available at the NDC Web site (*www.ndoc.org*). The reader also is referred to the NCI Web site (*www.nci.nih.gov*).

Acknowledgments

The National Biospecimen Network (NBN) Blueprint could not have been completed without the generous assistance of more than 100 participants who willingly volunteered their expertise and insights. In addition to those who participated in discussions on the integration of the document at the NBN Design Team and Advisors Meeting in Bethesda, MD (Appendix F), the following experts contributed substantively to this document: Lucille L. Adams-Campbell, Jules Berman, Col. Kenneth Bertram, Marianna Bledsoe, Andrea Califano, José Costa, Martin Ferguson, David Fitzgerald, Michael Katz, David Kerr, Kim Lyerly, Mary McCabe, Mark McGrath, Alan Rabson, Cathy Ratcliffe, and Sheila Taube. Special thanks are due in particular to Roger Aamodt, Alan Buckler, Dan Crichton, William Grizzle, and Julie Kaneshiro for their detailed comments on earlier drafts of the Blueprint. Donna Bonner, Ellen Eisner, Nina Ghanem, Charles Goldwaithe, Nina Goodman, Donald Johnsey, David Korn, Jo-Ann Kriebel, Carl Roth, Karl Schwartz, Sudhir Srivastava, and Jim Vaught also provided helpful input.

Funding for the development and production of this report was provided by the National Dialogue on Cancer (NDC) and the National Cancer Institute (NCI). The support from Anna D. Barker, Connie R. Curran, Allan Erickson, Matt Mangan, Mark Mendenhall, Hal Moses, and Andrew von Eschenbach, affiliates of the NDC, is gratefully acknowledged. Special acknowledgements are extended to Anna D. Barker, Kenneth Buetow, Gregory Downing, Julie Schneider, and Andrew von Eschenbach with the NCI for their extraordinary dedication and expert direction throughout the NBN Blueprint development process, and to Travis Earles and Tracey Nathaniel for their administrative assistance.

Special thanks go to the Constella Health Sciences staff who served as module coordinators and lead science writers: Andrew Friede, Ruth Grossman, Rachel Hunt, Rose Maria Li, and Susan Stern. Other Constella Health Sciences staff who assumed responsibilities for editing, document formatting, graphics, meeting planning, and management are gratefully acknowledged for their contributions: Beverly-Jean Cambridge, Mary Cadette, Patti Collins, Mary Conley, Bruce Dennis, Patricia Downey, Anne Fabiszewski, Kerry Groome, Ann Hennigan, Chris LeGrand, Tracy Lyles, Elisa Madonna, Keith Maynard, Julie McLaughlin, Harry Milman, Bonnie Ortiz, Alecia Saah, Jennifer Solomon, Diane Stidham, Ann Thompson, Thomas Timmons, and Louise Wides.

Finally, we wish to recognize the tremendous efforts of cancer advocacy groups for their invaluable participation in the fight to cure and prevent cancer.

Table of Contents

Foreword i

Acknowledgments iii

Executive Summary vii

Acronyms xiii

1. Why the National Biospecimen Network? 1

2. Management of Ethical and Legal Issues 13

3. Biospecimen and Data Collection and Distribution 29

4. Bioinformatics and Data Management 49

5. Communications 67

6. Governance and Business Models 79

7. National Biospecimen Network and Public Health 99

8. Demonstration Project 103

References 107

Glossary 111

Appendices
- A: Case Study: The Role of Biospecimens for Discovery of a Targeted Cancer Therapy
- B: Case Study: Biospecimens for Validation of a Biological Target and Discovery of Alternative Therapeutic Uses
- C: Case Study: Cancer Tissue Samples Key to Development of High-Precision Genomic Diagnostic Test
- D: Interview Instrument for RAND Evaluation of Selected Existing Tissue Resources
- E: National Cancer Institute and National Dialogue on Cancer Research Team Questionnaire to Assess Biospecimen Needs and Response Frequencies
- F: National Biospecimen Network Blueprint Design Team and Advisors Meeting Participant List: July 28–29, 2003
- G: Examples of Potential Needs and Uses for Biospecimens by Constituency

H: Guidance for Informed Consent for Tissue Repositories
I: The National Cancer Institute Tiered Consent Form
J: Examples of Ethics Oversight Groups
K: Best Practices for Repositories: International Society for Biological and Environmental Repositories
L: Vignette: Life Cycle of a Biospecimen Through the National Biospecimen Network
M: Advanced Analysis Techniques
N: Nonprofit, Government-Funded Example: The Cooperative Human Tissue Network
O: Nonprofit, Eventual Private/Public Funding Example: United Kingdom National Cancer Tissue Resource
P: Nonprofit, No Government Funding Example: The SNP Consortium
Q: For-Profit, Minimal Government Funding Example: Ardais Corporation
R: For-Profit, Primarily Private Funding Example: First Genetic Trust
S: Private/Public Funding Example: International Genomics Consortium Expression Project for Oncology (exp*O*)
T: Private/Public Funding Example: Organ Procurement and Transplantation Network

Executive Summary

Why the National Biospecimen Network?

Recent advances in the fields of genomics and proteomics are providing new ways to derive more valuable data for cancer research from human biospecimens. Advances in informatics can now support the compilation and analysis of genetic and clinical data on an unprecedented scale. Applications of genomic and informatic technologies to interrogate biospecimens represent unparalleled opportunities for the discovery and development of new cancer diagnostic, therapeutic, and preventive agents. For scientists to take full advantage of the power of new informatics tools to compare results across multiple investigations requires genetic and clinical data to be available in a standardized format. In brief, to accelerate discovery for cancer research, scientists need timely and equitable access to biospecimens that are annotated with clinical information and collected specifically for genomic and proteomic studies, as well as a platform for comparing analyses with the results of other researchers.

At the same time, the advanced technical capabilities that enable the linkage of genetic data with clinical information have raised ethical, legal, and social concerns. Individual contributors of biospecimens and data for clinical research must feel confident that the privacy of their medical information will be honored by the research community.

The National Biospecimen Network (NBN) provides a key infrastructure to harness the potential of new technologies for cancer research, while ensuring that the privacy interests of biospecimen donors are preserved. It creates a comprehensive framework for sharing and comparing research results through a robust, flexible, scalable, and secure bioinformatics system that supports the collection, processing, storage, annotation, and distribution of biospecimens and data using standard operating procedures based on best practices. This combination of characteristics is vital to fully support emerging scientific opportunities to accelerate progress in prevention, diagnosis, and treatment of cancer.

The NBN concept represents an innovative approach to provide biological materials and clinical information to advance translational research. Although many existing tissue repositories in the United States collect and store millions of specimens for many types of scientific investigations, the contents of these repositories are frequently collected and stored under varying conditions, making it difficult for scientists to compare or pool genomic and proteomic results from biospecimens across institutions. Because many of these samples were initially collected for a broad range of uses, the amount of clinical information associated with these biospecimens also varies widely, and is rarely detailed or longitudinal. In addition, many of these biospecimens may not have appropriate donor authorization for genetic studies. Finally, current access to existing specimens is often uneven, further impeding scientific progress.

The heterogeneity among existing repositories poses a challenge to support genomic and proteomic research, particularly concerning the ability to conduct and compare large numbers of biospecimens to capitalize on genomic and proteomic technologies. The NBN model is designed

to standardize resources to overcome these obstacles. However, it should also build on existing resources to function efficiently and cost effectively.

With the recognition that a national tissue resource, although ambitious, is necessary to realize the promise of genomics and proteomics for the prevention and cure of cancer and other diseases, the National Dialogue on Cancer Tissue Access Working Group, in collaboration with the National Cancer Institute, commissioned a Design and Engineering Blueprint for an NBN (the "NBN Blueprint"), with the following goal:

> ***"to establish a national, pre-competitive, regulatory compliant and genetic-privacy protected, standardized, inclusive, highest quality network of biological sample(s) banks; supported by and developed via novel financial and other partnerships with cancer survivors and advocates, the private sector and non-profit organizations as appropriate; that is shared, readily accessible, and searchable using state-of-the-art informatics systems (e.g., amenable to molecular profiling capability)."***

The Design Team outlined essential requirements of the NBN and made specific recommendations. Taken together, these requirements and recommendations provide a framework for realizing the vision of the NBN to be the first nationwide, standardized biospecimen resource designed to facilitate genomic and proteomic research. They are summarized below.

Management of Ethical and Legal Considerations

The relationship between the patient and the NBN is critical, since the patient is the potential donor of specimens and, as such, is the mainstay of the NBN. The NBN should promote a conceptual "chain of trust," with links in the chain that begin with the patient and include the institutional review boards, those responsible for the collection and storage of tissues and data, and researchers who use the resource. Each link will be entrusted with the responsibility to ensure privacy, safety, and compliance not only with applicable rules and regulations, but also with the wishes of donors. The strength of the entire chain is the assurance that the NBN can give to potential donors that responsible use will be made of their biospecimens and associated data.

The NBN informed consent process must ensure that potential participants understand how their specimens may be used, and should use tiered consent procedures that provide individuals with options for levels of participation. In general, donors will not be apprised of specific research results derived from their particular specimens, but information about general scientific discoveries made through the use of biospecimens will be publicly available.

Specimen allocation should be equitable and ensure appropriate use by qualified researchers. The NBN should create a system whereby the NBN has permission to use, and in turn gives permission to use, all biospecimens and associated data. The NBN should avoid asserting any reach-through rights to intellectual property developed through researcher use of tissues and associated data, as attaching such rights may hinder access by certain users and slow research progress.

Biospecimen and Data Collection and Distribution

An overriding principle for the NBN must be that biospecimens for banking (collected for storage in and distribution by the repository) are obtained only after all patient diagnostic needs have been met, and subject to appropriate bioethical structures and procedures to ensure patient protection. The NBN would be distinguished from existing resources for tumor tissue and other specimens by highly standardized procedures for collection, processing, storage, annotation, and distribution. The NBN would be developed to provide biospecimens and clinical information in compliance with Federal, state, and local regulations.

The NBN should have a comprehensive representation of a broad diversity of disease and human populations. Biospecimen donors should therefore reflect the broad range of ethnicities, socioeconomic groups, and other demographic subgroups in the United States. The NBN should pursue the selection of collection sites that will increase the genetic and geographic diversity of its biospecimens.

In particular, the Design Team made the following key recommendations:

> **Recommendation 1.** The NBN should be organized as (a) a decentralized network of collection facilities with regional storage, possibly of nonprofit, tissue-repository organizations located near academic medical centers and community-based hospitals that serve large and diverse patient populations, and (b) as a virtual data repository networked across the nation.

Incentives tailored to each kind of source should be developed to encourage many entities to participate. For example, community hospitals, which could conceivably provide the largest volume of specimens for the NBN, would need incentives and assistance to develop the experience, infrastructure, and understanding of research necessary to establish viable collection centers.

> **Recommendation 2.** Specimens from all cancer types should be collected (with matched normal specimens, whenever possible), but the NBN should be structured to provide the quantity and diversity of biospecimens required to meet researcher needs.

The repository should consist of high-quality biospecimens appropriate for genomic and proteomic studies, and the type of biospecimens stored in the repository should be determined by an ongoing review of researcher needs.

Best practices should be incorporated and/or developed for every aspect of biospecimen and data collection, processing, storage, and distribution in the operation of the tissue repository, and should be consistently applied through the use of standard operating procedures that would be monitored. Biospecimens and data should be collected from sources meeting NBN criteria, while applying standardized clinical annotation. A minimal dataset would be established for each specimen, with collection of additional longitudinal data for a high percentage of specimens, and provision of genomic- and proteomic-based data. The annotation of clinical and pathological data about the biospecimens would be quality controlled and standardized across collection sites. It is recognized that the costs for these associated data are likely to be substantial, and success in obtaining these data will require innovative solutions.

In the NBN, distribution of specimens would be guided by a Biospecimen Utilization Review Committee using a peer review process that would evaluate researchers' needs against competing demands for specimens.

Recommendation 3. It should be expected that validated, investigator-derived data using NBN resources be submitted to the NBN and linked back to original NBN tissue samples.

An expanded dataset, created by the return of this experimental data to the NBN, could be made available to all investigators. It is recognized that this function will take substantial time and effort to establish, and the NBN will have to create incentives to encourage researchers in the public and private sectors to submit research data.

Bioinformatics and Data Management

The NBN should support open research access for the scientific community and enable searches of the databases via the Internet. To support the enhanced research capabilities of the NBN, bioinformatics platforms should incorporate computational analysis tools. At the same time, it should protect privacy and confidentiality of biospecimen donors. The NBN informatics strategy should include, whenever possible, existing vocabularies and common data elements. The architecture should have the ability to scale as the volume of data increases, and to adapt as new datasets emerge. To the extent possible, the NBN data architecture would build upon best practices from existing repositories and other genomic-based research resources. Development and implementation will start with a centralized database, and then move to a more decentralized, yet interconnected model as the system matures, capabilities are strengthened, and requirements clarified.

Recommendation 4. While recommendations about the general architecture and common data elements require broad input, to maximize efficiency, a central architect should be designated to build and manage the bioinformatics infrastructure.

It is proposed that this central architect have authority over project personnel, budget, design, and system architecture, and be accountable to the governing principles of the NBN. Critical benchmarks of success for the NBN data system include ease of data entry and retrieval, highly responsive user support, and commitment to NBN Quality Assurance (see also *Governance and Business Models,* below).

Recommendation 5. The NBN bioinformatics system should be standards based (e.g., Systemized Nomenclature of Medicine, Health Level Seven, or Minimum Information About a Microarray Experiment for Data; Internet for communications) to enable data and information exchange among system components and the researchers who use them.

In designing the bioinformatics system, a standards-based approach will allow flexibility to employ individualized approaches, while reducing the difficulty involved in developing a comprehensive system that links diverse components of the NBN. The NBN should not impose specific requirements for databases or hardware systems by the operational units.

Communications

The NBN must have the support of its constituents—patients/potential donors, researchers, clinicians, academic institutions, hospitals, and commercial interests. This support will require a

well-developed, comprehensive communications plan initiated at the earliest stage of the NBN project and sustained from development through implementation. Through an extensive program of education and training, outreach, and public relations, the NBN must publicize its purpose, availability, and accomplishments. The information campaign will help to manage expectations and provide information about NBN policies that will encourage widespread confidence and trust in the new system.

Recommendation 6. Effective communications among stakeholders should be considered a high priority for the NBN.

Communications planning should begin as early as possible and should be monitored at the operations level (see also *Governance and Business Models,* below). Communications should be broad and comprehensive to meet a variety of communications needs. It should be directed toward diverse stakeholders, and should be structured to ensure consistency of the NBN message throughout the planning, development, and implementation of the NBN program.

Recommendation 7. The NBN should employ an evidence-based model in planning NBN communications.

Under this model, planners identify the various audiences and determine their particular needs. Based on that information, the planners design a strategy to tailor messages to each specific audience, using appropriate media and effective channels of delivery. The conduct of a broader survey of the potential user population should be a priority to ascertain a more detailed and accurate picture about potential research usage trends, need for additional services, cost sensitivities for tissues and data, and advanced analysis services.

Recommendation 8. The NBN should design a communications strategy that clarifies the NBN's role, sets realistic expectations among its different constituencies, and encourages participation by patients, clinicians, and researchers.

This communication strategy should develop ways to encourage participation by donors, particularly from underrepresented groups, and to ensure long-term commitment by all groups for the collection of high-quality longitudinal data. The NBN should develop model methods for obtaining informed consent from patients, and model materials (including forms) explaining the consent process. The NBN must also place a strong emphasis on encouraging clinicians to participate in the system, and on encouraging researchers to use the resource as much as possible.

Governance and Business Models

The NBN governance and business models must include strategies that permit the NBN to rapidly establish a firm foundation and facilitate its growth into a self-sustaining organization. The NBN structure must be highly flexible and capable of expansion as the science progresses. The management strategies must be sufficiently defined to attain the organization's goals, yet flexible to meet the likely evolution of research opportunities and needs.

The Design Team recognized that a sound business plan is needed to effectively operationalize the NBN and partnerships with existing institutions.

Recommendation 9. The NBN Operations Center should be a not-for-profit organization.

The Design Team suggested that a non-for-profit organization provide stewardship of the NBN because it will collect detailed clinical and genetic information about biospecimen donors. In addition to enhancing credibility and maximizing public trust, a not-for-profit model would allow substantial flexibility in accepting funds from public and private sources. The Foundation for the NIH, or another not-for-profit organization, could serve as an incubator for the NBN until the resource could become self-sustaining.

Recommendation 10. The NBN governance should be organized at three levels, to include a Board of Governors, the NBN Operations Center, and its business units. The proposed business units would include the following: Research Administration and Support, Specimen and Data Acquisition, Storage and Distribution/Basic Analysis, Advanced Analysis, Bioinformatics and Data Management, and Patient Relations.

Quality Assurance, Bioinformatics and Data Management, and Communications are core efforts that would be prominent at the NBN operations level. Bioinformatics and Data Management and Communications will also be represented in the business units. Business unit work (their own governance structures permitting) can be carried out by any private or public entity—including existing businesses or research consortia—through a competitive process.

Access to biospecimens should be administered by a neutral, streamlined, peer review system facilitated by a Biospecimen Utilization Review Committee that provides timely review of requests and distribution of samples with minimal administrative burden. Data and samples should be distributed in a clearly articulated and equitable fashion based primarily upon the quality of the proposed research. The Biospecimen Utilization Review Committee would represent a component of the Research Administration and Support Business Unit, but would be monitored at the Operations Center level as part of the quality assurance process.

Demonstration Project

The Design Team recommended initiating the NBN as a demonstration project to test key aspects of the system. A 3-year demonstration project would provide sufficient time to enable the stakeholders to establish and assess the needs of researchers; create an ethical and legal framework; collect, analyze, and distribute biospecimens; and manage the information to suit the needs of the research community.

Acronyms

AACR	American Association for Cancer Research
AAMC	Association of American Medical Colleges
ABL	Abelson Murine Leukemia viral oncogene homologue
ACS	American Cancer Society American College of Surgeons
AFIP	Armed Forces Institute of Pathology
AJCC	American Joint Committee on Cancer
AMA	American Medical Association
AMIA	American Medical Informatics Association
APC	adenomatous polyposis coli
APHA	American Public Health Association
API	Application Programming Interfaces
BCCA	British Columbia Cancer Agency
BCR	breakpoint cluster region
BIGR	Biomaterials and Information for Genomic Research™ (Ardais Corporation)
BIO	Biotechnology Industry Organization
BPH	benign prostatic hyperplasia
BRCA	breast cancer
BU	business unit
caBIO	cancer Bioinformatics Infrastructure "Objects"
caDSR	cancer Data Standards Repository
CALGB	Cancer and Leukemia Group B
caLIMS	Cancer laboratory information management system
CAP	College of American Pathologists
CCOP	Community Clinical Oncology Program
CCSS	Childhood Cancer Survivor Study
CDC	Centers for Disease Control and Prevention
CDMRP	Congressionally Directed Medical Research Program
cDNA	complementary DNA (deoxyribonucleic acid)
CFR	Code of Federal Regulations
CGH	comparative genomic hybridization
CHTN	Cooperative Human Tissue Network
CIRB	Central Institutional Review Board
CLIA	Clinical Laboratories Improvement Act
CMAP	Cancer Molecular Analysis Project
CML	chronic myeloid leukemia
CMS	Centers for Medicare and Medicaid Services (formerly HCFA or Health Care Financing Administration)
CoC	Commission on Cancer
COG	Children's Oncology Group

CRF	Case Report Form
DCP	Division of Cancer Prevention (NCI)
DH	Department of Health (UK)
DoD	Department of Defense
DOT	Department of Transportation
DUMC	Duke University Medical Center
EC	European Commission
ECOG	Eastern Cooperative Oncology Group
EDRN	Early Detection Research Network
EOC	Ethics Oversight Committee
ESC	Executive Steering Committee
EU	European Union
FDA	Food and Drug Administration
FGT	First Genetic Trust
FISH	fluorescence *in situ* hybridization
FNA	fine needle aspiration
FNIH	Foundation for the National Institutes of Health
GCI	Genomics Collaborative, Inc.
GCP	Good Clinical Practices
GIST	gastrointestinal stromal tumors
GLP	Good Laboratory Practices
GMP	Good Manufacturing Practices
GOG	Gynecologic Oncology Group
H&E	hematoxylin and eosin
HCFA	Heath Care Finance Administration (now CMS)
HHS	Department of Health and Human Services
HIMSS	Healthcare Information and Management Systems Society
HIPAA	Health Insurance Portability and Accountability Act
HL7	Health Level Seven
HMO	health maintenance organization
HRSA	Health Resources and Services Administration
IATA	International Air Transport Association
ICD	International Classification of Diseases
ICF	Informed Consent Form
IGC	International Genomics Consortium
IHC	immunohistochemistry
IMAGE	Integrated Molecular Analysis of Genomes and their Expression
IOM	Institute of Medicine
IP	intellectual property

IRB	institutional review board
ISBER	International Society for Biological and Environmental Repositories
ISO	International Standards Organization
IT	information technology
IWHS	Iowa Women's Health Study
J2EE	Java 2 Enterprise Edition
LCM	laser capture microdissection
LIMS	Laboratory Information Management System
LOINC	Logical Observation Identifiers Names and Codes®
MIAME	Minimum Information About a Microarray Experiment
MRC	Medical Research Council (UK)
MTHFR	5,10 methylenetetrahydrofolate reductase
NAPBC	National Action Plan for Breast Cancer
NBAC	National Bioethics Advisory Commission
NBN	National Biospecimen Network (or "the network")
NCCLS	National Committee for Clinical Laboratory Standards
NCDB	National Cancer Database
NCI	National Cancer Institute
NCICB	National Cancer Institute Center for Bioinformatics
NCRI	National Cancer Research Institute (UK)
NCRN	National Cancer Research Network
NCTR	National Cancer Tissue Resource
NDC	National Dialogue on Cancer
NDI	National Death Index
NHS	National Health System (UK)
NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
NLM	National Library of Medicine
NOTA	National Organ Transplant Act
NSABP	National Surgical Adjuvant Breast and Bowel Project
NTRAC	National Translational Cancer Research Network (UK)
OCT	optimal cutting temperature
OHRP	Office for Human Research Protections (formerly OPRR)
OMG	Object Management Group
OPO	Organ Procurement Organization
OPRR	Office for Protection from Research Risk (replaced by OHRP)
OPTN	Organ Procurement and Transplantation Network
OR	operating room
OSHA	Occupational Safety and Health Administration

OSI	Open System Interconnection
PCR	polymerase chain reaction
PHI	protected health information
PhRMA	Pharmaceutical Research and Manufacturers of America
PHS	Public Health Service (U.S.)
PRC	Processing Resource Centres
PRIMR	Public Responsibility in Medicine and Research
PSA	prostate-specific antigen
PXE	pseudoxanthoma elasticum
QA	quality assurance
QC	quality control
RFP	request for proposals
RT-PCR	reverse transcription-polymerase chain reaction
SEER	Surveillance, Epidemiology, and End Results
SNOMED	Systemized Nomenclature of Medicine
SNOMED CT	Systemized Nomenclature of Medicine – Clinical Terms
SNP	single nucleotide polymorphism
SOP	standard operating procedure
SPIN	Shared Pathology Informatics Network
SPORE	Specialized Program of Research Excellence
TARC	Tissue Acquisition Resource Centres (UK)
TARP	Tissue Array Research Program
TAWG	Tissue Access Working Group
TDU	terminal ductal unit
TK	tyrosine kinase
TMA	tissue microarray
TTR	Tumor Tissue Repository
TURPS	transurethral resection of the prostate
UK	United Kingdom
UML	Unified Modeling Language
UNOS	United Network for Organ Sharing
USAMRMC	U.S. Army Medical Research and Materiel Command
VA	Department of Veterans Affairs

Module 1

Why the National Biospecimen Network?

1.

Why the National Biospecimen Network?

This module presents the rationale for the National Biospecimen Network (NBN) system, describes how it is expected to differ from extant tissue resources, and identifies the resource requirements and challenges associated with the development of an NBN that will be of maximum utility to researchers. This module focuses primarily on researcher needs, and seeks to provide detailed requirements for key components of the NBN. Recommendations for implementation are discussed in subsequent modules.

1.1 Background

Participants in the groundbreaking National Dialogue on Cancer (NDC) Research Team Forum I held in March 2002 identified *access to appropriately collected and annotated tissue* as a critical need for fully capitalizing on new genomic and proteomic technologies to accelerate progress against cancer.[1] The group identified the lack of such access as one of the major barriers to realizing the promise of developing targeted cancer diagnostics, preventives, and therapies. The opinions expressed at the Forum I meeting echoed previous public comments of prominent researchers, and the conclusions of several advisory committee reports to the NCI in a variety of cancer research areas, including brain tumor, leukemia, lymphoma, myeloma, lung cancer, and gynecologic cancers.[2]

Several recent examples illustrate how access to relatively large numbers of tissue samples has played a pivotal role in oncology drug development, and these examples underscore the likely benefits from implementing a standardized system by which researchers and clinicians may gain access to biospecimens and associated data. The development of trastuzumab (Herceptin®) is a success story that demonstrates the potential of biomarkers in the rational design and development of cancer drugs that could not have been realized without access to tissue samples. The clinical benefits of trastuzumab would almost certainly have been insufficient for FDA approval if the agent had been tested in unselected patient populations (Appendix A). The Gleevec® story demonstrates how alternative uses for a drug can be discovered through investigations conducted with tissue samples. Gleevec® originally was developed for the treatment of chronic myeloid leukemia. However, screening of tissue samples for c-kit activation identified gastrointestinal stromal tumor patients as potential clinical benefactors (see Appendix B). Finally, as researchers unravel links between molecular pathways and specific cancers and treatments, thereby discovering yet untold uses for many existing therapies, standardization of tissue collection and analysis will become increasingly important for linking various independent observations. Tissue samples played a pivotal role in the development of laser capture

[1] For more information about the NDC, see: *www.ndoc.org/issue_teams_c_research.html*

[2] See the *Report of the Brain Tumor Progress Review Group* (2000), p. 26; the *Report of the Leukemia, Lymphoma, and Myeloma Progress Review Group* (2001), pp. 47-48; the *Report of the Lung Cancer Progress Review Group* (2001), pp. 10, 32-33; and the *Report of the Gynecologic Cancers Progress Review Group* (2001), pp. 4-5.

microdissection, a breakthrough technique that facilitates the precise, reproducible, and accurate transfer of tissues for analysis (see Appendix C).

In response to the challenge articulated by the NDC Research Team—a self-selected group of individuals involved in cancer research, drug development, delivery, and commercialization, as well as representatives from patient advocacy organizations—met in Washington, D.C. on August 26-27, 2002, on January 7, 2003, and again as part of the NDC Forum II meeting on March 5-7, 2003, to further investigate the barriers involved in tissue access and to explore possible avenues for improvement. This group, the NDC Tissue Access Working Group (TAWG), sought to design an approach that could meet the Nation's research needs for biological specimens, and to present options for moving forward. It was at the Forum II meeting that Cathy Ratcliffe from the National Translational Cancer Research Network (NTRAC) presented the United Kingdom (UK) experience with the development of their National Cancer Tissue Resource (NCTR), and TAWG members were provided a copy of the NCTR strategic plan.[3] The UK experience effectively accelerated the United States. Blueprint development process by several months.

During their deliberations, the NDC TAWG members reinforced the conclusion that the development of a national tissue resource, although ambitious, is necessary to realize the promise of genomics and proteomics for the prevention and cure of cancer and other diseases. Unparalleled advances in dissecting the genetic changes and molecular mechanisms that ultimately produce cancer have provided, for the first time, compelling reasons to pursue target-specific interventions. It is now well-recognized that there is a high degree of disease heterogeneity, that sample characteristics and preparation will impact results, and that large samples are required for robust design and rigorous conclusions. To reflect this understanding, the NDC TAWG defined its goal as:

> ***"to establish a national, pre-competitive, regulatory compliant and genetic-privacy protected, standardized, inclusive, highest quality network of biological sample(s) banks; supported by and developed via novel financial and other partnerships with cancer survivors and advocates, the private sector and non-profit organizations as appropriate; that is shared, readily accessible, and searchable using state-of-the-art informatics systems (e.g., amenable to molecular profiling capability)."***[4]

Building on the ideas and vision shaped by the NDC TAWG, an effort was initiated to create a Design and Engineering Blueprint for an NBN (the "NBN Blueprint"), a biospecimen resource envisioned for optimizing and accelerating the development of new interventions for cancer. The major goals of the NBN Design Team were to articulate the unmet needs that the NBN seeks to address, to clarify the NBN customer base, including the role of patients and advocates as well as commercial interests, and to describe the desired processes that the NBN would engage in to meet its goals. The Design Team met in Bethesda, MD on May 28, 2003, to achieve a collective

[3] Knox K. and Ratcliffe C. (2002). *A Strategic Framework for Establishing a National Cancer Tissue Resource for Cancer Biology and Treatment Development.* UK: National Translational Cancer Research Network Coordinating Centre (September).
[4] TAWG Meeting Summary, August 26-27, 2003, p. 4. http://www.ndoc.org/TASummMtgSumAB.pdf

understanding of the purpose of the NBN Blueprint document and the process for its development, to agree on the objectives, key questions, issues, concerns, and types of recommendations to be addressed in each module, and to agree on a general framework for completing the NBN Blueprint. The May 28 meeting also gave participants an opportunity to discuss the challenges (including institutional and other barriers) as well as opportunities to integrate with and adopt best practices from existing systems in the United States and from the UK's NCTR model.

The NBN Design Team's deliberations were enhanced by site visits in May 2003 to tissue resource operations considered to provide state-of-the-art facilities with optimal data access processes as well as an intense period of conference calls during the months of June-August involving the Design Team members and advisors for each particular module topic. The NBN Design Team also benefited from the RAND evaluation of selected existing tissue resources: an exercise that included description of the types of tissue users and distribution practices, and queries to ongoing tissue resource managers of any unmet needs or quality control (QC) issues (see Appendix D for interview instrument). A questionnaire was administered at the American Association for Cancer Research meeting in Washington, D.C. in July 2003 that collected information from meeting participants about their anticipated uses for and reactions to the development of the NBN, as well as their willingness to pay (see Appendix E). This information also assisted the Design Team in their deliberations.

On July 28-29, 2003, the NBN Design Team and invited experts convened to discuss overarching issues, integration of the modules, and final recommendations that should appear in the Blueprint. (See Appendix F for the list of participants to the July 28-29 NBN Blueprint meeting.) Earlier versions of this Blueprint have undergone extensive review by outside experts selected by Constella Health Sciences in consultation with the sponsors, to help ensure the accuracy and relevancy of information provided in this report and to capture the broadest representation possible from diverse viewpoints.

The Design Team identified five primary areas where the NBN could bring value to researchers: (1) Standardized collection of large numbers of fresh/frozen cancer specimens; (2) Accurate, highly standardized clinical annotation and associated data; (3) Prompt and equitable specimen accessibility; (4) Informatics platforms to facilitate sharing of data and results; and (5) Protection of patient privacy. Items one and two specifically address the variation in collection procedures and annotation that currently inhibits uniform comparisons among tissue collections, and items three and four address the frustration expressed by many researchers and advocacy groups about the lack of biospecimen resource sharing.[5]

Additionally, the Design Team envisioned the NBN as serving a number of well-defined niches for researchers. Full implementation of the NBN might well see centralized, advanced analyses of a subset of specimens, as well as the collection of standardized longitudinal data for a high percentage of specimens. Not all researchers will have the same needs; thus, some subsets of samples will be accompanied by more extensive longitudinal data, while other subsets will have

[5] Zitner A. (2003). Whose DNA is it, anyway? Many people, hoping for medical advances, give genetic material. But some researchers' refusal to share samples has donors up in arms. *Los Angeles Times* (July 18).

undergone one or more types of advanced analyses; combinations of these subsets also will exist. However, all specimens should be characterized by the major areas of value outlined above.

1.2 Purpose of the National Biospecimen Network

The NBN is envisioned to be the first national, standardized tissue resource in the United States designed to facilitate genomic and proteomic research, with open access to cancer researchers across the country. Although several countries—including the UK[6], Iceland,[7] and Japan[8]—are investing in nationally coordinated specimen collection, banking, and dissemination systems specifically designed to support genomic and proteomic research, no effort has been attempted on a comparable scale in the United States.

The NBN can facilitate a range of scientific activities that could lead to new genomic- and proteomic-based interventions for cancer, including target identification and validation, the development of new biomarkers and diagnostics, and pharmacogenomic analyses. Recent breakthroughs in the biomedical sciences have produced a wealth of new knowledge about the diagnosis, treatment, and prevention of cancer. Biospecimens, which historically have been a key element of cancer research,[9] have now assumed a central position in the application of genomic technologies to new interventions for cancer. As researchers unravel the roles of particular

[6] The UK National Cancer Tissue Resource, launched by the National Cancer Research Institute, will receive approximately $8 million over 5 years from three primary sources: The Department of Health; the charity, Cancer Research UK; and the Medical Research Council (Spinney L. [2003]. UK launches tumor bank to match maligned Biobank. *Nature Medicine*, Vol. 9, No. 5 [May]: 491.). Another tissue specimen collection effort, the European Human Frozen Tissue Bank (TuBaFrost), seeks to provide high-quality frozen tumor tissue accompanied by a solid diagnosis. The system will have a centralized database containing representative histologic images and code-linked patient data. The information will be accessible in a restricted public domain but freely available to eligible European researchers (See: *www.tubafrost.org*). The UK is supporting a genetic database to help study gene-environment interaction in common diseases. UK Biobank plans to start recruitment in 2004 of up to half a million participants between the ages of 45 and 69 who will be asked to contribute a blood sample, lifestyle details, and medical histories to create a national database of unprecedented size to study genes, environment, and health. The Wellcome Trust, the Medical Research Council, and the Department of Health have committed an initial $73 million over 10 years for the UK Biobank project (Medical Research Council. [2002]. The UK Biobank study gets funding go-ahead. January – June 2002 news archive. *www.mrc.ac.uk/txt/index/public-interest/ public-news-4/public-news_archive/public-news_archive_1_2002/public-biobank_uk.htm*). Biobank Japan was launched in 2003, with the explicit goal of moving earlier investments in pharmacogenomics research closer to clinical applications.

[7] The Icelandic Health Sector Database cost between $135 million and $250 million to build. Under the 1998 law establishing the health record database, DeCode (the exclusive licensee) does not need individual consent for use of private medical data, but the database must meet security and privacy standards set by the government's Data Protection Commission. DeCode has collected disease data and DNA samples, with full consent, from 80,000 Icelanders—or close to one-third of the population (McCaffrey P. [2003]. Iceland's database tussle. *Bio-IT World* [April 1] or: *www.bio-itworld.com/news/040103_report2255.html*).

[8] Biobank Japan aims to create a large-scale DNA repository, with blood samples from some 300,000 individuals, which will be linked to a database containing clinical information. Although it appears that many of the details remain to be worked out, the Japanese government has committed about $180 million over a 5-year period (Triendl R. [2003]. Japan launches controversial Biobank project. *Nature Medicine*, Vol. 9, No. 8 [August]: 982). See also: Asian Technology Information Program (2003). *ATIP03.042: The Biobank Japan Project*, July 23.

[9] National Bioethics Advisory Commission (1999). *Research Involving Human Biological Materials: Ethical Issues and Policy Guidance, Volume I: Report and Recommendations of the National Bioethics Advisory Commission*. Rockville, MD (August): 19.

biomarkers and cellular pathways in specific cancers, biospecimens will help link observations from the laboratory to disease processes observed in the physiological setting.

As shown in Figure 1-1, the NBN is designed to provide a link between epidemiological investigations that identify genetic and environmental risk factors and clinical trials that directly test new interventions for cancer. Several existing efforts fall within the scope of the sectors shown in Figure 1-1. For example, the UK Biobank, Biobank Japan, and DeCode Genetics in Iceland support large-scale genetic epidemiology efforts, whereas the U.S. Cooperative Oncology Groups collect many biospecimens in the context of clinical trials.[10] A recent initiative to develop an NCTR in the UK shares several key elements with the NBN concept. As discussed further in an accompanying report on "best practices,"[11] several existing repositories include individual aspects of the NBN vision; however, the NBN would uniquely integrate a specific combination of features needed to translate basic genomic and proteomic research into clinical discoveries for cancer patients in the United States.

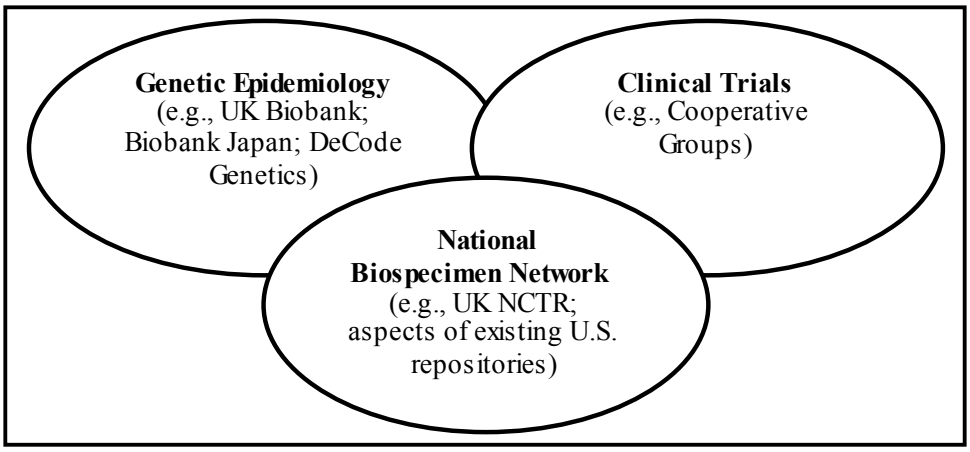

Figure 1-1: The Relationship of the NBN to Other Biospecimen Collection Initiatives

While the postgenomic era holds great promise for the use of biospecimens, it also is changing the biospecimen needs of cancer researchers. It is estimated that more than 300 million specimens representing more than 150 million cases currently are stored in the United States, with over 20 million new specimens added each year.[12] However, many of these samples are not collected, stored, or annotated in a manner that is compatible with genomic analysis. Furthermore, recent Federal regulations, such as the Health Insurance Portability and

[10] For an annotated guide to the Web sites of the NCI Cancer Cooperative Groups, which conduct trials around the country and in Canada and Europe, see the Guide to Adult Cancer Cooperative Groups at: *www.nci.nih.gov/clinicaltrials/finding/cooperative-group-web-sites/page2*

[11] Eiseman E., Brower J., Olmsted S., Clancy N., and Bloom G. (2003). *Case Studies of Existing Human Tissue Repositories: "Best Practices" for a Biospecimen Resource for the Genomic and Proteomic Era.* RAND Science and Technology (August 28).

[12] Eiseman E. and Haga S.B. (1999). *Handbook of Human Tissue Sources: A National Resource of Human Tissue Samples*. Santa Monica, CA: RAND.

Accountability Act of 1996, have reconfigured privacy issues. Researchers often find themselves navigating a complex maze of intellectual property rights, liabilities, and other sociological barriers that currently impede the sharing of tissue samples for research and discourage clinical investigations. As a consequence, current programs and practices collectively fall short of meeting the research community's most pressing needs in genomic and proteomic research.

1.2.1 Researcher Needs

Cancer researchers have called repeatedly for biospecimens to be collected using standardized protocols, so that results can be reproducible and comparable. They seek greater research data accessibility through an open, Web-based platform, while remaining committed to the proposition that the collection and use of biospecimens and associated data must meet the highest possible ethical standards for protecting the privacy and confidentiality of the donor. They also recognize that the usefulness of biospecimens is maximized if accompanied by relevant demographic, social history, clinical, pathology, and longitudinal data, as well as genomic and/or proteomic data. A searchable, Web-based bioinformatics system therefore is seen as crucial for facilitating scientific discovery. Many investigators also have expressed a desire for the services that accompany tissue sample analysis, such as tissue microarrays and DNA or RNA assays. The suggested approach for incorporating all of these features into an NBN is described in this Blueprint report.

The potential sources for biospecimens are expected to be derived primarily from academic medical centers and community hospitals. The potential users would be primarily scientists and researchers at academic institutions, government agencies, and biotech and pharmaceutical companies. The potential uses of biospecimens and associated data are many, including for the following purposes:

- Target- and validation-discovery of molecular correlates
  - Primarily using RNA or protein analysis methods (large and small scale)
- Genomic analysis
  - Mutation screening
  - Loss of heterozygosity and amplification studies
  - Methylation studies
- Validation of diagnostic or therapeutic antibodies, or nucleic acid probes
- Pharmacogenomic analysis

These areas involve both major and minor cancer types, as well as specimens from primary and metastatic cancer sites. Descriptions of how particular NBN constituencies might use biospecimens are found in Appendix G. When determining which products to provide to its users, NBN must address the tissue amounts required, the tissue quality required (e.g., ranging from standard clinical quality to RNA grade to protein grade), the types and numbers of tissues required (e.g., primary and metastatic sites), and the format of tissue (e.g., sections of tumors to tissue microarrays) that would be most useful to researchers for each of the above purposes.

1.2.1.1 Commercial Interests

Pharmaceutical companies comprise an important component of the customer base for the NBN. Easier access to well-annotated cancer samples could help make oncology more attractive to pharmaceutical companies and will enhance investment in developing anticancer therapies. Advances in genomics are likely to continue to segment histologically defined cancers into better-defined subsets, which may result in smaller market segments. However, these smaller but genetically defined market segments may offer the possibility for better patient responses, and ultimately may pose less risk to private sector investments if identified during the early stages of the drug development process. Other commercial users would include a broad range of companies developing predictive and diagnostic products directed at cancer and other diseases. These include companies developing diagnostic/prognostic/therapeutic antibodies, as well as companies testing new technologies. Commercial companies need access to well-defined clinical samples in order to fully develop targeted agents and new technologies. The NBN will provide access to biospecimens for these industry customers.

1.2.1.2 Academic Centers

Academic medical centers will be a principal source of the operational and technique-related expertise for the NBN, as the vast majority of current research resources are located at academic health centers. Academic researchers, by virtue of their numbers, will also constitute the primary user base of the NBN.[13] Currently, many tissue access systems provide access to specimens and data only for researchers within the institution in which specimens are collected.[14] The NBN seeks to broaden access and standardize procedures for obtaining specimens.

1.2.2 Existing Resources

While a number of biospecimen resources exist at selected government, academic, nonprofit and for-profit institutions in the United States, there is no national, standardized, openly accessible biospecimens repository and database that is available to researchers who are interested in pursuing genomic or proteomic research. The NBN Blueprint will allow for the development of a system that will increase access across the country to these important biospecimens and associated data, while at the same time streamlining the collection and analysis of these samples from existing resources. In fact, elements of current biospecimen resources will be incorporated into the NBN through a "best practices" framework. The resource is intended to be shared openly among researchers at public and private institutions throughout the country, without the competitive or intellectual property constraints that are often barriers to resource sharing.

Although a large number of biospecimens exist in repositories today (there currently are approximately 350 organizations), materials are in various states of usefulness and readiness. In addition, no overarching standards exist, fresh/frozen tissue is not always readily available, and

[13] See *3. Biospecimen and Data Collection and Distribution* for a discussion about the different incentives needed in order for academic and community hospitals to participate.
[14] Eiseman E., Brower J., Olmsted S., Clancy N., and Bloom G. (2003). *Case Studies of Existing Human Tissue Repositories: "Best Practices" for a Biospecimen Resource for the Genomic and Proteomic Era.* RAND Science and Technology (August 28).

variations in specimen collection procedures and annotation across specimen collections are the norm (see Table 1-1).[15]

Preliminary findings from the RAND study[16] suggest that, while all studied repositories collect paraffin-embedded samples, some repositories have a relatively small collection of fresh frozen samples, and the clinical and longitudinal annotation is uneven. Although samples are used for genomics and proteomics studies after distribution, most repositories do not proffer genomics/proteomics data. The variability (or lack) of appropriate donor-informed consent and sample tracking capabilities limits resource utility. Finally, repository design is integrally linked to its original collection objectives, resulting in little cross-repository standardization.

Table 1-1: Summary of Current Limitations and the Ideal NBN Prototype

Limitations of Existing Systems	Ideal NBN Prototype
Wide variation in tissue collection, processing, and storage techniques, and difficulty obtaining sufficient samples for large-scale genomic and proteomic studies of rare cancers	Single, nationally coordinated network of pathological and normal tissue collection, employing standardized procedures for storage and distribution, as well as collection of associated clinical data
Nonuniform (or nonexistent) bioinformatics systems that are incapable of remote searching and data entry	Coordinated and centralized bioinformatics system for all aspects of specimen and data collection and dissemination
Restricted access to researchers outside institution at which specimens are collected	Extensive, external specimen-sharing is required of NBN collection centers on a national scale
Reluctance to share exhaustible specimen supply	Emphasis on collecting *inexhaustible* data from specimens
Consent procedures that are variable and may be insufficient for future genomics/proteomics research	Standardized consent for all specimens tailored to genomic and proteomic studies

As summarized in Table 1-1, the ideal NBN prototype is to have (1) a single, nationally coordinated network of pathological and normal tissue collection, employing standardized procedures for storage and distribution and collecting associated pathological, clinical, demographic, social history, and longitudinal data and (2) a coordinated and centralized

[15] Eiseman E. and Haga S.B. (1999). *Handbook of Human Tissue Sources: A National Resource of Human Tissue Samples*. Santa Monica, CA: RAND.
[16] Eiseman E., Brower J., Olmsted S., Clancy N., and Bloom G. (2003). *Case Studies of Existing Human Tissue Repositories: "Best Practices" for a Biospecimen Resource for the Genomic and Proteomic Era.* RAND Science and Technology (August 28).

bioinformatics system for all aspects of specimen and data collection and dissemination. The NBN is envisioned to have extensive sharing of specimens by NBN collection centers on a national scale, an emphasis on collecting *inexhaustible* data from specimens, and more or less standardized consent of all specimens tailored to genomic and proteomic studies. Like the UK NCTR, management of the databases and access to them would be monitored by an oversight body, whose function will be to safeguard the interests of all participants. The resource will be available to scientists and medical researchers, but there will be strict controls in place to protect the confidentiality of participants.

The NBN does not anticipate supplanting the existing tissue collection resources in the United States; rather, it seeks to fill a niche not served by current resources. Existing systems vary by the nature and intent of their collection protocols, the extent of their data, and their consent procedures. Some existing systems will align well with the goals of the NBN. Such systems may have an interest in participating in the NBN, and their involvement and expertise will be welcomed. Specialized Programs of Research Excellence (SPORE) grantees and Cooperative Group programs, for example, offer many of the characteristics that the NBN desires, and may have unmet needs that may be filled by alliances with the NBN. (See *3. Biospecimen and Data Collection and Distribution* and *6. Governance and Business Models* for additional information on this issue.)

Meaningful, broad molecular profiles of cancers can be developed optimally from tissues and clinical data that are collected using rigorous, highly standardized procedures. Some existing systems may not be candidates for genomics/proteomics research or future advanced technology purposes, but they are valuable in their own right. For example, banks like the National Surgical Adjuvant Breast and Bowel Project clinical trial bank are invaluable for their tight linkage between detailed clinical information and tissues; however, most of the samples were collected some time ago and are mostly fixed tissues embedded in paraffin. Similarly, the Armed Forces Institute of Pathology offers great depth and expertise in pathology diagnosis from fixed tissues, but it is not a state-of-the-art frozen tissue bank with validated clinical annotation, and never was intended to be such. These existing systems will continue to serve their specific users' needs.

1.3 NBN Requirements

Access to appropriately collected and annotated biospecimens is critical to accelerating progress against cancer in the postgenomics/proteomics era. Paraffin-embedded tissue is adequate for the clinical diagnosis of a specific cancer in an individual patient, and someday, with the advent of new technologies, it also may be used for a wide range of genetic studies. However, many state-of-the-art, molecular genetics-based technologies initially require fresh/frozen tissues, and successful identification and credentialing of drug targets or confirmation of diagnostic markers often depend on connecting tissue samples with a patient's characteristics at initial presentation and after appropriate followup. Additionally, what is developed as the optimum system today may not be what is needed after 5 years. Therefore, any system must be forward thinking, capable of expansion, and flexible, and it must anticipate potential technologies that are yet to be developed. The NBN system must have the capacity to grow as the information base grows and evolve as technologies advance.

It is the goal of the NBN to provide researchers from industry, government, and academe with a standardized, inclusive, high-quality network of biological samples that is shared, readily accessible, and searchable, using state-of-the-art informatics systems. The NBN must understand current and anticipated needs of all involved research communities, and it must provide a product that researchers will both desire and use. Output must be broadly available and readily accessible to users. Longitudinal clinical data must be periodically updated and should include data points focusing on therapeutic modalities, response measures, and outcomes. The system must incorporate high levels of QC, as the quality of the biospecimens and the accuracy of the resulting data will determine how relevant and usable the samples and data are to researchers.

To meet researchers' needs, the NBN must collect from patients with cancer sufficient numbers of tissues, blood, serum, and plasma in a manner that maintains the architecture of the tissues and the molecular integrity of DNA, RNA, and proteins in the biospecimens. The NBN should collect and provide detailed clinical (including longitudinal) and eventually genomic data for biospecimens, in addition to providing access to the biospecimens themselves. Efficient distribution of biospecimens and data to researchers would require an equitable peer review system for biospecimens and an integrated, searchable bioinformatics system for data. Long-term preservation of data would also need to be addressed.[17]

The Design Team identified the following, overarching requirements in order for the NBN to meet researcher needs:

Biospecimen and Data Collection and Distribution

- Biospecimens for banking (collected for storage in and distribution by the NBN) should be obtained only after all patient diagnostic needs have been met, and should be subject to appropriate bioethical structures and procedures, to ensure patient protection.
- The repository should consist of high-quality biospecimens appropriate for genomic and proteomic studies, and the type of biospecimens stored in the repository should be determined by an ongoing review of researcher needs.
- Annotation data (clinical, pathological, demographic, and social history) should be accurate, quality-controlled, and standardized across collection sites.
- The collection of longitudinal data should be strongly supported, and should include relevant biomarker measurements, if available. It is recognized that the costs for these data are likely to be high, and success will require innovative solutions.
- Genomic and/or proteomic testing may be performed on a subset of biospecimens by the NBN. Both the testing and specific subsets should be responsive to the needs of the users and flexible to changes in the research environment.

[17] Long-term preservation of data is a difficult challenge and one with which many sciences are struggling. An online catalog of specimens and an archive of integrated biomedical research data are distinct systems that need to be understood and built accordingly. Establishing long-term links and knowledge bases drives the need for establishing "active" archives. The Earth Observation System, which has spent years addressing this in earth science, may be a source for further insight on these issues (Personal communication, D. Crichton, August 23, 2003).

- The NBN should have a comprehensive representation of a broad diversity of disease and human populations. Biospecimen donors therefore should include a broad range of ethnicities, socioeconomic groups, and other demographic subgroups.

Bioinformatics and Data Management

- The repository should support open research access and be searchable and mineable via the Internet and incorporate computational analysis tools. The technology should be amenable to sharing appropriate clinical and longitudinal data, and at the same time should protect the donors' privacy and confidentiality. The repository should be available to a broad researcher base, and should associate clinical and experimental data with the specimens.
- The database should support integration and expansion, by establishing strict standards for data contributors and developing platforms that will expand and extend as the science grows. It should address the different vocabularies and data collection structures inherent among scientific communities (e.g., genomics, pathology) using common data elements.
- The database should support the exchange of information; it should capture data generated through use of the resource (both primary data and data interpretations), but should share and restrict data according to specific rules established by the NBN.
- Although likely to be challenging, it is important that validated, investigator-derived data be returned to the NBN and linked back to original NBN tissue samples. An expanded dataset, created by the return of this experimental data to the NBN, could then be made available to all investigators.
- The architecture should have the ability to "scale" as the volume of data increases, and it should have the ability to "extend" as datasets and types of data change. The architecture should provide interfaces that enable the construction of data mining and extraction tools, providing a comprehensive computational and data analysis environment.

Communications

- The conduct of a broader survey of the potential user population should be a priority, to develop a more detailed and accurate picture of potential research usage trends, the need for additional services, cost sensitivities for tissues and data, and advanced analysis services.
- Data and nomenclature standards being created for many types of research results should be incorporated into the NBN protocols.

Governance and Business Models

- The system must be highly flexible and capable of expanding as the science progresses.
- Development and standardization of collection, processing, storage, and distribution procedures, as well as QC and quality assurance monitoring, should be of paramount importance at all stages of the process, to allow for comparison of specimens from various collection sites.
- Data and samples should be distributed in a clearly articulated and equitable fashion, based primarily on the quality of the proposed research. Access to biospecimens should be controlled by a neutral, streamlined peer review system that is facilitated by a Biospecimen Utilization Review Committee. The tissue access system must include timely review of requests and distribution of samples, with minimal administrative burden.

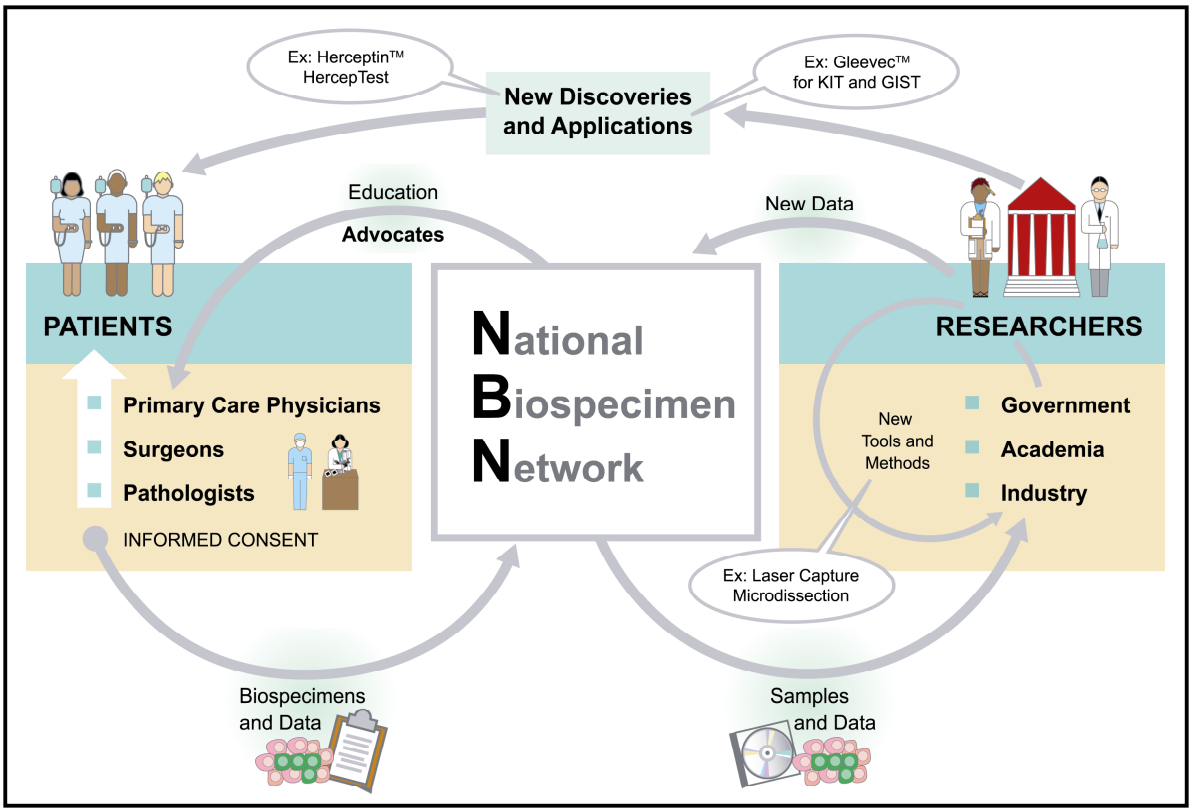

Figure 1-2. An Overview of the National Biospecimen Network

In short, the NBN seeks to create a data and tissue repository that provides aggregated, mineable information from a large number of biospecimens collected on a national basis. NBN specimens and data are expected to be highly valued and accessible to cancer researchers in both the private and public sectors. As depicted in Figure 1-2, researchers need tissue specimens and associated data as the basis for scientific discovery. The process will begin with the patient, who is the potential donor of this precious material. Voluntary health organizations can help by educating the population about the benefits of tissue donation, so that the concept is not a foreign one at an inopportune time (i.e., when a patient is diagnosed with a possibly fatal disease). After the patient provides informed consent, pathologists and surgeons will be involved in collecting, processing, and storing the specimen and providing associated data. The biospecimens and data will enter the NBN network; the NBN then plays a role in distributing the samples and associated data to researchers for use. Researchers will be invited and encouraged to return data derived from the NBN sample back to the NBN, in order to build up the national resource. The requirements proposed in this module should be considered long-term goals, with certain components of the NBN expected to become available over time.

Module 2

Management of Ethical and Legal Issues

2.
Management of Ethical and Legal Issues

An overarching issue for the National Biospecimen Network (NBN) is to operate with the highest possible ethical standards and legal compliance. This module examines ethical and legal issues facing patients, clinicians, and scientists, which must be considered when formulating plans for the development of the NBN. It presents the ethical framework under which the NBN must operate in order to ensure maximum protection for donors of tissues and thus encourage widespread participation in the system. The module also considers the issues of intellectual and other property rights associated with biological specimens and medical information.

2.1 Introduction

An overarching issue for the NBN is to operate with the highest possible ethical standards and legal compliance. In addition, maintaining high ethical standards will help ensure support and participation from patients while protecting their rights, and it will help garner the active participation of surgeons, pathologists, researchers and other supporters.

The relationship between the patient and the NBN is critical, since the patient is the potential donor of specimens. With the increases in genomic and proteomic research and heightened concerns related to genetic privacy, it is even more important for the NBN to effectively address patient concerns, particularly legal and ethical issues, and particularly with respect to interpreting and managing clinically derived molecular information. Some potential donors also may be concerned about residual rights to their tissues and medical information.

The primary ethical driver underlying the creation of the NBN must be the recognition that the needs of the patient always must come first—as is true in medicine in general (*primum non nocere*: First, do no harm). There is a need to balance scientific progress as a public good with patient protection as an individual right. Within the United States, however, the complex mix of social, cultural, and religious backgrounds of the American population has led to heterogeneous views on blood, organ, and tissue donation. In general, however, there is ample evidence from the National Bioethics Advisory Commission (NBAC) literature and other published information that most Americans are in favor of their specimens being used for research. Because of the complexities of the American landscape, it is critical that the "informed consent process" be of the highest quality. The collection of the specimens must minimize privacy risks to patient-donors and must be cognizant of their interests and needs. Finally, the consent process must also anticipate the needs of future discoveries and therapeutic advances.

2.2 Background

In its 1999 report, NBAC stated that:

> Any ethically sound policy for research uses of human biological materials must reflect a defensible balance of the ethical reasons that support greater control over

> the use of human biological materials and stronger protections for subjects, on the one hand, and the ethical reasons that support greater access to samples for purposes of conducting clinically beneficial research and/or clinical interventions, on the other hand. These reasons will vary in weight and impact depending on the identifiability of the sample sources and on the probability and magnitude of various wrongs and harms that may occur.[1]

An understanding of the balance between the interests of society and the rights of individual patients, which is central to the ethical use of human tissues and the use of clinical outcome information for research, must underscore the development of any national tissue resource.

Human biological materials and associated health information fall into a number of categories, based on when and where they were collected and the extent to which they can be associated with their donors, or how "identifiable" they are. Different ethical standards guide the handling of the different categories of materials, and existing legislation requires different standards of care to be exercised, depending on the circumstances. Planning the NBN therefore requires an understanding of the categories of specimens that will be needed to create an effective network, the risks associated with each category, and the steps that can be taken to reduce these risks while still collecting useful information.

2.2.1 Identifiability of Biospecimens and Information Collected and Used for Research

Within a biospecimen repository, the identifiability of specimens may range from truly unidentifiable specimens to identified specimens. The identifiability of specimens obtained by researchers from these repositories can also vary, depending on the needs of the specific research study for which the specimens are being sought. The more identifiable the specimens are, the greater the risk to donors' privacy and confidentiality. Therefore, the policies and procedures established for the collection and distribution of specimens should reflect the level of risk and possible harm posed to donors.

The two Federal regulations that address this issue of identifiability, and therefore pertain to the creation and operation of a biospecimen repository, are the Federal Policy for the Protection of Human Subjects (also known as the "Common Rule," which is codified for the Department of Health and Human Services [HHS] at subpart A of Title 45 CFR part 46) and the Standards for Privacy of Individually Identifiable Health Information (also known as the "Privacy Rule," which is codified at Title 45 CFR Part 160 and subparts A and E of Part 164). The Common Rule pertains to human subject research that is conducted or supported by HHS, or conducted under an applicable assurance approved by the Office for Human Research Protections (OHRP).

Under the HHS regulations for the protection of human subjects at 45 CFR 46.102(f), "human subject" is defined as a living individual about whom an investigator conducting research obtains (1) data through intervention or interaction with the individual, or (2) identifiable private

[1] National Bioethics Advisory Commission. (1999). *Research Involving Human Biological Materials: Ethical Issues and Policy Guidance, Volume I: Report and Recommendations of the National Bioethics Advisory Commission.* Rockville, MD (August), p. 51.

information. Private information includes information about behavior that occurs in a context in which an individual would reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record).

If an investigator obtains private information about a living individual for research purposes that retains a link to individually identifying information, such private information is not ordinarily considered to be individually identifiable to the investigator if (1) the investigator and the holder of the individually identifying information sign an agreement prohibiting the release of individually identifying information to the investigator under any circumstances, or (2) there are other legal requirements prohibiting the release of the link to the investigator.[2]

Under the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, covered entities may use or disclose, without restriction, any health information that is deidentified by the removal of 18 specified identifiers enumerated in the Privacy Rule.[3] The covered entity also must have no actual knowledge that the remaining information could be used alone or in combination with other information to identify the individual who is the subject of the information.

---

[2] Department of Health and Human Services. (2003). *Institutional Review Boards and the HIPAA Privacy Rule* is available at: *privacyruleandresearch.nih.gov/IRBandprivacyrule.asp.*

[3] These 18 identifiers are:

1. Names
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP Code, and their equivalent geographical codes, except for the initial three digits of a ZIP Code if, according to the current publicly available data from the Bureau of the Census:
   a. The geographic unit formed by combining all ZIP Codes with the same three initial digits contains more than 20,000 people
   b. The initial three digits of a ZIP Code for all such geographic units containing 20,000 or fewer people are changed to 000
3. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older
4. Telephone numbers
5. Facsimile numbers
6. Electronic mail addresses
7. Social security numbers
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers, including license plate numbers
13. Device identifiers and serial numbers
14. Web addresses
15. Internet protocol address numbers
16. Biometric identifiers, including fingerprints and voiceprints
17. Full-face photographic images and any comparable images
18. Any other unique identifying number, characteristic, or code, unless otherwise permitted by the Privacy Rule for reidentification

Covered entities may also use statistical methods to establish deidentification instead of removing all 18 identifiers. Covered entities using the statistical method to establish deidentification may obtain certification by "a person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable."[4] There is a "very small" risk that the information could be used by the recipient to identify the individual who is the subject of the information, alone or in combination with other reasonably available information. The person certifying statistical deidentification must document the methods used as well as the result of the analysis that justifies the determination. A covered entity is required to keep such certification, in written or electronic format, for at least 6 years from the date of its creation or the date when it was last in effect, whichever is later.

Under the first method of deidentification under the Privacy Rule, unique identifying numbers, characteristics, or codes must be removed if the health information is to be considered deidentified. However, the Privacy Rule permits a covered entity to assign to, and retain with, the health information a code or other means of record identification, if that code is not derived from or related to the information about the individual and could not be manipulated to identify the individual. The covered entity may not use or disclose the code or other means of record identification for any other purpose and may not disclose its method of reidentifying the information. For example, a randomly assigned code that permits reidentification through a secured key to that code would not make the information to which it is assigned Protected Health Information (PHI) because a random code would not be derived from or related to information about the individual and because the key to that code is secure.

The issue of identifiability of biospecimens and information is an important one to the operation of NBN. Tissue samples are more useful to researchers when accompanied by demographic and clinical information, some of which may make them identifiable. In determining the extent to which it will collect and maintain identifiable samples and information, NBN must balance the needs of its end users with the ethical requirement to protect the confidentiality and privacy of health information and samples, and must adhere to all existing regulations. While NBN itself is not directly subject to HIPAA, tissue collection sites almost certainly will be. NBN must be able to assure its partners in the system that they will not be at risk for privacy violation through their relationship with NBN.

2.2.2 Ethical Considerations in Developing the NBN

There are a number of key ethical issues that the NBN must consider when developing its operating principles and procedures. The act of collecting and storing human biological materials and related health information in publicly available repositories and archives poses few risks to donors, and those are primarily social risks. Developing strategies to mitigate these risks is a fundamental ethical (and sometimes legal) responsibility of NBN. Communicating the steps taken to the potential donor population will be important to encouraging patients to participate in

[4] Department of Health and Human Services (2003). *Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule,* NIH Publication Number 03-5388.

the NBN. This section discusses some of the ethical issues that concern donors, which must be considered in the NBN design.

2.2.2.1 Possible Harm at the Time of Tissue Donation

Collection of specimens must first meet the needs of patient diagnosis. If specimens over and above that required for clinical care are taken at surgery or biopsy, then the donation might pose an increased risk to the donor. NBN will seek only the amount of tissue that poses minimal risk to patients.

2.2.2.2 Loss of Privacy and Confidentiality

Potential donors may be concerned that information about them derived from tissues and medical records may be released and harm them. One of the primary concerns is that employment and insurance discrimination might result from exposure of information about health history, genetic makeup, or familial predisposition to disease. Although there are anecdotal reports of employment termination and denial of insurance coverage based on genetic predisposition to disease, there is no definitive research on the extent of these risks; however, the perception is real and the risk exists.

2.2.2.3 Donors' Access to Research Results

The Privacy Rule provides research subjects (in this case, donors) with certain rights about how their health information is used and disclosed. With few exceptions, the Privacy Rule guarantees individuals access to their medical records and other types of health information, to the extent the information is maintained by the covered entity or its business associate within a designated record set. Research records maintained by a covered entity may be part of a designated record set if, for example, the records are medically related or are used to make decisions about research participants.

Patients may wish to receive information derived from research using their tissues. Such information includes not only the results of research at issue, but also discoveries made when preparing the specimens for banking, which may have medical significance for the donors. For example, reexamination of the tissues by repository pathologists or researchers may lead to a different diagnosis than that made by the hospital pathologist. In establishing the operational procedures and developing the informed consent process, it will be critical to decide how these situations will be handled.

Donors may also have questions about ownership, access to, and control over the specimens they have been donated. Advances in medical science are likely to produce more sensitive diagnostic and prognostic tests. Patients may wish to store tissues, such as tumor cells, against the possibility of better tests to predict the likelihood of recurrence or progression of disease, or sensitivity to treatment in the case of relapse.

It has been suggested that biospecimen banks agree to inform donors of future discoveries and therapeutic advances as a *quid pro quo* for tissue and information contributions. Factors that complicate this arrangement and may make it impractical include:

- The deidentification of tissues and data that make it difficult to recontact donors

- The progression of a patient's disease, and subsequent treatment that may make it difficult to determine the relevancy of the new information
- The very small likelihood that clinically or statistically validated results would be available during the course of the patient's active disease
- The sheer magnitude of the task of maintaining a valid patient-contact database.

The availability of research outcomes poses another, and seemingly contradictory, concern. There is a risk that physicians may use preliminary research data communicated to them about their patients for clinical decisionmaking before clinical usefulness is validated, posing a different risk to the patients' well-being. Recent debate over the breast cancer (BRCA) gene mutations demonstrates the risk in sharing results prematurely. Early results suggested that the presence of the BRCA mutation was strongly associated with the development of BRCA. Some women, in particular those from populations known to carry the gene at a higher rate, were tested for the presence of the mutation, and some opted for prophylactic surgical intervention, even though the benefits for such an approach had not been demonstrated. Subsequent research indicated that the association was less clear than originally supposed, and that the benefits of prophylactic intervention overall remain unclear.[5]

2.2.2.4 Conflicts of Interest

A related area of potential concern to donors is the possible conflict of interest that arises for physicians who have incentives to collect particular tissue specimens, perhaps to further their research agenda or for the benefit of the institution. A physician choosing to undertake a more invasive procedure may appear to be pursuing this course because he or she desires the specimen. This issue is becoming increasingly pressing in light of new, noninvasive technologies that are used with increasing frequency to diagnose cancer (e.g., radiology, laparoscopic surgery), making more invasive procedures that result in collection of more tissue less necessary.

2.2.3 Informed Consent

Informed consent is a key mechanism NBN will employ for protecting the rights of donors: its goal will be to ensure respect for persons, mutual understanding of research procedures, risks, rights, and responsibilities, and continuous voluntary participation. The process will be designed to provide information about a research protocol that the potential donor can understand and use to make an informed decision about participation. Rather than being a simple form to sign, informed consent is an educational process between the investigator (or tissue collector in this case) and the prospective subject (or the subject's legally authorized representative).

Five elements of informed consent that are derived from fundamental ethical principles have been summarized by the NBAC:

[5] Eisen A. and Weber B.L. (2001). Prophylactic Mastectomy for Women with BRCA1 and BRCA2 Mutations – Facts and Controversy. *N. Engl. J. Med.*, Vol. 345, No. 3 (July 19): 207-208.

- Informed consent must include **full disclosure** of all anticipated relevant risks and benefits of the research. Derived from the principle of respect for persons, a participant has the right to know the future use planned for his or her tissues or medical information.
- The principle of respect for persons also demands that the choice to participate be made **voluntarily** and that a clear **statement of choice** (an expressed decision) to participate in the research be made by the potential participants.
- The informed consent process performs an assessment and gives assurances of the **competence** of the subjects to make a decision regarding whether to participate in the research and to **comprehend** the relevant risks and benefits to the potential participants. The requirement for assurances of competence and comprehension are derived from recognition of individuals' autonomy under the principle of respect for persons.[6]

Currently, a hierarchy of informed consent, from more open to more restrictive, exists in practice. However, as testing methodologies advance, researchers likely will need access to expanded minimum datasets for each sample to support deeper, more productive research. Meaningful dialogue around the risks and benefits of augmented informed consent, up to and including explicit universal consent, will need to take place among patients and advocates, oncologists, surgeons, pathologists, scientists, institutional review boards (IRBs), and other regulatory agencies. Patient advocacy groups have played a key role in the development of the model consent forms and in educating patients about the benefits and precedents of consent. (Various types of guidance on developing informed consent are included as Appendix H.) Advances in medical research will require that the consent process be flexible and capable of handling new demands placed on it by ever-greater needs for access to patient information.

Informed consent for medical research in the United States is based on the opt-in model. Under this model, potential tissue donors give specific consent to participate in the research described in the informed consent document. Another approach to consent is the opt-out model. Opt-out models presume consent, unless a person specifically elects not to participate.[7] Under the HHS regulations for the protection of human subjects, this model would only be permitted if informed consent could be waived by an IRB as stipulated at 45 CFR 46.116 (c) and (d). Although an opt-out scheme would facilitate the creation of a national tissue repository, within the United States it is not considered acceptable to require all patients to participate in a national tissue system if they have a biopsy. Patients might prefer that their samples go to alternative tissue banks for their own future use or specimen research, or they might object to banking of their tissue entirely. However, it may be possible to have patients check a box that directs the transfer of

[6] Faden R.R. and Beauchamp T.L. (1986). *A History and Theory of Informed Consent,* New York, NY: Oxford University Press.

[7] Opt-out consent models have been used in Europe for the collection of health information and specimens for research. Iceland has created the Iceland Healthcare Database, which is correlating genetic and medical information for the entire population of Iceland. Patients not wishing to participate in the system may specifically request not to do so. The existence of the choice to opt out of the system has led some to argue that those staying in the system have given their broad consent. Although this system is clearly an extremely valuable source of information, some bioethicists have argued that presumed broad consent is not truly informed consent, because patients probably have not considered all of the issues related to the possible use of their information. One benefit of an opt-out system is that it increases participation. In the United Kingdom, there has been a call for establishing an opt-out organ donor program, wherein all adults would be considered organ donors unless they request otherwise.

residual tissues to a national resource. Many ethicists agree that consent for unspecified use of their specimens is an appropriate way to obtain consent for a biospecimen bank such as the NBN. However, informed consent or a waiver of informed consent for specific studies using specimens and identifiable data from the NBN may be necessary if identifiable specimens are to be used by researchers. (The National Cancer Institute's [NCI] tiered consent is discussed in Appendix I.)

2.2.4 The Role of IRBs

The role of an IRB is to determine that the anticipated benefits of research are worth the risks to patients**.** The IRB is expected to protect the interests of human research subjects, and as such, IRBs can be expected to play an important role in the development of the NBN informed consent process and protocols for collection of samples and associated patient data.

The OHRP delineated the role of IRBs in HHS-conducted or -supported human tissue repositories in its 1997 guidelines, "Issues to Consider in the Research Use of Stored Data or Tissues." The guidance states that operation of a repository and its data-management center should be subject to oversight by an IRB. The IRB should review and approve protocols that specify the conditions under which specimens and data may be accepted and shared, ensuring that there are adequate provisions to protect the privacy of subjects and maintain the confidentiality of data. The guidance also states that an IRB should review and approve sample collection protocols and informed consent documents for distribution to tissue collection sites and their local IRBs, if specimens are sent to off-site repositories.

If a licensing approach is taken, submitting institutions would be granting usage rights to biospecimens when they are turned over to the NBN, but not relinquishing control of these specimens. The Early Detection Research Network (EDRN) has addressed parts of this issue in deploying its distributed national specimen-sharing infrastructure. EDRN established a common protocol for submission to IRBs that has greatly decreased both the variability and the time required to achieve IRB approval.[8] NBN would want to take steps to simplify this process as much as possible, in order to increase the incentive for institutions to participate.

2.2.5 The HIPAA Privacy Rule

The HHS established the Privacy Rule to safeguard the privacy of individually identifiable health information, as required under HIPAA (*www.hhs.gov/ocr/hipaa*). Many states also have enacted statutes to protect the privacy of health information, and to prohibit discrimination on the basis of genetic information in the provision of insurance or employment.

The Privacy Rule restricts the use or disclosure of PHI by covered entities (health plans, health care clearinghouses, and certain health care providers). Many research organizations that handle

[8] See National Cancer Institute. (2002). *The Early Detection Research Network: Translational Research to Identify Early Cancer and Cancer Risk.* Second Report. Division of Cancer Prevention. (October); *www3.cancer.gov/prevention/cbrg/edrn/edrn_report2002.pdf;* or Kincaid H., Crichton D., Winget M., et al. (2003). *A National Virtual Specimen Database for Early Cancer Detection.* Paper prepared for presentation to the Sixteenth IEEE Symposium on Computer-Based Medical Systems, June 25-27, 2003, New York.

individually identifiable health information will not have to comply with the Privacy Rule, because they are not covered entities; however, they still may be affected if they rely on covered entities for research support or as sources for individually identifiable information for research.

There is a difference between informed consent and the authorization process (i.e., authorizing the use of PHI). The informed consent process provides research subjects with an understanding of the study and of any anticipated risks and/or benefits. This may include a description of how the confidentiality of records will be protected. The Privacy Rule authorization is a permission that discloses how, why, and to whom the PHI will be used and/or disclosed for research. The informed consent and authorization request can be combined into one form.

Under the Privacy Rule, the development of a repository or database for future research projects falls within the definition of research. If the organization creating or maintaining the repository or database is not a covered entity, then HIPAA does not apply (although other state and Federal regulations may come into play). Complications may arise if tissues are maintained in a distributed network of entities, some of which are covered entities and some of which are not. If the data are to be stored in a centralized location, its operations may be affected by whether it is considered a covered or noncovered entity. (More details about allowable uses of PHI for research can be found at: *privacyruleandresearch.nih.gov/pr_02.asp*/).

2.2.6 Property Rights to Biospecimens within the NBN System

A primary issue in the property rights discussion is the question of who will ultimately profit from the donated biospecimens.[9] Profits most generally stem from inventions derived from biospecimens plus associated analytic information, rather than from the biospecimens or baseline information alone. These inventions are protected as intellectual property (IP). Current case law would support the interpretation that only value-added contributions qualify as IP.[10] Simple possession of specimens does not confer IP rights. Some have argued that if the tissue specimen could have been obtained from another patient and cannot be considered novel, it is not subject to IP protection.

The NBN team agreed that, under the Common Rule and based on the principle of autonomy, there is a legal and ethical requirement to allow donors to withdraw their specimens from the repository before the specimens are distributed for research. This can be accomplished by assigning a unique code to the sample at the collection site, which links donor to specimen. Donors could apply to the collection site for withdrawal of their samples. The site then could communicate to the NBN repository the donor's request to withdraw a specific coded sample. This is the practice currently being followed by some repositories. However, it is impossible to withdraw from the system those samples that have been distributed to researchers.

[9] Federal and state laws are not always in agreement on these issues. NBN operations core and legal staff will need to examine current laws as they relate to Federal and state laws and regulations to ensure compliance.

[10] Moore v. Regents of the University of California, 51 Cal. 3d 120, 271 *Cal. Rptr.* 146, 793 P.2d 479, 15 USPQ2d (BNA) 1753 (1990) (reversing and remanding trial court action sustaining defendants' demurrers), *cert. denied*, 111 S. Ct. 1388 (1991).

2.2.7 Potential Sources of Liability

NBN will have many potential sources of liability, most of which are unlikely to happen, and many of which can be mitigated with well-considered data-sharing agreements.[11] These potential sources include the following:

- Negligence in the collection, storage, and dissemination process
- Adverse outcome to the patient during the collection process
- Exposure to biohazard risk
- Tissue samples anticipated for the repository are cancer tissues and correspondingly matched healthy tissues. Because of this collection strategy, specimens will not necessarily be tested for other pathologies. It is possible that some specimens might constitute a biohazard risk, and liabilities might arise in the transfer of affected materials regarding possible exposure of individuals who are working with the materials.
- Violation of patent protections by a third party
- Breaches of privacy and confidentiality
- Lawsuits by groups of donors over access to profits. There is a slight risk that individuals (or, more likely, groups) could initiate a lawsuit to try to gain access to profits (or possibly information) from a patent/license that was based in part on their specimens or information. Even if it is not the holder of the patent, a biospecimen repository system that is licensing use of the biospecimens may be at risk in this situation.

2.3 NBN System Requirements and Recommendations

2.3.1 NBN Must Establish a Chain of Trust in Specimen Collection and Handling

NBN must encourage specimen donation to the system. It is recommended that NBN build the system on a conceptual "Chain of Trust," which begins with the patient and runs through to the researcher. Each link in the chain of the NBN will be entrusted with the responsibility to ensure the privacy and safety of donors, and to comply with their wishes. Entities that are involved in each step of the process—from the IRBs that approve collection of samples and data, through the parties responsible for collection and storage of tissues and information, to the qualified researchers who use the samples and data for appropriate research purposes—will have specific obligations to protect the donors. Operational policies, procedures, and structures can be conceived as the materials out of which strong links in the chain of trust are forged. The strength of the entire chain is the assurance that the NBN can give to potential donors to promote their trust for the responsible use of their biospecimens and associated data.

[11] None of the twelve repositories interviewed by the RAND team reported any lawsuits against them related to the issues listed. See Eiseman E., Brower J., Olmsted S., Clancy N., and Bloom G. (2003). *Case Studies of Existing Human Tissue Repositories: "Best Practices" for a Biospecimen Resource for the Genomic and Proteomic Era.* RAND Science and Technology. (August 28).

2.3.2 NBN must Reflect a Diverse Population

One of the challenges facing NBN will be ensuring the participation of the broad range of racial and ethnic groups that make up the American population. In order that a system be just, it must allow for the equal sharing of both risks and benefits. In the NBN case, this will mean both participation through donation and enjoyment of the benefits of the research, whether the research leads to an increased knowledge about a particular health risk or, more importantly, an improvement in health outcomes from group-specific disorders.

It is recommended that NBN pursue the selection of collection sites that will increase the genetic and geographic diversity of its biospecimens, in order to accelerate scientific progress that will benefit the diverse American population and help reduce health disparities.

2.3.3 Ethical and Legal Management of NBN

Appropriate ethical and legal management will be key to developing and operating the NBN. The Design Team recommends the early creation of a Bioethics and Legal Advisory Board to guide these activities, an approach followed by many biospecimen banks. It is recommended that membership be drawn primarily from a group of external content experts. Other members could be drawn from past and potential donors, as well as from parts of the NBN, including members of the Governing Board, Operations Center, and Business Units. The Board would report to the CEO, who would be charged with its formation and (ideally) would chair it. Like quality assurance (QA), it would be an important executive function about which the Governing Board would receive regular reports. Key activities of the Bioethics and Legal Advisory Board would include:

- Development/review of the informed consent document and process
- Development of guidelines for the distribution and transfer of samples
- Review of the communications/education/outreach plans to ensure that appropriate target audiences and messages have been selected
- Monitoring of the regulatory landscape (examples of existing ethics oversight committees are included as Appendix J).

2.3.4 Effective Informed Consent Process

It will be critical that essentially the same fundamental informed consent process be followed at all collection sites. NBN will require an approach that can be implemented early in the process, preferably at the first contact between potential donors and medical personnel; the approach also will be subject to QA. To implement this plan, these questions must be answered:

- What are the critical elements of consent?
- How will the consent process be administered?
- When and where will consent be obtained?

- Who will administer the consent process?
- Will the NBN consent be separate from other consents (such as the surgical consent)?

Numerous Federal agencies, professional organizations, and academic institutions have developed approaches to informed consent. In planning the NBN, it will be important to evaluate these approaches and determine an effective model, or to identify the elements needed for an effective process.

The most efficient place for administering the informed consent may be when a patient is admitted to a community-based facility for a diagnostic biopsy or treatment. Under these circumstances, potential donors, as patients, are asked to sign a clinical informed consent for the surgical process. The 1999 NBAC report recommends that the consent for collection of specimens for research should be kept separate from consent for treatment/diagnosis. This report also points out that asking patients to choose among a series of options in a tiered consent process, just at a time when they are faced with the possible diagnosis of a serious illness, may not be in the best interests of the patients. In this case, it was suggested that it might be appropriate to use a more general consent for prospective use of tissue samples.

While it was originally thought that NBN would utilize a uniform document, this may prove to be difficult because:

- Removal of different types of tissue from different sources represents varying levels of risk to patients and will require different levels of informed consent
- IRBs have not routinely approved standardized consent forms and processes in the past
- Collection sites are unlikely to accept a standard form.

The Design Team noted that the consent process and the form itself would vary, depending on the context. It is suggested that NBN consider the tiered consent form used by the NCI cooperative groups (which has been highly successful) in the eventual design of the NBN consent process. It will be important for NBN to identify the minimum information requirements for both the informed consent document and process, which will be a requisite for collection site participation.

2.3.5 Protection of Donors' Privacy

NBN will require the protection of the privacy and confidentiality of donors and their family members throughout the entire process of collection, storage, and use of biospecimens and information. The Design Team suggests that this most likely will be accomplished through the design of a secure bioinformatics system that offers maximum security for the protection of health information in the NBN database. Further information regarding suggested security procedures may be found in *4. Bioinformatics and Data Management*.

2.3.6 Alignment with HIPAA Privacy Rule

The exact impact of the HIPAA Privacy Rule on the NBN will depend on a number of factors. It will need to be determined whether the NBN repositories and databases themselves are (or will

be located) at covered entities. Further, the extent to which NBN's various business units (such as the collection sites and researchers using sample and information from NBN) are covered entities also will potentially affect the operations of NBN. In general, the Privacy Rule prevails unless state laws relating to privacy of health information are both contrary to and more stringent than the Privacy Rule. In order to develop a process to reliably comply with HIPAA, it will be necessary to understand state requirements and to develop an approach that meets the most stringent requirements.

An element of the Privacy Rule that may affect NBN is the requirement to account for certain disclosures of PHI for up to 6 years from the date of the request. Data disclosed as part of deidentified data or minimal datasets with data use agreement are not subject to this type of accounting; NBN may consider use of these types of data sets. Although there are several methods for accounting for research-related disclosure (if this is necessary), all methods will require that the process developed to manage the data disclosures within the NBN be designed in advance.

Legislative and regulatory relief may facilitate the NBN's establishment. The Design Team recommends that a mechanism be established within NBN to monitor both the regulatory landscape and any guidance from the Office for Civil Rights and the OHRP (perhaps via the Bioethics and Legal Advisory Board).

2.3.7 Intellectual Property Rights

Some believe that IP issues will represent a major barrier for the creation of an NBN because of concerns that third parties can demand royalties. To address these concerns, the following IP-related criteria should be built into the operations of the system:

- The governing body of the NBN should avoid asserting any patent or other enforceable IP rights on materials it distributes, as asserting IP rights would hinder open access.
- Use of NBN tissues or information and subsequent patenting should not prevent future use of materials or information in the repository.
- A strong disclaimer should be used to ward off third-party claims (an adequately protective disclaimer should put the onus on the customers to obey all laws and restrictions established within the market).
- To the extent possible, a free market approach with respect to IP issues should prevail.

The Design Team recommended that the NBN create a licensing system whereby NBN has permission to use—and in turn gives permission to use—biospecimens and information. The NBN should encourage researchers to patent discoveries made with the use of its resources. However, use agreements with researchers should make clear that allocations for patents that would block the use of any NBN biospecimens and information by others are not allowed.

While NBN may not wish to assert any ownership rights to tissues or information, it may, in the course of developing the infrastructure, develop processes, process improvements, or technologies that are patentable. NBN may wish to apply for patent protection for these inventions, or it may wish to publish them, making them publicly available and preventing

another entity from patenting them. The latter approach is consistent with a desire to facilitate public research on cancer. The former would, however, assist NBN in becoming self-sustaining. However, use agreements should be reasonably designed to ensure that ultimate patent holders agree to not interfere with the ability of others to utilize the NBN resources.

The existence of the repository should serve to remove IP barriers for all participants. The NBN's goal should be to develop a system with unencumbered use of the information and samples, providing that peer review, privacy, payment, and similar constraints are met. Reach-through rights to future discoveries will be a disincentive to potential users, who must consider the possible return on investment.

2.3.8 Engagement of Patient Advocates

Patient advocacy groups have been an increasingly powerful voice in protecting the interests of their members and bringing to light issues surrounding particular diseases. There are a number of characteristics of patient advocacy groups, which could make them effective partners in developing and implementing the NBN. Typically, patient advocacy groups:

- Recognize the benefits of participating in the research process
- Serve as a voice for the patient population
- Provide access to the patient community (their membership)
- Have considerable influence over their constituencies
- Are powerful special-interest lobbyists
- Have an interest in the education of the patient population.

The advocacy community has played a key role in promoting research in cancer, genetic disorders, and other diseases. Although advocacy groups vary in focus, size, and scope of activities, their input can be invaluable to strengthening the design, development, and implementation phases of the process, and to maximizing responsiveness and relevancy of the NBN for patient needs. It will be important to design roles and expectations to guide their participation in the process, so that their contributions are directed where they can be most effective. For example, patient advocates could be asked to play a key role in identifying or reviewing ethical and socio-cultural concerns of potential donors; in defining an optimal informed consent process and raising sensitivity about practices and procedures that can discourage participation; and in developing and helping to implement an education program about the resource, tissue donation, and participation in research generally.

2.4 Summary of Key Requirements

Adherence to the highest possible ethical standards and legal considerations will be critical to the success of NBN. The assurance that the donors' best interests are at the heart of the system—from the standpoint of both reducing the burden of cancer and protecting them from harm to the greatest extent possible—will help to ensure the support and participation of the broadest donor population possible. It also will help garner the active participation of researchers and other supporters, who might be concerned with possible risks in using the system if such principles were not at the foundation of the system. The key tactic to effecting this strategy will be the

establishment of a Bioethics and Legal Advisory Board early on, to guide the start-up NBN activities. Other recommendations can be roughly divided into ethical and legal issues.

Ethical

- The establishment of a "Chain of Trust" to ensure the privacy, safety (protection from harm), and compliance with the wishes of the donors. Other links in this chain include the patient, IRBs, those responsible for the collection and storage of tissues and information, and researchers.
- The NBN informed consent process must ensure that potential participants understand how their specimens may be used and should use tiered consent procedures that provide individuals with options for levels of participation.
- In general, donors will not be apprised of specific research results derived from their particular specimens, but information about scientific discoveries made through the use of biospecimens will be publicly available.
- NBN should pursue the selection of collection sites that will increase the genetic and geographic diversity of its biospecimens.
- Specimen allocation should be equitable and ensure appropriate use by qualified researchers in qualified organizations.

Legal

- NBN will adhere to applicable Federal, state, and local rules and regulations.
- NBN should avoid asserting any reach-through rights to IP that are generated through researcher use of tissues and associated data, since attaching such rights may hinder access by certain users and slow research.
- NBN should create a system whereby it has permission to use, and in turn gives permission to use, all biospecimens and information.
- NBN will follow all applicable laws and regulations (including but not restricted to the HIPAA Privacy Rule) that restrict the use and release of identifiable medical information to insurers, employers, and others.

Module 3

Biospecimen and Data Collection Distribution

3.
Biospecimen and Data Collection and Distribution

This module describes the principles governing the collection, processing, storage, and distribution of materials and data for the National Biospecimen Network (NBN), *after* the patient has provided informed consent. NBN system requirements with respect to its organization, users and contributors, tissue sources, collection, annotation, inclusion of longitudinal data, possibility for advanced analyses, storage, distribution, safety, and incentives to participate are further detailed. This module also reviews existing tissue banks and explains how the NBN is expected to differ from and interact with them.

3.1 Introduction

The National Dialogue on Cancer (NDC) Tissue Access Working Group (TAWG) called for the establishment of a nationally coordinated, standardized, inclusive, high-quality network of pathological and normal tissue collection, storage, and distribution, with associated clinical data, that is accessible through a user-friendly informatics system. It also envisioned the need to develop and implement national standards for the proposed system in areas such as sample collection, annotation, storage, and distribution. Such standards would also be developed for training, site monitoring, sample tracking, and quality management. The implementation of standard collection protocols is expected to minimize experimental variability and accelerate scientific progress. Variability in acquisition, processing, and storage of tissue may contribute to undesired experimental variability in whole-genome analysis. One of the uncompromising propositions of the NBN must therefore be the collection of biospecimens and associated data according to known and widely accepted standards, and in conformance with ethical and legal requirements (see *2. Management of Ethical and Legal Issues*).

Although the NDC TAWG thought a fresh start would be optimal to achieve the goals of the NBN, the NBN Design Team recognized the complementary value of many existing resources and their usefulness for a range of purposes. The RAND Corporation review of selected existing resources observed that some existing collections might be useful for some research, but collection methods varied with not all storing frozen tissues, making comparisons across resources difficult. Most repositories designed for research purposes collect demographic and diagnostic data from tumor registries and/or patient medical records. It will be necessary for the NBN to critically evaluate existing resources for their specific research utility, and to work with valuable extant resources to maximize benefits to research.

3.2 Background

3.2.1 Existing Resources

Thousands of tissue repositories currently exist within the military, private industry, hospitals and hospital consortia, cooperative groups, research programs, and academic institutions.

However, there is great variance among these existing resources in their focus (e.g., treatment or research), the resources available for tissue testing and annotation, and the ability to control the use of tissue samples. Many of these specimens were not collected or stored for research, but are essentially pathology archives. These existing repositories share no specific standards of patient consent procedures, acquisition, storage, or quality control (QC); are not able to share easily either data or tissue; and are not linked in terms of tissue distribution.

Many existing tissue banks have served research well and are valuable for advancing knowledge through the first stages after discovery. For example, the Cooperative Breast Cancer Tissue Resource has more than 9,000 annotated samples and is a major source of tissue from patients who have participated in randomized trials. Specimens from clinical trials in many organ systems, mainly routine formalin-fixed paraffin-embedded tissues, are available in the repositories of the major clinical cooperative group banks. High-quality specimens are also stored in Specialized Program of Research Excellence (SPORE) repositories and laboratories participating in the National Cancer Institute (NCI) Director's Challenge: Toward a Molecular Classification of Cancer.

RAND is examining a number of existing tissue resources to help clarify the nature of available resources.[1] Preliminary results confirm that few common standards exist among resources in terms of how specimens are collected, stored, processed, and distributed. This is partly because the repository design is integrally linked to an institution's objectives, and these resources were designed to serve multiple purposes including: Prospective collection, storage, and distribution; standardized tissue microarrays; diagnosis; and study-driven banking. The RAND team also observed that anticipating a repository in the initial study design, rather than assembling one after the fact, increases the likelihood of standardization across samples.

Most of the resources examined by the RAND team receive a substantial amount of public funding. In general, the repositories are integrating new technologies, and tissues from most repositories are already being used for genomics/proteomics research. All of the repositories are actively addressing ethical, legal, and privacy issues. These resources also market their availability primarily to the scientific community, and the majority of consumers are academics. Other highlights from the RAND study include the following:

Biospecimen Collection and Annotation

- All the repositories store paraffin-embedded tissue; many store frozen as well.
- All resources record pathological data. Some also record clinical data (taken from chart review).

Bioinformatics and Data Management

- Most repositories use bioinformatics systems as a repository management tool.

[1] Eiseman E., Brower J., Olmsted S., Clancy N., and Bloom G. (2003). *Case Studies of Existing Human Tissue Repositories: "Best Practices" for a Biospecimen Resource for the Genomic and Proteomic Era.* RAND Science and Technology. (August 28).

- None of the repositories allows public access to individual patient data.
- Every repository has its own network security system.

Privacy and Ethical Issues

- Types of consent vary: Full consent, general surgical consent, study-dependent consent, and exempt.
- Institutional Review Board (IRB) review for tissue use was required for all resources.
- In most cases, donors (patients and well volunteers) did not receive compensation.

3.2.2 Best Practices

Collection of specimens must first meet the needs of patient diagnosis, and procedures should emphasize quality and follow standardized protocols to the extent possible. New national standards based on best practices would be implemented as the basis for the new network processes and upgrading of existing resources. These standards would cover every aspect of the system—collection, freezing/fixing, storing, and shipping. Such standardization would improve comparative research, encourage the use of common technologies and testing methods, and make it easier to merge data and conduct multidisciplinary research. The development and implementation of national standards would require systems to assure compliance and oversight, and provide incentives to ensure success. Quality assurance (QA) policies and procedures for the NBN could be enforced by periodic site visits by a QA committee. Discussion of this subject is found in *6. Governance and Business Models*.

A potential model for creating and enforcing widely accepted standards and protocols nationally is that of the organ-transplant banks and their associated organizations. Although the purpose of these collection methods is primarily to maintain viability for transplant rather than to serve research, the cumulated wisdom of the organ-transplant banks may still provide insights that can be applied to the NBN.

The American Association of Tissue Banks (AATB) has developed a system for procurement, standardization, storage, and access to tissue, and has an existing structure and network design. It publishes standards to help ensure that the conduct of tissue banking meets acceptable norms or technical and ethical performance, and provides technical information that describes procedures to foster reasonable and responsible approaches to recovery, processing, preservation, and distribution of transplantable tissue.[2] The National Committee for Clinical Laboratory Standards is developing standards for tissue collections, and the International Society for Biological and Environmental Repositories (ISBER) is producing a comprehensive set of best practices for tissue banks that will cover collection, storage, retrieval, packing, and shipping, and treatment of human subjects (see Appendix K for the ISBER proposed best practices).

[2] For more details about the AATB, see *www.aatb.org*.

Another example of common practices is the set of standardized cancer protocols or formatting standards for 42 tumor types to be used for the evaluation of surgical pathology specimens, developed and published by the College of American Pathologists. The American College of Surgeons Commission on Cancer (CoC) will require that pathologists at CoC-approved cancer programs use the checklists in their surgical pathology reports starting January 1, 2004.[3] All operations at sites in the NBN and at the central site would be covered by standard operation procedures to be refined as the network is formed.

3.3 NBN System Requirements

3.3.1 Organization of a Tissue Collection Network

The NBN Design Team recommended that the national tissue resource be organized as (a) a decentralized network, possibly of nonprofit, tissue-repository organizations located near academic medical centers; and (b) a virtual repository with data networked across the nation. The approach should involve standardized acquisition, storage, and distribution of tumor tissue and other samples (including, but not limited to, healthy adjacent tissues, buffy coat cells, serum, plasma, and urine) by trained personnel whose primary responsibility is to meet NBN objectives. Tissue would be obtained from patients who had given consent to dedicated, appropriately trained personnel using standard protocols. The use of complex technologies that form the basis for molecular, genetics-based studies would dictate a priority on fresh-frozen tissue acquisition. The samples would be clinically annotated and "deposited" within a repository of both tissue samples (likely a distributed physical network) and derived data (likely a virtual network feeding a centralized national database). In addition, the NBN could contribute to the standardization of existing repositories, improve flow of specimens to researchers, broaden the specimen types that are collected, and develop data sharing processes and platforms to support more effective specimen distribution.

To help communication to broad audiences, the Design Team found helpful a vignette that follows the life cycle of a biospecimen through the NBN system for highlighting the salient considerations from tissue donation consent by the patient, through researcher access to specimens and data, and finally to efficacious treatment products. A summary of the scenario is included as Appendix L.

3.3.2 Users and Contributors of Biospecimens

The proposed NBN is designed to serve scientists and clinical researchers from academia, industry, and government. Access to both tissue and data derived from tissue should be broadly available. Extensive external specimen sharing would be required of NBN collection centers on a national scale. Additionally, a single, centralized policy for biospecimen distribution that guarantees timely peer review of research applications seeking to access the resource needs to be developed. Specimen allocation is further discussed in *6. Governance and Business Models*.

[3] Practicing pathology cancer protocols are available by body site at *www.cap.org/cancerprotocols/protocols_index.html*

It is imperative that the NBN have the full confidence and support of tissue donors, who are integral to the building of the NBN into a national asset. The fundamental principle of informed consent governs the use of materials and information that patients provide. These principles and associated requirements are discussed in *2. Management of Ethical and Legal Issues*. Experience has shown that patients generally are willing to donate their tissue for research.[4] If new drugs are to be developed, commercial entities (pharmaceutical and biotech companies) should be allowed full access to the resources of the NBN, and should not be prohibited from making a profit on their investments.

Given the importance of studying health disparities across socioeconomic groups, the NBN should be cognizant of fairly well-documented sociocultural factors that underlie the willingness of different racial, ethnic, and cultural groups to donate biospecimens and participate in health studies. For example, differences in willingness to donate blood, organs, and tissues for transplantation have had the unfortunate consequence of reducing the availability of appropriate organs for certain groups. If such factors also affect the willingness of some groups to donate tissues for research, the NBN may face a similar problem. There may be reluctance on the part of racial and ethnic minorities to donate to national repositories because of lack of trust in the system.[5] The NBN must make every effort to recruit a population of donors diverse in racial, cultural, and socioeconomic characteristics if it is to support research that will help to further elucidate genetic and environmental factors contributing to health disparities.

3.3.3 Tissue Sources

Eiseman and Haga (1999) estimate that "more than 307 million tissue specimens from more than 178 million cases are stored in the United States, accumulating at a rate of more than 20 million per year" (p. xvii). They place the estimate of available specimens collected specifically for research purposes at around 2.3 million. Sources of tissue may include academic medical centers, large community hospitals, and existing repositories. Community hospitals, the primary point of diagnosis for roughly 80 percent of cancers, would provide the surgical volume to be a major source of tissues for a national network. The community setting also reduces the potential selectivity that can occur in academic medical centers. Greater use of community hospitals would also allow researchers using the NBN to benefit by access to more demographically diverse patient populations.

Despite these many benefits for a national resource, community hospitals would need incentives and assistance to develop the experience, infrastructure, and understanding of research necessary to establish viable collection centers, particularly since these organizations conduct little or no research on tissue. It remains an empirical question whether deploying a group of experts to

---

[4] Ardais has well over 90 percent of patients providing consent to donate tissue for research. (Personal communication, A. Buckler, June 18, 2003.) Also, in 3,140 preoperative interviews with surgical patients in the United Kingdom, only 38 (1.2 percent) of patients refused to allow their tissue to be used for commercial research. Jack J.L. and Womack C. (2003). Why surgical patients do not donate tissue for commercial research: A review of the records. *BMJ*, Vol. 327, No. 7409: 262. Also see Appendix D of Volume II: Commissioned Papers of the NBAC Report: *http://www.georgetown.edu/research/nrcbl/nbac/hbmII.pdf.*

[5] Personal communication, L. Adams-Campbell, July 2, 2003.

community hospitals to help raise their levels of pathology and surgical expertise will greatly improve medical care in their communities, provide incentives for hospital and medical personnel participation, and allow the repository to reach the 80 percent of cancer patients who are served in community hospital settings. What is more clear is that the full engagement of community hospitals will require that human resources, coordination, and oversight challenges be overcome, and that appropriate incentives exist to encourage their participation. The NBN might consider financial support for infrastructure expansion, training, and salary support for new staff. These incentives would be tailored to the unique needs of community hospitals.

Another model is to have one or two larger community hospitals paired with a regional academic medical institution with experience in tissue resources. Institutions more experienced in research could assist in establishing the infrastructure and training to develop quality tissue collection programs. Criteria for selecting collection sites should include the following: Adequate specimen flow, good followup, willingness to comply with collection protocols, and diversity in geographic and other demographic characteristics of patients. It should be noted that diversity in the resource does not necessarily require that every site be diverse.

The NBN might consider recruiting other specimen collection sites through merit-based competitions based on NBN specifications (see discussion in *6. Governance and Business Models*). Any qualified organization would be encouraged to participate. Existing specimen resources will be eligible to join the NBN if they meet the NBN certification criteria, including rates of longitudinal data collection and diversity of tumor characteristics. Thus, existing resources such as the SPOREs, Clinical Cooperative Groups, and Cancer Centers could play an important role in the NBN. It will be important to develop incentives for these sites to participate in the NBN, as they provide high levels of expertise and they together can contribute to the diversity of the patient or donor pool in terms of cancer types, geography, and demography. The NBN could supplement current funding to encourage participation in broader tissue research efforts and provide a ready source of high-quality specimens, including frozen tissue, with clinical annotation, correlative/translational research results, and followup data. The NBN might also develop special access provisions for clinical trial-based samples, since these samples will likely have a unique importance to a particular study group. Restricted access may be beneficial and an incentive for clinical trial groups to associate with the NBN.

3.3.4 Specimen and Data Collection

The Design Team recommended that specimens from a broad range of cancer types be collected initially and that the NBN should be viewed primarily as a biospecimen bank with general guidelines on types and quantities of biospecimens available, rather than a prospective targeted collection source where specific types of specimens are collected only after a researcher has made a request for them. The goal should be to collect and bank the number and types of specimens needed by the research community, not just the ones easiest to obtain. Some cancers would be more commonly represented, at least initially, because they produce more abundant tissue, occur more frequently, or are less “damaged” by adjuvant therapies. It is possible to apply statistical models to determine the optimum amount of each tissue type, as collection for all types of tissue should not need to expand indefinitely. As it becomes very expensive to continue to collect specimens once a sufficient number has been banked to meet research needs, the NBN

should be cognizant of the number and amount of banked specimens needed to meet researcher needs. The size of tissue fragments available to the repository would vary with the type of tumor. The availability of metastatic and uninvolved tissue, expected specimen size, and most likely institutional source is shown for 15 cancer types in Table 3-1. This information, along with the incidence of specific cancers and annual deaths, can provide parameters for evaluating the adequacy of biospecimen inventories.

The suitability of the tissue samples for genomic and proteomic research varies with the organ system. Organs and their neoplastic derivatives that have high contents of proteolytic and other degrading enzymes such as pancreas, kidney, and liver, tend to degrade rapidly following ischemia and vary in their usefulness for research, especially for studies at the mRNA level. The usefulness for research of any tissue obtained from surgery varies with the time the organ is isolated from its vascular supply at surgery and how fast the tissue for research is cooled from body temperature. Nevertheless, many tissues remain useful for extended times and, even after death, remain potentially viable. For example, pituitary fibroblasts and hematopoietic cells have been grown in culture from tissue obtained from autopsies performed up to 48 hours after the deaths of patients.

Specimens not used by the NBN could be available for other collection efforts or would be discarded. Some experience suggests that limiting collection of tumors in the operating room (OR) to specific types has adverse effects on the overall collection process. On the other hand, the Cooperative Human Tissue Network (CHTN) has demonstrated that specimens can be effectively targeted to meet needs without compromising the collection process.[6] Tissue would be divided into sections and would be preserved in several ways (e.g., paraffin block: formalin fixation; paraffin block: ethanol fixation; frozen: optimal cutting temperature compound; frozen: no preservative) that would allow it to be used in a variety of techniques. A small portion of each section would be taken before preservation for QC purposes.

Most samples are (and likely will continue to be) gathered in the OR, or in the surgical pathology laboratory associated with the OR; consequently, surgeons and pathologists currently have initial control over the samples. The NDC TAWG suggested that the preferred scenario would have highly qualified, trained personnel (employed by the organization or consortium that would govern the new national network, working under the supervision of a pathologist) present in the OR/surgical pathology laboratory. Such dedicated staff would have the responsibility for monitoring the surgical schedule, arriving at the OR/surgical pathology laboratory at a preset time, taking possession of the tissue in an appropriate container, bringing it to the pathologist for evaluation, and overseeing initial processing in a standardized fashion. (Patient prescreening and consent would have occurred before the patient entered the OR.) A training and oversight program would be required for these individuals. This scenario would maximize the potential for following NBN data collection and associated annotation standards.

A less costly scenario that would rely upon independent surgeons and pathologists to comply with NBN collection standards was considered less preferable because of reduced NBN control over the handling of the samples and data. The TAWG also observed that surgeons and

[6] Personal communication, R. Aamodt, August 30, 2003.

pathologists have roles in patient care that might preclude their taking responsibility for the initial tissue processing. Surgeons are legally obligated to provide specimens to pathology for diagnostic evaluation, so pathologists at the submitting institution would be responsible for the initial evaluation of tissue and should determine which samples are available for research.

Table 3-1. Characteristics of Tissue Sample Collection by Type of Cancer

(1) Type of Cancer	(2) Incidence 1999 (new cases)	(3) Annual Deaths 2001	(4) Tissue Amt Available from biopsy or surgical resection	(5) Likely Institutional Sources	(6) Adjacent Normal Tissue	(7) Metastatic Tissue
Breast	201,500	41,844	0.2 gm Small	Community hospitals	Challenging with fewer mastectomies	Yes—from lymph node dissection
Lung and Bronchus	188,900	156,005	1-2 gm Medium	Large Cancer Centers	Alveolar—yes Bronchus limited	Unlikely
Prostate	195,400	30,714	0.2 gm Small	Widely available	Yes – but may have BPH	Unlikely
Colorectal	155,300	56,799	1-3 gm Med/Large	Community hospitals	Yes	Yes
Non-Hodgkin's Lymphoma	50,900	22,340	0.5 gm Small	Limited sources	Usually Not	Unlikely
Kidney	32,300	12,084	5 gm Large	Large Cancer Centers	Yes	Unlikely
Liver and Intrahepatic bile duct	13,000	13,263	1-2 g Med/Large	Large Cancer Centers	Yes	Unlikely
Ovarian	24,000	14,361	5 gm Large	Widely available	Omentum Only	Yes
Cervical	13,500	4,064	0.5 gm Small	Widely available	Yes	Limited
Pancreas	29,900	29,723	In resections this will be large	Large Cancer Centers, ~3,600 resection/yr U.S.	Yes—from resections	Very limited
Bladder	60,600	12,115	0.5 gm Small	Large Cancer Centers	Limited from cystectomy	Yes
Esophagus	14,000	12,509	0.2 gm Small	Large Cancer Centers	Yes—in small quantities with limited quality	Limited—needed for staging
Uterus	36,800	14,361	0.5 gm Small	Widely available	Yes—Widely	Yes
Stomach	21,200	12,340	1 gm Medium	Widely available	Yes—Widely	Yes
Oral cavity and Pharynx	29,100	7,638	0.5 gm Small	Widely available	Yes	Yes

Sources: [Col 2] based on incidence rates from Surveillance, Epidemiology, and End Results (SEER) registries and population figures in U.S. Cancer Statistics Working Group, 2002. [Col 3] Arias E. and Smith B., 2003, p. 15. [Cols. 4-7] Personal communications, W. Grizzle and S. Hewitt, July 2003.

The anticipated utilization of the specimens will determine how they need to be processed and what associated data should be collected. For each donor with a specific cancer, the NBN would collect high-quality specimens, sometimes with adjacent non-neoplastic tissue, and quality-controlled tumor RNA and DNA. These specimens would ideally be paired with serum, plasma, buffy coat, DNA, and urine samples. Quality-controlled clinical and pathology annotation, treatment, response-to-treatment, and outcomes data would be linked to the specimens.

It was suggested that a standardized kit be developed and utilized, with associated protocols for tissue collection procedures, equipment, and supplies. Similar kits, with established protocols and a procedure manual, are being used by the Biopathology Center, Children's Research Institute, in Columbus, Ohio, and at other institutions. This center provides centralized histopathology and tissue bank services for the pediatric division of the CHTN program, the Childhood Cancer Survivor Study, the Children's Oncology Group (COG), and the Gynecologic Oncology Group. Although the NBN can learn from best practices of child cancer groups, the Design Team suggested that NBN focus on tumors from adult patients. Pediatric tumors are already well covered by programs operated by the COG.

Pathologists are responsible for the diagnostic evaluation of patient surgical specimens and must first take what is needed for clinical (diagnostic) purposes. If enough tissue remains for donation to the NBN, the pathologist will evaluate representative sections of the samples designated for the repository for QC purposes. Excess tissue would be embargoed in the event that it is needed for additional diagnosis and patients have the right to revoke consent for unused samples.

Potential challenges in sample collection include the increase in preoperative neoadjuvant therapy, the increasing use of biopsy techniques that yield small specimens, and the impact of surgical clamping and the resulting anoxia on a specimen's suitability for advanced analytical techniques, particularly RNA analysis.

Tissue specimens would include snap frozen tissue and other necessary samples (including but not limited to matching adjacent tissues of grossly normal appearance, serum, plasma, buffy coat, and urine), with appropriate tumor representation and accurate depiction of patient background (e.g., age, ethnicity, race, gender, socioeconomic status, and geographical location). A portion of the specimens would be fixed in an appropriate reagent to allow RNA expression analysis. Other specimens would be snap frozen, thus preserving the ability to analyze specimens at the proteomic level. Tissue microarrays could be prepared using selected specimens. A model for this could be the Tissue Array Research Program within the intramural program of the NCI, which provides high-density tissue arrays for researchers throughout the country. Other forms of processing might be available. The goal should be that tissue is processed (annotated and frozen) within 15 to 30 minutes of removal from the patient. Tissue collection techniques and protocols would be validated in the demonstration project phase (see *8. Demonstration Project*).

3.3.4.1 Quality Control Considerations

Human tissues used in research must be of as high a quality as possible, with sufficient annotation to be able to determine appropriate uses. Initial steps should confirm the tissue type, check to see that the tumor is present and the percent cellularity, and look for signs of

degradation. QC should also, at a minimum, consist of preparation of a stained slide obtained from tissue adjacent to a section. This would be analyzed at the central site even if storage were maintained locally.

It is suggested that the NBN utilize an industry-accepted bioanalyzer to evaluate all specimens coming from the OR to determine the quality of RNA and protein. Specimens should be prioritizied based on the results of this testing. It should be possible to collect a fair amount of chip data using this technique, and provide a printout of RNA characteristics with each specimen.

An independent, offsite pathology review should also be a part of any QC procedures. QC procedures should confirm that involved tissue has the correct diagnosis, and that uninvolved tissue is actually free of the tumor or disease process. Samples should be validated histologically, as well as any molecular derivatives produced, prior to distribution. The successful production of high-quality tissue microarrays would provide one indication of the physical quality of the specimens. The ISBER Best Practices document provides an informative description of QC requirements (see Appendix K).

Ultimate responsibility for QA and QC activities would rest with the appropriate NBN entity, in coordination with the collection site (see also *6. Governance and Business Models*).

3.3.5 Annotation and Clinical Data

To accelerate drug and cancer therapy development, more complete datasets are needed. Requirements for data must be defined with input from researchers, as researcher needs should dictate the types of data that are most valuable to collect. Also, the magnitude of both the currently available and projected data, and researchers' ease of access, should drive the design of the NBN database; the needs of users for data should be anticipated and well served by the selected bioinformatics platform. Ideally, information to be tied to tissue samples via identifiers should include the following:

- Demographic data and social history (e.g., smoking, alcohol use), including familial cancer history
- Diagnostic and clinical information
- Pathology reports
- Initial staging procedures
- Tissue collection procedures
- Treatment data
- Information that could track the patient in the future for clinical outcomes.

Investigators need to know the pathologic diagnosis of the tumor and characteristics of associated tissues. Clinical annotation of specimens should ideally include site of origin (e.g., lung), primary diagnosis (e.g., adenocarcinoma), secondary descriptors (e.g., poorly differentiated), tumor size and stage, a copy of the blinded surgical pathology report, and any

additional diagnostic studies. For every case (multiple specimens), a base set of patient demographic and clinical data should be collected. The demographic data would ideally include age, race, ethnicity, sex, state of birth, state of current domicile, and urban/rural classification. Social history should ideally include detailed tobacco usage, detailed alcohol usage, and potential toxic (occupational, previous chemotherapy or radiation therapy) exposures. Clinical history also should include (if possible) familial histories of cancer, with detailed, more specific histories focused on each tumor type. Data collection should be standardized and QC enforced across biospecimen collection sites. It is recognized that such an extensive data collection will be costly.

Except for outcome data available from tumor registries and use of questionnaires, treatment and outcome data are extremely difficult to obtain and might not be available, even though these would be very valuable for research. New policies and protocols would be needed to facilitate the submission of these data, ensure uniformity, ensure patient privacy, and track treatment and outcomes. Tracking recurrence is difficult because patients often do not return to the same hospital for treatment of the second tumor incidence. Standardization, QC approaches, and monitoring would be required at all stages of the process. Additionally, the database system should be designed to grow and evolve as technologies advance and the information base grows. The system should also be capable of collecting longitudinal data, and the methods to be used to collect longitudinal data must be carefully planned. This is further discussed in *4. Bioinformatics and Data Management*.

Various groups that might offer models for best practices in annotation include the following:

- American Joint Committee on Cancer (AJCC) or American Cancer Society Tumor Registries and State registries
- Clinical Cooperative Group banks
- Children's Oncology Group (COG)
- College of American Pathologists
- Familial cancer groups/Cancer Family Registries program
- Shared Pathology Information Network
- Military or Veterans Administration (VA) systems
- Health Maintenance Organizations
- Single Nucleotide Polymorphism (SNP) Consortium
- Early Detection Research Network (EDRN)

Coding standards, databases, and communications protocols—key to making these data useful—are discussed in *4. Bioinformatics and Data Management*.

3.3.6 Advanced Analysis

It may be reasonable and desirable for the NBN to provide advanced analysis services to researchers. To respond to researchers' demands for the collection and cataloging of baseline

genomic data, it will be necessary to define the data that are in highest demand, how they will be accessed, and at what price. Possible advanced analysis services might include proteomic and genomic-based services (SNP profiling, mRNA profiling, proteomic profiling) and advanced tissue approaches (laser capture microdissection, tumor cell lines, and advance biospecimen production). A discussion of baseline genomic and proteomic analysis techniques likely to be employed by researchers is found in Appendix M.

Collection of genomic data, while highly desirable to researchers, would dramatically increase cost and management issues. The informal questionnaire administered by the NDC Research Team and the NCI at the AACR in July of 2003 indicated that approximately 70 percent of respondents expressed an interest in obtaining gene expression by DNA microarray (mRNA transcript expression profiles of about 30,000 genes) along with specimens; however, the fees respondents were willing to pay for these studies were insufficient to cover even 50 percent of the costs for such data (see Appendix E for complete results from the survey). The Design Team recommended that a professional survey of researchers be commissioned to further expand the findings of the AACR survey, with particular emphasis on the priority of researcher needs for advanced analyses and cost sensitivities.

3.3.7 Longitudinal Data

The usefulness of the biospecimen resource would be fully realized only after high-quality genomic and proteomic data are linked with clinical outcomes, events that may occur months to years after collection of the samples.

Longitudinal data useful for research should include specific treatment, time to progression, and survival outcome. Relevant biomarkers should be used, if available. Suggested data elements include primary chemotherapy (including chemotherapeutic agents); primary radiation therapy; primary surgical therapy; secondary therapy; concomitant drugs; narcotic pain medication usage; and date of death. Chemotherapy and radiation therapy would ideally specify doses, cycles delivered, dose modification, response, toxicity, reason for termination of therapy, and time to progression of disease. The NDC-NCI questionnaire results (Appendix E) suggest that researchers want all of these data elements.

Longitudinal data collection holds enormous promise for scientific discovery but is very resource intensive. The substantial costs involved in realistically collecting the most potentially useful longitudinal data must be anticipated. Statistical expertise for incorporating such diverse followup data into interpretation of laboratory results also will be needed.

There are several ongoing potential sources of longitudinal data that can offer opportunities and insights for the NBN, including the VA medical centers, clinical cooperative group banks (of which the COG is one), the National Cancer Database (NCDB), selected academic centers, and the Woman's Health Initiative. The COG is considered an excellent example of a successful model system, featuring a uniform treatment protocol, collection of biospecimens, and tracking for outcomes. Although the NCDB, run by the American College of Surgeons CoC, collects longitudinal data, it does not have a legal mandate to enforce cooperation, and its ability to share the data once it is collected is not clear. A partnership between the NCDB and NBN could

provide the NBN with outcomes data linked to tissue sample data, particularly if a legal mandate could be established to facilitate sharing.

Collecting longitudinal data presents a variety of challenges. Heterogeneity of current therapy, quality and nature of followup data, and changing therapies with availability of new agents have an impact on the utility of followup data. A periodic link must be sustained from the patient to the tissue sample, but patient privacy must be maintained. Experience has shown that collecting longitudinal data requires dedicated onsite data managers and a flexible information technology (IT) structure. Using existing resources within hospitals and clinics—for example, clinical trial data managers or state tumor registry personnel—may be possible. The Design Team recommends that an onsite data manager for each collection site be designated to maximize the collection of longitudinal and outcome data, as well as ensure the accuracy and validity of clinical annotation.

Epidemiologists have shown that it is possible to collect longitudinal data over a period of years with minimal loss to followup.[7] Individuals who are part of a study and know that they are part of a study do not mind being recontacted. The response rate tends to be high, and the loss to followup is low. The combination of active contact, change-of-address systems (operated by the Postal Service), checks on cancer registries, and the National Death Index means that only a small percentage is lost to followup. Sometimes individuals fail to respond at one contact attempt but resurface on the next contact. Additionally, acceptable loss to followup rates can be defined *a priori*. The NBN should build on current epidemiological techniques and expertise to maximize its success in collecting this valuable information. Additionally, a well-considered marketing campaign, coupled with a privacy-protected Web site, could encourage patients to provide the NBN with updated address information.

Some specific treatment, response, and time-to-progression information (or time between treatments, as a surrogate) might be obtained from chart reviews, although additional investigation may be necessary to determine more effective ways to obtain this type of

[7] A response rate of 90 percent was achieved for each biennial questionnaire of the Nurses' Health Study and Health Professionals Follow-up Study (Feskanich D., Ziegler R.G., Michaud D.S., Giovannucci E.L., Speizer F.E., Willett W.C., et al. [2000]. Prospective study of fruit and vegetable consumption and risk of lung cancer among men and women. *J. Nat. Cancer Inst.,* Vol. 92:1812-23.). In an article on colorectal cancer in the Adventist Health Study, followup using mailed annual questionnaires was completed for 97 percent of the participants (Singh P.N. and Fraser G.E. [1998]. Dietary risk factors for colon cancer in a low-risk population. *Am. J. Epidemiol.,* Vol. 148: 761-774.). One group (Flood A., Velie E.M., Chaterjee N., Subar A.F., Thompson F.E., Lacey J.V., Jr., et al. [2002]. Fruit and vegetable intakes and the risk of colorectal cancer in the Breast Cancer Detection Demonstration Project follow-up cohort. *Am. J. Clin. Nutr.,* Vol. 75, No. 5:936-43.) reports a study followup rate of 90.8 percent for their study cohort from the Breast Cancer Detection Demonstration Project. Another group (Goldbohm R.A., Van den Brandt P.A., and Dorant E. [1994]. Estimation of the coverage of Dutch municipalities by cancer registries and PALGA based on hospital discharge data. *Tijdschr. Soc. Gezondheidsz*, Vol. 72: 80-84) reports a followup rate of at least 96 percent for 6.3 years of followup based on data from the Netherlands Cohort Study. For the Iowa Women's Health Study (IWHS), response rates (of known living) after the 1986 baseline have been: 1987 (91 percent), 1989 (89 percent), 1992 (83 percent), and 1997 (79 percent). Thus, 12 years into the IWHS, and with information from the National Death Index, study principals are fully knowledgeable or in touch with well over 80 percent of the original cohort (Personal communication, J. Potter, August 4, 2003).

information. Survival data can be obtained from SEER or high-quality hospital or state tumor registries. Linking the NBN database to existing state tumor registry data has been considered, but may require solving additional privacy and IT challenges. Hospital and some state tumor registries can provide accurate data for patients who remain in the local area, and also provide information on who and what percentage of patients is lost to followup at that location. Because state and hospital registries vary in their followup success, and mobility rates of patients differ in various locales, the average stay in a medical system can be determined and the failure rate in longitudinal collection can be built into the NBN system expectations.

Why Are Longitudinal Data so Important?
Example: Colorectal Cancer

The current standard-of-care for patients with Dukes B stage colorectal cancer is surgery (alone). However, 20 percent of patients so classified go on to have recurrences, and may have benefited from adjuvant therapy; unfortunately, there is no currently accepted method for identifying that 20 percent. Longitudinal clinical outcome data associated with biospecimens from patients with colorectal cancer would be invaluable for studying the correlation between gene expression patterns and long-term outcomes. These data could lead to development of a test to more accurately stage these cancers, and thus to make more informed recommendations for treatment (surgery alone or surgery plus adjuvant therapy).

3.3.8 Storage

A decentralized network of tissue repositories would require solutions to logistical issues such as inventory control and ease of access. The design of the physical plants of the network would require highly standardized procedures and practices. These would include power outage/electrical backup, operation of liquid nitrogen freezers (including procedures to buy, store, and use liquid nitrogen; ample backup; and safety equipment), and shelving and materials for room-temperature storage. Issues such as heating/air conditioning and security systems, physical layouts, and geographical locations (e.g., seismic areas and local transportation centers) also should be considered.

Storing all frozen specimens in the vapor phase of liquid nitrogen freezers will ensure that physical and chemical reactions within tissue are slowed so that the specimens remain stable for use over many years, and biomarkers of interest to researchers preserved. If cryopreservation for viable cells is required, biospecimens should undergo controlled rate freezing to prevent damage by ice crystallization.

A high-quality inventory system using bar coding and an inventory database should be employed so that sample location can be tracked. Tracking and life cycle management of biospecimens is critical. Written procedures for all facets of facility management and security will be required.

The use of secondary storage sites is complex. A collection site would remove specimens from temporary storage, package the specimens, and ship them to a central site. Central sites would unpack the specimens, check identification of each specimen, store them according to

identification, and wait for usage. Some specimens should remain at the collection site as a redundant measure to avoid failures in storage. All participating tissue collection sites that possess the expertise to collect specimens according to NBN requirements would need to be trained in NBN storage methods. Storage methods would be validated in the demonstration project phase (see *8. Demonstration Project*).

3.3.9 Distribution and Access

A clear research need is timely and equitable access to biospecimens and associated data without undue administrative burden, as well as a prescribed mechanism for rapid turnaround of requests. Research would have to be supported through a uniform candidate biospecimen identification and distribution system, rather than one that requires the NBN to negotiate with a series of institutions. Distribution of scarce tissue resources should be prioritized through an evaluation of needs at the national level, and a peer review process should determine distribution.

To facilitate this, the NBN will provide access to tissue specimens through a Biospecimen Utilization Review Committee. Members of this committee would be recruited from the research community, advocacy groups, industry, and possibly government. The NBN Biospecimen Utilization Review Committee would utilize a peer review system to set priorities as to how the collected tissue should be allocated, and to guarantee fair and equitable access. Evaluation of researchers' needs and capabilities against competing demands for specimens would be required to assure maximal benefit of the NBN. The importance and quality of the laboratory studies should be commensurate with the value of the specimen, its associated annotated database, and followup data. As part of the peer review process, the Research Administration and Support Business Unit might consider factors such as the applicant's willingness to redeposit data and past collaboration with the NBN.[8]

Most investigators would probably use fewer than 100 specimens per year of one particular type of tissue. Some entities might use up to 1,000 samples per year (for example, a large pharmaceutical company or large academic medical center conducting many concurrent studies). All studies should justify the number of specimens required for the specific proposed study. Also, the Biospecimen Utilization Review Committee should determine if the tissue needed might be obtained from other sources (e.g., a prospective collection for which outcome data are not available, if the designated study does not require outcome data).

It is important to clarify that the roles of the NBN Tissue Utilization Committee and IRBs/industrial review boards are distinct. The IRBs provide assurances that human subject regulations and policies are being followed. The NBN Tissue Utilization Committee would look at the quality of the proposed research and determine its priority for access to specimens.

[8] See *6. Governance and Business Models* for further discussion about business units and their functions.

Deidentified data associated with the biospecimens would be made available to industry, academia, and government users through a transparent IT platform that is Web based, searchable, secure, and workable across virtually all systems (see *4. Bioinformatics and Data Management*.)

3.3.9.1 Shipping

Designated carriers who are reliable and who routinely ship human tissue samples would be used for all shipments from the NBN repository to the research users. Packages would be bar coded and could be tracked via software provided by the carriers, and the carriers would have acceptable procedures for resolution of shipment problems. The carriers would ensure that paperwork accompanying shipments would be standardized and compliant with regulations. Standard operating procedures would be developed for shipping and would be provided to members of the network.

Facilities applying to be part of the NBN would have to comply with regulations for storage and shipment of human tissue to domestic and overseas destinations. The International Air Transport Association (IATA) regulates overseas shipments, and the U.S. Department of Transportation, which regulates domestic shipments, has now adopted the IATA standards.[9]

3.3.10 Safety

Biohazards and other safety issues must be considered by personnel who collect tissues, by investigators who receive tissues, and by laboratorians who work with tissues. The NBN must ensure that persons who come into contact with specimens procured by its collection facilities are trained properly in how to handle potentially hazardous materials. In particular, all specimens must be handled as if infectious. It is an Occupational Safety and Health Administration (OSHA) requirement that organizations that employ persons handling human tissues establish a written safety program and a comprehensive training program to protect personnel from blood-borne pathogens.

To be an approved site within the NBN, a facility should demonstrate its compliance with the OSHA regulations—e.g., regulations for handling blood-borne pathogens (29 CFR Part 1910)—and state and local biosafety requirements. As part of this compliance, safety and emergency procedures, including a facility safety plan with standard operating procedures and standardized personnel training, would be developed.[10] This responsibility would probably lie at the Business

[9] Examples of general requirements for shipments and packaging of diagnostic specimens can be found at *www.olao.od.nih.gov/packaging_instr.html*. This Web site also details the requirements for packing instructions in accordance with the IATA dangerous good regulations. The Department of Transportation, U.S. Postal Service, and IATA are working to align their regulations to eliminate conflicting requirements. A new Federal Department of Transportation Regulation for Shipping of Medical Diagnostic Specimens was implemented on February 14, 2003. See the Texas Veterinary Medical Diagnostic Lab Web site (tvmdlweb.tamu.edu) for additional information. The complete regulation can be found at *hazmat.dot.gov/67fr-53118.pdf.*

[10] An example of such a document, a Bloodborne Pathogens Exposure Control Plan, prepared by the Division of Safety, Florida Department of Labor and Employment Services, to facilitate compliance with OSHA's blood-borne pathogens standard, can be seen at *www.cdc.gov/niosh/elcosh/docs/d0300/d000378/d000378.html.*

Unit level, with a corresponding QA matrixed responsibility at the NBN Operations Center level. Fulfillment would include site visits, the distribution of written and online materials, and training.

Organizations such as CHTN require investigators to sign an agreement that they will educate their staff about the proper handling of biohazardous materials, and sign an agreement that indemnifies the tissue procurement and distribution facility from any claims, costs, or damages resulting from the use of tissues provided by that facility. It is recommended that the NBN consider similar agreements with persons who will receive specimens.

3.3.11 Submission of Research Data

Creating a framework for sharing and comparing research results in a timely fashion would add substantial value to the NBN. Therefore, it is recommended that validated, investigator-derived data be returned to the NBN and linked back to original NBN tissue samples. Encouraging the entry of validated data in a standardized format (e.g., outcomes, possibly research results) back into the NBN system would create a rich, valuable, and unique resource available to all investigators, providing new possibilities for the corroboration of research findings across methodologies; the emergence of successful new therapeutic/preventive targets; and the acceleration of important scientific breakthroughs. The research database will also be useful for selection of related specimens remaining in the bank. The data would be able to be used repeatedly, unlike the actual tissues, and would serve as a stimulus for participation in the NBN, further enhancing the benefits for the scientific community.

The tissue resource and associated data would be made maximally useful if investigators were required to submit their experimental data to the database after a set period. The Design Team recommended that a multitier policy on submission of research data be considered, as whole-scale submission of data may not be acceptable for some researchers. However, multiple levels of acceptable validation exist among industry and academia, which may limit the usefulness of the database to some groups of customers. Therefore, researcher-submitted data, unless validated to the standards of all constituencies, may suffer from lack of utilization. Standards to ensure as much data validity as possible will be necessary. When such data should be submitted, and the appropriate level of data to be submitted (e.g., broad profiling genomic data versus focused pathway mapping studies) by various types of researchers (academic or industry) with sufficient validity standards should be determined by the Research Support Business Unit. The Design Team offered the following guiding principles:

- As close to practicable after publication, researchers would be encouraged and invited to submit data (especially DNA, RNA, and proteomic data) to the NBN.
- Researchers would be required to report all publications resulting from the use of NBN samples, as well as reference the source of the samples in their report (as is required by CHTN and other resource sharing facilities).
- If DNA, RNA, and proteomic data are published, the raw data must be available for public use, similar to publication standards set by major scientific journals.

Creation of such a data resource, while of great potential value, certainly presents many technical and policy implementation challenges. It will be necessary to create data standards to allow resubmission in a usable format into a centralized database. Fortunately, data and nomenclature standards are currently being created for many types of research results (e.g., Systemized Nomenclature of Medicine and AJCC for cancer staging; Minimum Information About a Microarray Experiment for gene expression data; and the common data elements created by the EDRN), and these standards should also be used by an information resource such as that suggested by the NBN. It will be difficult for the NBN to revalidate all submitted results, so steps must be taken to ensure submission of data that other researchers will find valuable and want to use. Researchers are often reluctant to submit data for various reasons, so incentives to encourage this submission may be necessary. Finally, it will take time for the NBN to address these challenges; consequently, the creation of a usable database with high value for researchers likely will occur later, rather than sooner. Specific recommendations about how this might be accomplished may be found in *6. Governance and Business Models*.

3.3.12 Incentives for Participation in the NBN

To implement the NBN, potential sites may need incentives to overcome possible institutional barriers. For example, they will need to adhere to strict policies, procedures, and standards; invest in new infrastructure and support (staff, equipment, training, etc.); and revisit the role of IRBs in the process. Incentives could be developed to encourage surgeons and pathologists to cooperate in increasing samples within the repository for research. For example, increased deposits of tissues could be rewarded with increased or priority access to tissue. Participating in the NBN, and thus having met certain quality standards, might make an institution's tissues more valuable for other research purposes. NBN participation—and the associated increased resources—might help existing repositories (e.g., SPOREs, clinical cooperative group banks, and cancer centers) to improve their capacity for their entire range of activities.

Another incentive would be provided by access to the NBN database that would be developed as part of this resource. Academic medical centers would want access for their researchers to use this new valuable resource, which would not otherwise be available. The data could be used repeatedly, unlike the actual tissue; thus, the data could serve as a stimulus for long-term participation.

Different incentives would need to be developed for community hospitals because their needs differ from those of academic centers. Absent resource constraints, community hospital pathologists are generally willing to participate for the satisfaction of contributing to the research enterprise. Providing a community hospital with funds to hire employees, such as technologists, would be likely to encourage their participation. Community hospitals and academic centers could be paired to share various resources. Alternatively, these hospitals could be tied in with clinical trial groups to involve them in clinical research.[11] Cost and reimbursements as incentives are discussed in *6. Governance and Business Models*.

[11] The Division of Cancer Prevention at NCI already has ties through the Community Clinical Oncology Program.

3.4 Summary of Key Findings and Recommendations

The NBN would be distinguished from existing resources for tumor tissue and other specimens by highly standardized procedures for collection, processing, storage, annotation, and distribution. The NBN would be developed to provide biospecimens and clinical information in compliance with Federal, state, and local regulations. In particular, the following should be noted:

- The NBN should be organized as (a) a decentralized network of collection facilities with regional storage, possibly of nonprofit, tissue-repository organizations located near academic medical centers and community-based hospitals that serve large and diverse patient populations; and (b) a virtual data repository networked across the nation.
- Best practices should be incorporated and/or developed for every aspect of biospecimen and data collection, processing, storage, and distribution in the operation of the tissue repository; and should be consistently applied through the use of standard operating procedures that would be monitored.
- Biospecimens and data should be collected from sources meeting NBN criteria, while applying standardized clinical annotation.
- Incentives tailored to each kind of source would be developed to encourage many entities to participate. Although community hospitals are the primary point of diagnosis for roughly 80 percent of cancers and could thus provide the needed volume, community hospitals would need specific incentives and assistance to develop the experience, infrastructure, and understanding of research necessary to establish viable collection centers.
- Well-trained personnel would collect tissue using NBN-provided protocols, applying standardized clinical annotation, with selected biospecimens undergoing advanced analysis.
- The NBN must ensure that persons who come into contact with specimens procured by its collection facilities are trained properly in how to handle potentially hazardous materials.
- Specimens from all cancer types should be collected (with matched normal specimens, whenever possible), but the NBN should be structured to provide the quantity and diversity of biospecimens required to meet researcher needs.
- Both fresh frozen and formalin fixed/paraffin embedded preparations should be used.
- A minimal dataset would be established for each specimen, with collection of additional longitudinal data for a high percentage of specimens, and provision of genomic and proteomic-based data.
- A professional survey of researchers should be commissioned to further explore researcher needs for advanced analyses and cost sensitivities.

- Distribution of specimens would be guided by a Biospecimen Utilization Review Committee peer review process that would evaluate researchers' needs against competing demands for specimens.
- It should be expected that validated, investigator-derived data derived from NBN resources be submitted to the NBN and linked back to original NBN tissue samples. An expanded dataset, created by the return of this experimental data to the NBN, could then be made available to all investigators.

Module 4

Bioinformatics and Data Management

4.

Bioinformatics and Data Management

The National Biospecimen Network (NBN) will require an information system to track biospecimens, support the collection and dissemination of key clinical and biological data associated with them, provide analytical capabilities for genomic and proteomic based research, and manage information needed for NBN administration. This module outlines the requirements and architectural issues for the NBN information system, and proposes approaches to establishing such a system using a combination of new development and adaptation of existing systems.

4.1 Introduction

The field of biomedical informatics encompasses the full range of technologies needed for clinical, biomedical, and genetics research, including applications for computational biology such as the study of probabilistic methods for gene analysis or techniques for studying protein folding. This module focuses on the particular requirements of the NBN for a multidimensional system that can capture and manage data serving a broad range of applications that begins at the bedside and progresses through the genetics laboratory.

The NBN's utility is maximized with a well-designed and powerful information system, and biospecimens will prove more useful for research when accompanied by annotation of relevant clinical data. By design, not only the quality but also the accessibility of the biospecimens and associated data must be considered to be of the utmost importance, as acquiring data without a way to use or share the information is pointless.[1] The Tissue Access Working Group and the NBN Design Team recognized that the development of the NBN information system as an open-access tool using scalable and extensible architecture is at the core of its efficient functioning and serves to buttress the NBN's purpose of supporting the integration and exchange of information for cancer research.

Identifying and implementing the best information technology (IT) architecture for the NBN will be a critical success factor and must be a priority in the NBN planning and budgeting process. This module reviews the architectural requirements for an informatics superstructure that supports searchable Internet-based approaches to an open-access but data-secure system; the collection and dissemination of biospecimens and data; reliable information exchange between NBN components and existing clinical information systems, subject to appropriate standards; data mining tools to facilitate scientific discovery; and management of the business functions of

[1] Maurer S.M., Firestone R.B., and Scriver C.R. (2000). Science's neglected legacy: Large, sophisticated databases cannot be left to chance and improvisation. *Nature,* Vol. 405 (May 11): 117-120.

the NBN. This system must include an optimal design for the protection of patient confidentiality that allows patients reasonable access to aggregated research results.

4.2 Bioinformatics Support of Existing Biospecimen Resources

There are a number of genomics-based resources that depend on robust IT systems. This section describes a few of the better known examples that can help guide the development of the NBN infrastructure or that have characteristics and/or the capability to cooperate in a linked system should the effort to build the NBN be focused on integrating existing (unconnected) systems. These are listed in alphabetical order, with more extensive discussions for some examples in the appendices.

Ardais Corporation

Ardais Corporation is a privately held clinical genomics company whose goal is to accelerate biomedical research by applying actual human disease, in the form of human tissue samples, as the discovery model in pharmaceutical research (see Appendix Q). The Ardais Biomaterials and Information for Genomic Research (BIGR™) System encompasses a unique repository, called the BIGR™ Library, with more than 170,000 research-quality tissue samples representing a broad diversity of disease. The samples are collected through the National Clinical Genomics Initiative, a strategic collaboration between Ardais and four leading U.S. medical centers.

Ardais has estimated that 45 percent of its staff are working in bioinformatics. As part of their comprehensive IT/bioinformatics system, Ardais identified a number of specific needs including:

- Structured collection of and quality control of clinical data
- Clinical data oncology specific to researcher needs
- Sample tracking
- Case linking
- Web-based access
- Experimental design
- Browsing
- Remote deployment
- Multisite coordination

To satisfy IT and bioinformatics requirements, Ardais created the BIGR™ System as a discovery platform for application to drug discovery and development. The system includes a centralized, shared clinical genomics repository that encompasses tissue samples, molecular derivatives, and associated clinical information accessed by an array of bioinformatics tools. The BIGR™ Library is a scalable, Web-deployable, Java-based system architecture that is managed from a central database, with strict controls based on individual user privileges. The Library is multi-institutional and will allow Web-based access to the repository on a researcher's desktop computer. Clinical and demographic data are gathered and reviewed to ensure consistency and

comparability, enabling researchers to navigate through a highly structured, consistent, and comparable library of materials and associated data. It is possible to search samples using patient attributes, diagnosis, tissue type, appearance, sample composition, pair-ability of diseased and matched normal from the same patient, in addition to viewing supporting data, including digital images.

First Genetic Trust

First Genetic Trust (FGT) is a business that develops IT solutions to address data, privacy, confidentiality, and ethical challenges in genomics and proteomics (see Appendix R). It is focused on supporting genetic research as a trusted third party; by providing a highly secure Web-based IT infrastructure for genetic banking; as a cornerstone of an integrated research solution for patient recruitment and informed consent; and for medical and genetic data acquisition, transfer, storage, and analysis. The patient, physician, investigators, administrators, and laboratory personnel can dynamically interface via the Web for patient education, information regarding the scope of the proposed research, and the consent process. The physician has similar access to aggregation of phenotypical clinical data and to obtaining clinical samples.

To address privacy and confidentiality protection, FGT has developed the enTRUST Genetic Banking System, with a Web-based architecture, and a highly secure, distributed genetic banking system. This system uses "Virtual Vault," Hewlett Packard's military-grade operating system, which leverages standard security technology for encryption and intrusion detection and exceeds both Health Insurance Portability and Accountability Act (HIPAA) and European Directive requirements for data collection, consent, accuracy, and security. A patient is assigned an encrypted electronic identifier that serves as a virtual private identity that is stored in one dataset; phenotypic or clinical information is stored in a second dataset; and genotypic data are stored in a third dataset. The three datasets are linked through the patient's virtual private identity.

FGT aggregates data via a Web-based architecture that interfaces with existing datasets. Data are accumulated, cleaned, aggregated, and stored in a repository. A common architecture in the system provides for distributed, centralized sample banking. The FGT research management tools are all Web-based. They include consent and reconsent modules (including information feedback to the patient, such as genetic counseling), clinical and genomic data capture, the ability to configure specific studies, sample logistics and banking, remote clinical data capture, study contract storage, and bioinformatics. Data representation standards support data exchange and mining, including aggregation of complex studies.

Since FGT is a trusted third-party banking technology provider, data access rights and policies are determined by the sponsor of the banking initiative. Public and "managed data access" models can both be accommodated. Histopathological image data are not currently available, but it is technologically feasible to provide them. The design protocol can be written to automatically aggregate clinical updates and secular outcomes.

NCI Center for Bioinformatics

The NCI Center for Bioinformatics (NCICB) has developed a comprehensive enterprise architecture for biomedical data management and several analytical tools to facilitate translational research, much of which could be leveraged for use in the NBN. Several of these components are described here. All are free, open-source software and are in production today (see *ncicb.nci.nih.gov*).

Bioinformatics Core Infrastructure. The NCICB infrastructure backbone is called caCORE; it consists of technology tools and services as well as an overall data management architecture (see *ncicb.nci.nih.gov/core*). It sets up common data elements and structured interactions between these elements through object models. caCORE has a common ontologic representative environment. It begins with a comprehensive cross-mapping of the various discipline vocabularies, first classified into a metathesaurus and then redistributed into a common framework, thus creating a National Cancer Institute (NCI) cancer thesaurus. From this thesaurus, 3,000-plus trial-structured data reporting elements have been extracted. Because the trial elements are based on a standard metadata repository, they can be defined, shared, and manipulated by the various communities.

This "multitier" system architecture is derived from modern software engineering designs and is implemented using the Unified Modeling Language (UML) and Java 2 Enterprise Edition. The features of the caCORE that make it a candidate for adoption by the NBN include:

- A multitier design that allows for the addition of multiple independent database sources that do not have to be physically colocated. This would enable a flexible deployment topology across multiple NBN sites.
- A modeling paradigm designed to be understood by nonprogrammers, which uses cancer bioinformatics infrastructure "objects" (CaBIO).
- Well-defined, documented application programming interfaces (APIs), which allow programming teams that did not develop the original architecture to use the full power of the system to write their own applications.
- Built-in support for many biomedical data types, including the human genome sequences and features, single nucleotide polymorphisms, gene expression patterns and sequences, therapeutic agents, clinical trial protocols, and many others.
- The Cancer Data Standards Repository, part of caCORE, which is a data element (metadata) database and tool suite. Such data elements are constructed from controlled vocabularies and thus provide for semantic consistency over time and across collection sites. The NBN could create and manage the data elements it needs for data collection and sharing using this resource.

Microarray Data Management and Analysis. The NCICB's caArray project consists of a minimum information about a microarray experiment (MIAME)-compliant microarray (caArray) database and tools for microarray data analysis and visualization. Originally built to support the NCI Director's challenge initiative, it is currently used in several NCI research programs. The

goals of the project are to make microarray data publicly available and to develop and bring together open-source tools to analyze and visualize these data. The caArray database connects to CaBIO, permitting access to a variety of genomic, cancer model, and clinical trials information. The primary interface to caArray is the Gene Expression Data Portal, a facility for uploading and retrieving microarray experiment data as well as performing some types of analysis (*gedp.nci.nih.gov*). The project has recently released a pathway visualization tool, which allows researchers to view the expression levels of genes on the array via visual pathway diagrams.

Two new applications will be released in September 2003: A gene expression data analysis workbench and a genomic viewer. The data analysis workbench is a desktop tool that will include a richer collection of analysis and visualization functions, including filters and normalizers, clustering tools, color mosaic images, dendrograms, and pathways. WebCGH is a Web-based tool for the analysis and visualization of Comparative Genomic Hybridization (CGH) data. The application will enable users to create whole genome plots using CGH data stored in the caArray database and focus on a chromosome or chromosomal region of interest.

Clinical Trials. The NCICB has constructed a clinical trial protocol management system for the Specialized Programs of Research Excellence (SPORE). The system supports administrative entry and tracking of trial protocols being reviewed and launched by the SPORE program. The NBN could conceivably leverage this system to manage research proposals for biospecimens.

Image and Other Data. Various information-capture modules are placed on top of this infrastructure (e.g., caEXPRESS, caIMAGE, caClinicaltrials, caModelsDB, and caLIMS). For example, caLIMS, the laboratory information management system (LIMS), describes how the data were collected. Image and pathology capture are managed by caIMAGE. A large collection of objects that describe clinical trials and the extant cancer model are provided by caClinicaltrials and caModelsDB. A prototype integration application to facilitate cross-disciplinary reasoning is the Cancer Molecular Analysis Project. It allows users to move from molecular profiles through clinical trials and is applicable to different fields. NCICB has constructed an image file management and delivery system called caIMAGE (*caimage.nci.nih.gov*) that could potentially be leveraged to support the NBN's histopathology image data needs. The NCICB system does not include a tissue management infrastructure for tissue inventory and control, as this is not part of its charge. Other organizations, however, may be able to provide off-the-shelf infrastructure for this purpose (e.g., Cooperative Human Tissue Network, Cancer and Leukemia Group B, Daedalus Software).

Security. The vast majority of the data in NCICB are already publicly accessible. For data requiring more secure access, particular authorization is required. The system uses Web-browser interfaces and an infrastructure distributed through a variety of formats that allows users to write external applications to "reach through" to the data. In essence, it is a repository that can be partially accessed with off-the-shelf technology.

Shared Pathology Information System

In 2001, NCI awarded two cooperative agreements to develop a Shared Pathology Informatics Network (SPIN), defined as a model Web-based system to access data related to archived human specimens at multiple institutions. Two groups were funded at $13.5 million over 5 years to work collaboratively on SPIN. Data will be accessed from existing pathology and other medical databases. The ability to automatically access information from medical databases is the first step toward the long-term goal of developing informatics systems to support NCI's efforts to improve researchers' access to human specimens and clinical data. The systems to be developed by the network will be able to respond automatically to authorized queries by identifying, obtaining, collating, and returning data for those cases that meet defined search criteria. Patients' names and other identifying information are to be encrypted or otherwise modified to protect patient privacy and confidentiality and to comply with applicable confidentiality regulations. In addition to improving access to clinical data, the system is expected to provide researchers with the means to quickly identify and determine the availability of archived specimens and data that meet their research needs.

Both projects are well into their second year, and notable progress has been made. Approximately 20 hospitals have now established peer-to-peer data sharing arrangements, and information on over a million pathology specimens is now accessible online. The current program will not actually make the specimens available to researchers, only information about them. SPIN is designed to provide a proof of principle. Actual specimen transfer would be the objective for a follow-on 5-year project under discussion.

The United Kingdom National Cancer Tissue Resource

One of the primary goals of the United Kingdom National Cancer Tissue Resource (NCTR) is to develop an information grid that can automatically and seamlessly incorporate all relevant data from each new patient into the appropriate database, input that patient's data into existing predictive models, and transmit that information to the clinician in the clinical environment (see Appendix O). As a first step, the NCTR proposed that the University of Cambridge, with the University of Leeds, the University of Glasgow, and the Peterborough Hospital Research Tissue Bank Network, lead the development of a pilot information system to include the development of an "informatics hub," which will build, maintain, and integrate heterogeneous and distributed databases (see Figure 4-1). Ultimately, these databases will include the specimen bank data, clinical and pathology data, clinical trials and outcome data, and research results. The base data would comprise a minimum dataset for the NCTR.

This hub will connect to present and future NHS information systems, including other clinical trials networks. There will also be a need to develop data mining tools that will synthesize these heterogeneous data sources into diagnostic and prognostic information on which clinicians can base decisions. The NCTR is working closely with commercial software developers to create research informatics platforms. Grid technology is expected to underlie the architecture, and it is recognized that there will need to be a security plan.

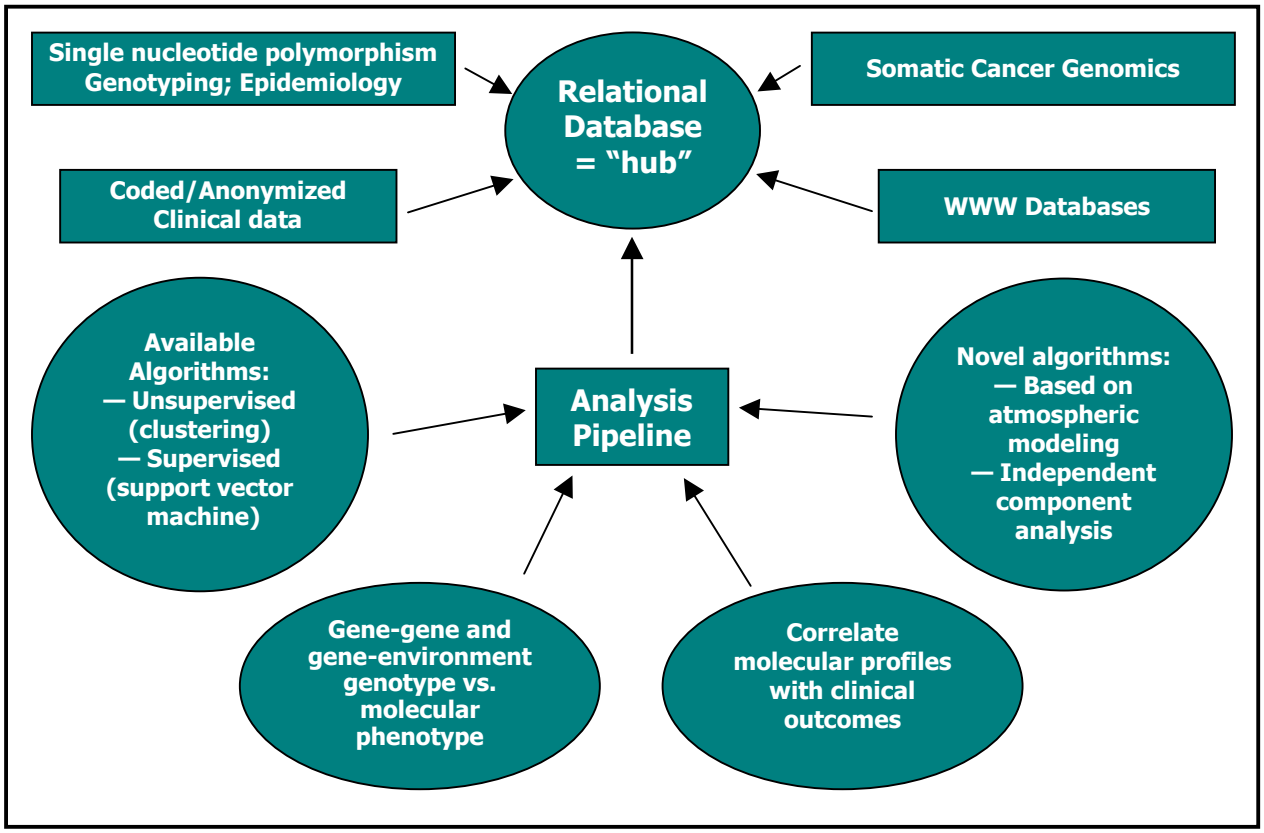

Figure 4-1. Adapted from the United Kingdom National Cancer Tissue Resource Bioinformatics Schema

4.3 NBN Bioinformatics System Requirements and Recommendations

There is an understandable tendency to start information systems from scratch; however, during the past 10 years the healthcare IT industry and the clinical and research pathology industries have made substantial investments in the development of sophisticated clinical and pathology information systems. Moreover, the Government and other organizations have developed information systems that may meet many of the requirements of the NBN information system, in whole or in part. Thus, it will be essential to identify the appropriate balance of "buy" versus "build." Truly critical system requirements should not be forfeited for shrink-wrapped solutions, and modifying commercial systems and paying ongoing licensing fees may dramatically increase the long-term costs of ownership. On the other hand, the NBN may be able to negotiate favorably with IT vendors if it can demonstrate sufficient volume of usage.

The buy/build/modify decision cannot be made until the architecture and requirements are specified. On the other hand, the requirements analysis needs to be informed by a thorough and up-to-date survey of existing systems, both commercial and government held, because requirements are not developed in a vacuum and features and functions that are deemed to be "absolutely required" or "desired" might be influenced by what is readily available (and the associated price). For example, 10 years ago "online instant access at no communications cost to the user" might have been deemed "expensive but desirable"; the Internet has made this a "no-cost must-have." Automated teller machines are an example of how data exchange standards can

revolutionize an industry. Similarly, consider how microarray technology has changed the requirements analysis from what it might have been a few years ago.

The size and complexity of the NBN information system will undoubtedly be impressive. There are clearly a number of challenges surrounding the building of big systems. The failure rate and cost of information systems are both exponentially related to size and complexity. Establishing an optimal information system management model is critical to the success of the NBN. The NBN Operations Center and the Board of Governors must understand that the informatics infrastructure itself will need to be adequately staffed and funded and ought to be supported as an informatics research enterprise, which is likely to promote quality improvements and attract the most talent and the associated research support that often accompany them.[2] It is clearly of paramount importance that a fully developed design be accompanied by realistic budgets, scalable implementation, and vigorous management of the social and political landscape via attendant authority.

4.3.1 Management of the NBN Information System

It was the sense of the Design Team that while general features and common data elements should incorporate input from diverse communities, information architecture design and implementation by committee does not work well for large software projects. It was recommended that a clearly empowered management model serve as the foundation of the NBN information system. Thus, the day-to-day management of the architecture design and development process must be guided by a strong and legitimized hand.[3] In essence, it is proposed that one individual have authority over project personnel, budget, design, and architecture and that this person be accountable to the governing principles of the NBN. It is important to establish authority early.

Although the decision-making authority for management and for design must be held by the system manager and architect, they will need to work closely on requirements analysis with:

- Bioinformatics counterparts who have developed biospecimen bank systems and who may be linked to this system directly or more loosely coupled via middleware in the spirit of grid computing
- Users (principally scientists and research administrators)
- Pathologists and laboratory scientists
- Biospecimen donors and users
- Pure data users
- Clinical information system experts

---

[2] See *6. Governance and Business Models* for discussion of the proposed management structure for the NBN.
[3] Brooks F.P., Jr. (1975). *The Mythical Man-Month.* New York: Addison-Wesley.

- Patients
- Bioinformatics experts with knowledge of similar systems

4.3.2 Design of the NBN Information System

It is recommended that the technical design of the NBN information system be created by the overall system manager's appointed architect. This section provides guidance on starting principles and will outline areas for which requirements will need to be developed. A detailed statement of the design is beyond the scope of this document.

4.3.2.1 Data Standards

The development of the NBN data standards could be challenging, primarily because the data standards and vocabularies that different user communities use are so heterogeneous (both within and between the groups), because the data structures and required reports are so inherently complex, and because of the need for longitudinal data. Any resulting models will have to take into account Health Level Seven (HL7) and UMLs and incorporate what is current at NCI. It is possible that some new data standards will need to be developed (hopefully as a variation of current ones). To illustrate, consider the wide variety of data representation required to encode such disparate data sources as encoded clinical pathology reports, text histopathology reports, demographic and clinical information, insurance records, clinical trials protocols, and microarray research results.

4.3.2.2 Minimum Dataset and Data Location

Assuring the availability of a minimal dataset should be an important element of the NBN system design. In addition, because modern computer communications systems make it possible to access data stored in many locations from many other (different) locations, it is unnecessary to build a single comprehensive database that would contain all NBN data. Nevertheless, connecting disparate information systems will likely be an evolutionary process, greatly facilitated by the more widespread adoption of data standards, such as the Systematized Nomenclature of Medicine (SNOMED), which has recently been adopted by the Department of Health and Human Services (HHS) as a standard.[4] A common lexicon of terms will not only create a new standard in health care and biomedical research but also enable researchers to mine databases with greater efficiency and added confidence that all available relevant information has been detected.

[4] Department of Health and Human Services. (2003). *HHS Launches New Efforts to Promote Paperless Health Care System.* Press Release (July 1); or *www.hhs.gov/news/press/2003pres/20030701.html*

4.3.2.3 Architecture

It is proposed that the architecture be a standards-based distributed system, with a central database at first that will evolve over time to be highly distributed, with the central database being limited to pointers and storage of any relatively stable, highly used administrative data. It was noted that sites that may be major data contributors (e.g., community hospitals) may be least equipped to handle a distributed data architecture.

The NBN architecture would specify standards at Open System Interconnection (OSI) Level Seven only and rely on HL7 standards wherever possible, developing new ones as needed.[5] Thus, the NBN information system would be built in a way that will not specify how a local site or Business Unit stores its data, what operating systems, hardware, or software it uses, or even what type of internal communications modalities it uses. Because it is certain that there will be data links to heterogeneous and distributed databases of ancillary clinical and pathological information, it would be wasteful over the long term to create duplicate local data stores and other capabilities except of selected, highly used data. On the other hand, for the first year or two, unless the system is well established, it is likely to be relatively centralized, and there will inevitably be some duplication.

However, it is strongly recommended that the NBN information system require that the data arrive in a certain format, and be packaged in a precise way, via the Internet. The NBN systems will know how to "open" the package and interpret the contents. Practically, this means that one of the first orders of business for the Bioinformatics Unit within the Operations Center will be the creation of a data model.

There may be times when the NBN will make recommendations with respect to software or hardware, based on experience. For example, it may learn that a certain software toolkit provides reliable data translations from Hospital Information System X to NBN standards. In addition, the NBN may develop specialized software (or even hardware) for use at local sites or within the Business Units, perhaps on a fee basis. Use of NBN-recommended or developed software, to the extent that it provides system-wide efficiencies, may be encouraged by incentives (e.g., a guarantee that the data will be formatted correctly or NBN will fix it; or perhaps a lower charge for another service). It is most likely that the NBN will develop software for local use if the NBN is asking for specialized data (e.g., on a rare disease) or for data for which standards do not yet exist (e.g., microarray data).

All communications will be Internet protocol-based, possibly with grid storage and job distribution capabilities strongly enabled; only later is it likely that the NBN will take advantage

[5] "Level Seven" refers to the highest level of the International Standards Organization's communications model for OSI—the application level. The application level addresses the definition of the data to be exchanged, the timing of the interchange, and the communication of certain errors to the application. The seventh level supports such functions as security checks, participant identification, availability checks, exchange mechanism negotiations, and most importantly, data exchange structuring (from *www.hl7.org*).

of parallel processing for advanced computation. The importance of a multitiered architecture has been emphasized by Working Group members.[6]

The system should probably be designed as a tightly coupled and extensible set of independent modules, each one associated with a specific set of functions. The design should fully support "Plug & Play" operations, where new modules (as well as new versions of existing modules) can be immediately deployed, as long as they interact with (and possibly extend) well-defined interfaces. Identifying these modules—and assuring their true functionality—will be one of the first jobs of the system architect. For example, all modules should be designed with robust and flexible key structures that allow them to be used in unanticipated fashions.

Vertical (functional) modules would allow for the easy addition of new functionality, e.g., a new data analysis algorithm or a new remote data capture module; horizontal (behavioral) modules can be used to drive the global behavior of the system by integrating the different functionalities via well-defined contracts and workflow patterns (Table 4-1). In practice, even the vertical functions, if properly developed, could be used throughout the system. For example, a strong education component could be included in the Specimen and Data Acquisition Business Unit as part of the informed consent process, in the Advanced Analysis Business Unit to bring a new technician up to speed, and in the Patient Relations Business Unit to help donors understand how their sample type is being used.

The NCICB platform, with its rich APIs and multitier architecture, could be integrated with other systems that provide the additionally needed functionality, such as a sample inventory management and tracking system. The NCICB platform currently does not include an impenetrable security and encryption mechanism; however, the Ardais model suggests that private patient data could remain at the primary specimen collection sites and that only deidentified clinical information needs to be transmitted into the biospecimen informatics network.

One important virtue of the modular approach is that well-designed (but extensible) security functions that protect confidentiality can be used throughout the NBN, enhancing overall security while allowing site-specific customization (e.g., e-consenting forms that might vary across hospitals). Essential functions are designed once to serve the NBN securely as a whole.

4.3.3 System Functions Supported by Bioinformatics

The NBN must rely on its bioinformatics system to support the overall integration and exchange of data for all of the other business units. Figure 4-2 provides a schematic view of the system functions mapped to the business functions outlined in *6. Governance and Business Models*, and this section provides a brief discussion of these system functions.

---

[6] Additional information on grid architecture may be found at the following Web sites: *www.grids-center.org/grids/grids_primer.asp; https://gsd.msfc.nasa.gov/FD40/papers/SpaceOps2002/spaceops02_p_t2_83.pdf.*

Table 4-1. Candidate "Plug & Play" Modules

Horizontal Modules (local functions)	• Authentication and authorization • Sample banking • Case report form design • Cohort/study management • Informed consent management • Data analysis • Reporting • Education • Questionnaire • Data warehousing	• Sample tracking • Remote data capture • Informed consent form design • Recontact • Clinical stratification • Knowledge management • LIMS integration • Study enrollment • E-signature • Genetic data banking
Vertical Modules (global behaviors)	• Workflow management • Study participant withdrawal • Role assignment	• Sample request management • Verification of data upload

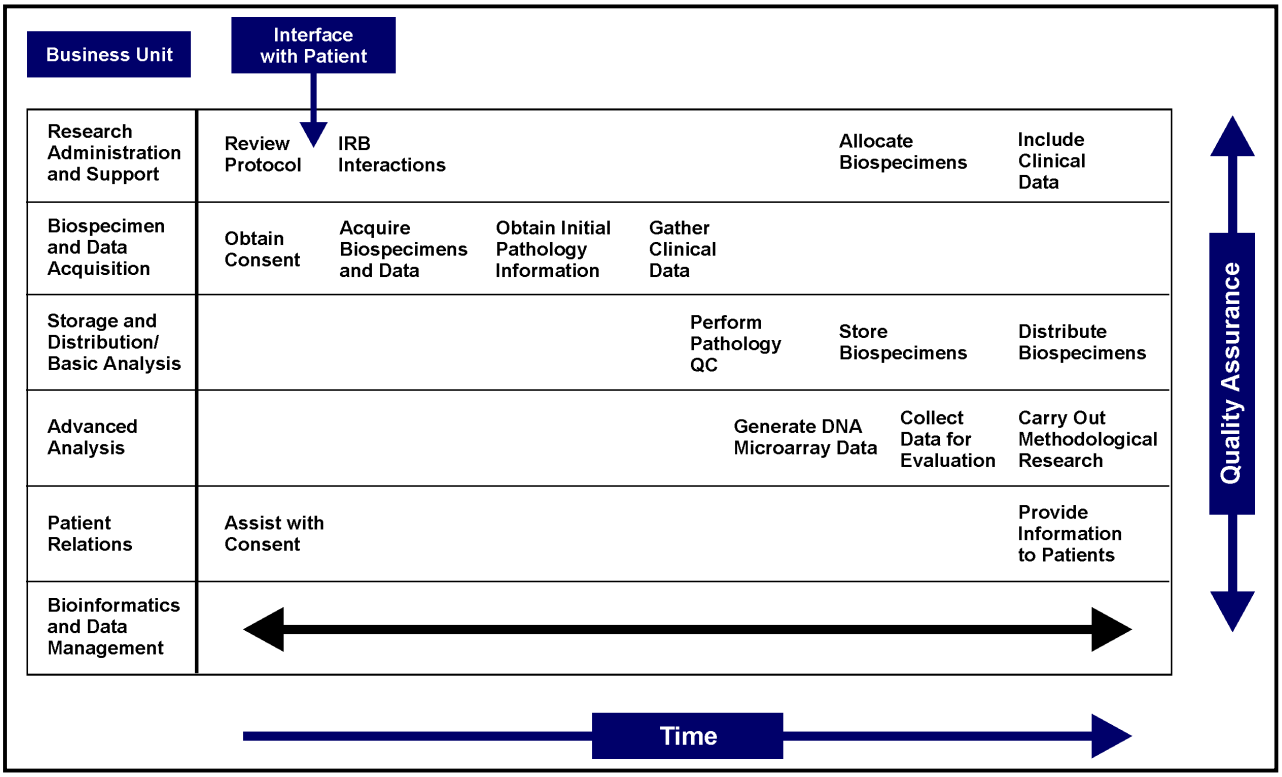

Figure 4-2. Mapping of Business Units and Their Key Functions to Bioinformatics

4.3.4 System Functions Supported by Bioinformatics

The NBN must rely on its bioinformatics system to support the overall integration and exchange of data for all of the other business units. Figure 4-2 provides a schematic view of the system functions mapped to the business functions outlined in *6. Governance and Business Models*, and this section provides a brief discussion of these system functions.

4.3.4.1 Research Administration and Support

Managing the research process will be greatly facilitated by the application of bioinformatics at every level. Tasks will include:

- Developing an e-consenting and reconsenting function, which will require the development of data standards for consenting information and associated electronic signature
- Developing a database of research activities, researchers, and grants creating a Web site about the NBN and its capabilities
- Administering the NBN Biospecimen Utilization Review Committee to provide equitable access to biospecimens and data
- Developing approaches to report activities to the NBN Operations Center and the Board
- Facilitating the identification of candidate specimens via user-friendly interfaces (e.g., users should not need to know Medical Subject Heading terms when looking for data about specific types of tumors)
- Developing a mechanism where data or analyses derived from NBN specimens can be added to the system for the benefit of future research

4.3.4.2 Biospecimen and Data Acquisition

The system will be required to manage the data for acquisition, basic analysis, storage, and distribution. In particular, the system will need to be able to identify biospecimens; add/delete/modify information; and track their location, availability, size, state (fixed, frozen, or both), and many other factors to be defined as NBN standards in *3. Biospecimen and Data Collection and Dissemination*. To support those broad activities, specific functionality will include the following:

- Informed e-consent and reconsenting management, tying the specific elements of the consent to the actual samples. Computerized informed consent forms should be defined.
- Integration with LIMS for tissue storage, retrieval, transformation, and shipping/receiving tracking
- Management of pathology reports and other longitudinal clinical data, including coding reports according to NBN standards, and ensuring that a true longitudinal picture of the patient can be created and studied

- Support for representing and monitoring the initial staging procedure and the standards for that procedure
- Tissue preparation (technology application) details and output
- E-signed workflow management to ensure that appropriate procedures are followed
- Developing a mechanism to communicate results to clinical providers

4.3.4.3 Storage and Distribution/Basic Analysis

It is anticipated that there will be regional storage and allocation of specimens, and that at these sites some preliminary analyses will be performed. At a minimum, systems will need to be created to:

- Receive, store, retrieve, and ship samples, and provide confirmation of receipt
- Manage inventory
- Conform with International Air Transport Association regulations

4.3.4.4 Advanced Analysis

A large amount of semistructured and unstructured bioinformatics data is expected to be generated from microarray analyses and other technologies. Important tasks in the bioinformatics arena will include the following:

- Implementing and enhancing existing standards for gene expression microarray data, tissue microarray data, and proteomics data—probably by participation in a larger process—will be an important part of the bioinformatics work of this project. The MIAME standards are Object Management Group approved, and several tools exist for creating XML documents for data exchange from a MIAME-compliant database.[7]
- Establishing detailed specifications for how to add or link these kinds of data to the NBN system and update, delete, and archive data will be required.
- Developing integration with LIMS or LIMS-like systems, via Logical Observation Identifiers Names and Codes (LOINC)®-based HL7 methods, will also be an important consideration.[8]

---

[7] Tissue Microarray Standards have been published recently. See Berman J.J., Edgerton M.E., and Friedman B.A. (2003). The tissue microarray data exchange specification: A community-based, open source tool for sharing tissue microarray data. *BMC Med. Inform. Decis. Mak*. 3:5.

[8] The LOINC database provides a universal code system for reporting laboratory and other clinical observations. Its purpose is to identify observations in electronic messages such as HL7 observation messages, so that when hospitals, health maintenance organizations, pharmaceutical manufacturers, researchers, and public health departments receive such messages from multiple sources, they can automatically file the results in the right slots of their medical records, research, and/or public health systems. For each observation, the database includes a code (of which 25,000 are laboratory test observations), a long formal name, a "short" 30-character name, and synonyms. The database comes with a mapping program called Regenstrief LOINC Mapping Assistant (RELMA™) to assist the mapping of local test codes to LOINC codes and to facilitate browsing of the LOINC results. Both LOINC and RELMA are available at no cost from *http://www.regenstrief.org/loinc/*. The LOINC medical database carries records for

- Developing a mechanism where data and/or analyses generated from NBN biospecimens by one user can be archived and shared with other potential users.

4.3.4.5 Patient Relations

It is to be emphasized that the NBN will maintain a firewall between the donor's identifying information and research materials. The bioinformatics system will act as an honest information broker. If patients are supposed to be able to request that their unused specimens be withdrawn from the repositories, then the NBN must record the specimens in a way that they can be linked to the original donor. If research results have implications for care, the *community* of NBN donors will be contacted, but never individuals (see *2. Management of Ethical and Legal Issues*). Thus, the system will need to provide an effective way to communicate key research findings to patients (biospecimen providers) and their families (see also *5. Communications*). This will involve:

- Developing a way to communicate to patients (and their families, as appropriate) findings with implications for treatment and counseling (genetic, participation in trials).
- Developing education and outreach materials for all the constituencies, including study participants, investigators, and the public at large.

4.3.4.6 Bioinformatics and Data Management

The NBN must be able to manage the information system tasks that cut across all the business units and that will be a shared (virtual) resource. The Bioinformatics and Data Management Business Unit supports a series of very general information system tasks that are discussed here.

Reporting. The design of the reports ultimately drives all system requirements for database applications. The NBN data architecture must allow the ability to:

- Facilitate researcher access to information.
- Enable or support the creation of longitudinal "virtual" studies to follow a cohort of patients through clinical outcome. (This is useful for early testing of a biomarker or diagnostic tool, but it requires modestly sophisticated information. Validation may require more sophisticated information because results might ultimately be tested in a clinical trial).
- Permit operational and management reporting.

>30,000 different observations. LOINC codes are being used by large reference laboratories and federal agencies, e.g., the CDC and the Department of Veterans Affairs, and are part of the HIPAA attachment proposal. McDonald C.J., Huff S.M., Suico J.G., et al. (2003). LOINC, A Universal Standard for Identifying Laboratory Observations: A 5-Year Update. *Clinical Chemistry,* Vol. 49:624-633.

Database and Data Model. It is likely that there will be a central core data facility, with a standard dictionary. SNOMED is a widely accepted standard, making its use within the NBN a yet more plausible choice. The NBN may have some dictionary requirements that go beyond SNOMED, and it is recommended that, whatever auxiliary functions are developed, the NBN dictionary cross-walk with other dictionaries. Similarly, the data system will need to have developed associated business logic and development and reporting tools. Parts of the system will be distributed. The distribution structure will evolve as technologies change and clinical data become more available (and more massive).

Candidates for some local stores include: Selected clinical (patient) data, trials data (research on these and related specimens, including data and bibliographic), and analytic data (histology and histochemical, microarray, and other genetic and chemical analyses). A detailed data model will need to be developed that includes at a minimum existing clinical, public health, insurer/Centers for Medicare and Medicaid Services, pathology, histology (possibly image), research, and bibliographic information. The model will have to allow patients to be followed longitudinally and linked to family members. It may also be desirable to collect data related to specific ethnic groups. It remains to be determined whether animal data related to the area of research should also be connected. It will be important to determine a hierarchy of information that would be required in a first version versus what is desired in what will probably be a series of later, more sophisticated versions.

Security/Controlled Access. The bioinformatics system should have strong authentication that supports the confidentiality and access rules determined by the governance body (which may in turn be advised by confidentiality and legal experts). Differential levels of data identification will be maintained in parallel, offering varying levels of access depending on the type of user (e.g., system administrator, NBN staff, researchers, patients).

User Support. User support will be a priority. There will be many types of users including researchers, the other business units, NBN Operations staff, and informatics staff at sites. User support will include helping with implementation and establishing minimal configurations (including needed communications infrastructures), updates, and user help. User support may also include a Web site, automated voice response, manuals, a help desk, and help with remote configurations. The system will only be as good as the support. The NBN will need to consider IT staffing at selected sites at a level commensurate with that of some private sector firms.

4.3.5 Quality Assurance

Demonstrating that quality measures have been set and attained (or that remedial actions have been taken) will be one of the most challenging aspects of both the program management and the bioinformatics tasks. Developing quality measurement tasks will include:

- Working with the QA Unit to define measurable objectives and pertinent regulations and procedures that the NBN must meet in supporting the bioinformatics infrastructure

- Developing an executive information system that reports on how each of the business units is doing in meeting their scientific, financial, and management objectives, and ensuring that the system is adding value
- Assuring regulatory compliance

4.4 Summary of Key Findings and Recommendations

The NBN will require an intelligent information system to track biospecimens, key associated clinical and biological data, and the information needed for research administration. The development of the system will be a major project that must be done simultaneously with the development of the NBN as a whole, rather than a being a standalone or later add-on. It will be essential to identify the appropriate balance of "buy" versus "build." Key features of the NBN information system are as follows:

- While recommendations about the general architecture and common data elements require broad input, to maximize efficiency, a central architect should be designated to build and manage the bioinformatics infrastructure. It is proposed that this central architect have authority over project personnel, budget, design, and architecture, and also be accountable to the governing principles of the NBN.
- The NBN bioinformatics system should be standards-based (e.g., SNOMED, HL7, or MIAME for data; Internet for communications) to enable data and information exchange among system components and the researchers who use them. In designing the bioinformatics system, a standards-based approach will allow flexibility to employ individualized approaches, while reducing the difficulty involved in developing a comprehensive system that links diverse components of the NBN. The NBN should not have specific requirements by the operational units for databases or hardware systems.
- Assuring the availability of a minimum dataset and location should be an important element of the NBN system design.
- Modularized components ("plug and play") will be used (and developed) to the extent possible, allowing the NBN data architecture to build upon best practices from other repositories.
- Critical benchmarks of success for the NBN data system include ease of data entry and retrieval, highly responsive user support, and commitment to the NBN Quality Assurance process.
- While Bioinformatics and Data Management will represent one of the business units of the NBN, it will also traverse and integrate all the business units and will be a shared (virtual) resource underpinning all of the NBN operations over time.
- Development and implementation will start with a centralized database, and then move to more a decentralized yet interconnected model as the system matures, capabilities are strengthened, and requirements clarified.

Module 5

Communications

5.
Communications

This module discusses an overall approach to developing a broad National Biospecimen Network (NBN) communications effort to attain support from the public, patients, and the research community. For the purposes of this discussion, NBN communications refers to all efforts—education and training, outreach, and public relations activities—to market the NBN to its key constituencies and to develop and maintain clear communication with all participants in the system.

5.1 Introduction

The success of the NBN will depend fundamentally on its ability to engage the research and clinical care communities and the public from the very outset of the project. Open and effective communication with all parties who might be involved with the NBN will be a critical tool for engagement.

Members of the National Dialogue on Cancer (NDC) Tissue Access Working Group (TAWG) recognized the importance of educating stakeholders as a means of encouraging participation in a national biospecimen system. Education about privacy, confidentiality, and other ethical issues should be an integral part of the overall master plan of a national biospecimen resource. Information about tissue banking and the NBN informed consent process must be well communicated to all participants, including patients, institutional review boards (IRBs), and professional associations. With effective education about the program, it will be possible to advance with public, professional, and patient support behind the issues to overcome sociopolitical barriers. The educational needs of each of the stakeholders will be different. These needs will drive the development of an overall education program about the national resource, and the separate components targeted to each of the major stakeholders.

5.2 Background

A well-planned, comprehensive communications strategy could effectively overcome barriers to support by the various groups from whom participation in the NBN is most highly sought.

This module examines:

- The nature of potential barriers to the collection and use of biospecimen tissue and related information
- The components of a successful communications strategy that could overcome these potential barriers
- Specific recommendations for how the NBN could develop a communications strategy, starting at the earliest stage of NBN planning, and maximize visibility and use of the resulting valuable resource

5.2.1 Considerations in the Acquisition and Use of Tissues

With better supported biospecimen resource sharing facilities, researchers will be better able to obtain appropriate biospecimen samples and related information so critical to scientific discovery. Researchers may fail to use a resource for a number of reasons. The usual and most obvious reason is that they are not aware that it exists. However, other more subtle factors may be at work. The costs may be prohibitive, the number of available samples may be inadequate for their research needs, or the samples may not be sufficiently annotated or linked to appropriate clinical data for their research needs.

On the supply side, most patients are willing and eager to participate, but may find few organized opportunities in the community hospital setting. Potential tissue donors should have ample opportunities to learn about the importance of tissue in clinical research, and the steps to take in order to donate; and to receive assurances that they will be apprised of the outcome of the research and about privacy and confidentiality issues.

A broad and effective NBN communications plan will help overcome these possible barriers by helping to manage expectations and correct misperceptions, and by fully acquainting all parties with NBN policies and procedures. In the past, efforts like the NBN may have been less than successful because communications planning was not considered a key part of the development process and start-up effort. Communication (particularly to patients) has not been a major priority for many existing resources.[1] By contrast, communications planning for the NBN will begin with the earliest planning phase and will be carried out in conjunction with the overall development of the project and at the Operations level (see *6. Governance and Business Models*). Communications, at every stage, will reflect other policy decisions taken by the NBN.

5.2.2 Multiphase Approach to the NBN Communications Efforts

The NBN tissue bank will involve a multistaged process of tissue collection, pathologic evaluation, preparation, storage, retrieval, and distribution for use. Each of these stages will have a different set of communications issues, key players, and desired outcomes. NBN communications programs will handle each stage as a discrete effort, with its own strategy, message, and evaluation mechanism. One of the first steps in developing an NBN communications plan will be a careful needs assessment for each stage of the process and for each stage in the development of NBN. The NBN communications activity will be an ongoing process. It will not be sufficient to develop a single campaign to alert researchers to the availability of a resource. Effective communications planning will have to encompass both the initial education campaign and ongoing education/outreach and public relations efforts.

The NBN will employ a well-established, evidence-based model in planning the NBN communications activity. The model, which has been used successfully in planning and executing large-scale communications efforts surrounding health and social issues, is broadly

[1] Eiseman E., Brower J., Olmsted S., Clancy N., and Bloom G. (2003). *Case Studies of Existing Human Tissue Repositories: "Best Practices" for a Biospecimen Resource for the Genomic and Proteomic Era.* RAND Science and Technology. (August 28).

referred to as social marketing. The approach is based upon classic marketing principles, which place an emphasis on careful preplanning, based on effective research. By using up-front research to identify specific targets, planners can divide potential audiences into subgroups. Planners can target specific messages to a particular group, using media and channels known to be effective in reaching that audience. This approach avoids the common mistake of developing an expensive outreach/education effort, only to discover that the messages chosen have no resonance with the target audience, and/or that the methods used to communicate the messages do not even reach the intended audiences.

The social marketing process is systematic and continuous; research-based decisionmaking at every phase feeds back into the process. There are seven general stages to the process, as depicted in Figure 5-1.

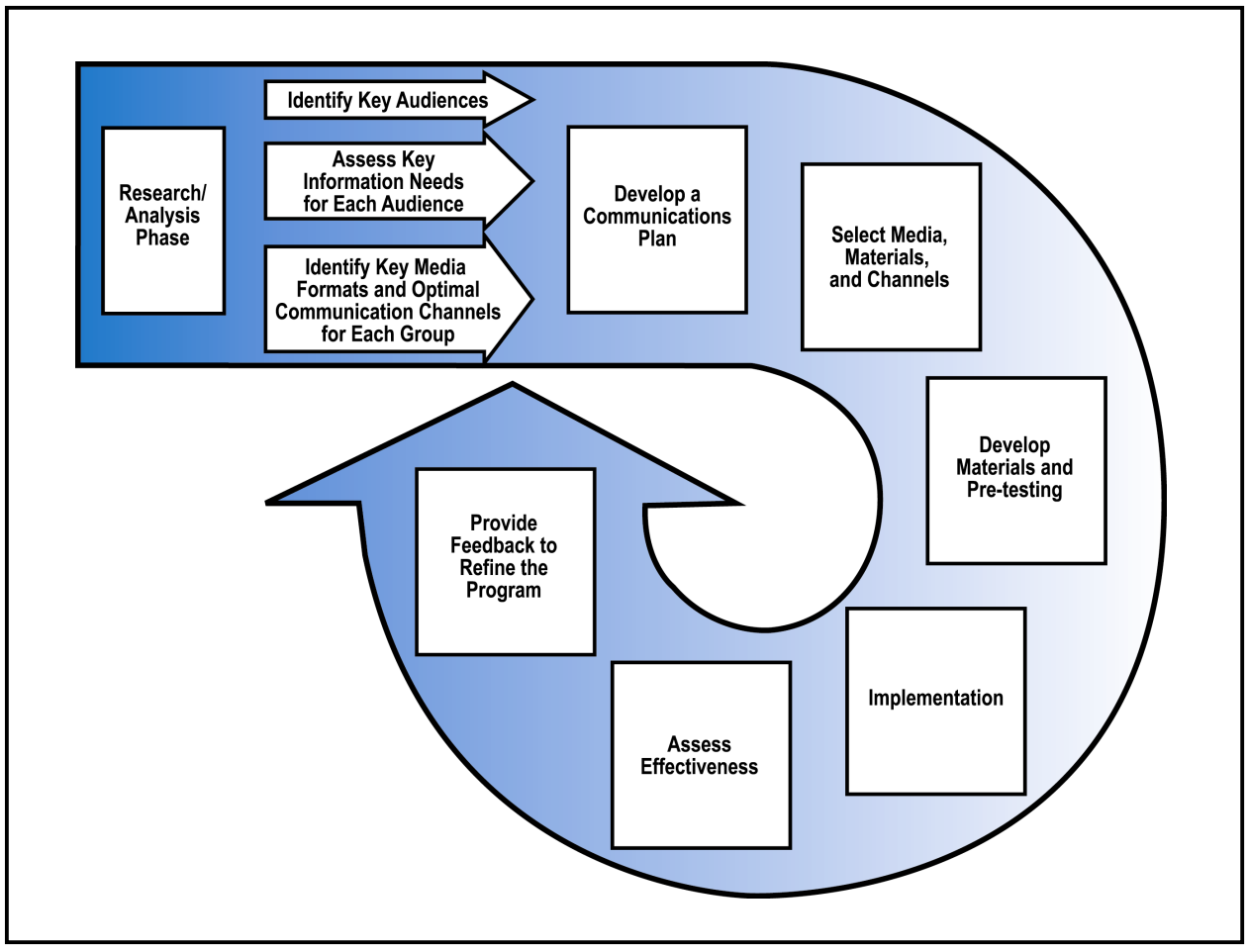

Figure 5-1: Evidence-Based Process to Develop a Communications Plan for the NBN

5.2.2.1 Analysis

The most important stage of the process is the preliminary analysis. In this early phase, the NBN will gather information to address the following questions:

- What is the NBN trying to communicate, and to whom?
- Who needs information about the NBN, and what specific information do they need?
- When is the information needed, and in what context?
- What barriers exist to communication?
- What are the optimal ways to reach various constituents with respect to both the type of media and the channels through which the media will pass?

Patients will have different needs at different times in the process. For example, potential donors will need information about research and the NBN in general, and about the benefits of participation. They may need additional information about how to donate specimens and about their rights as donors at the time of surgery, and again after their tissue has been banked. The planning process must take into account these disparate needs and their timing, and should develop a systematic strategy that meets the needs effectively and efficiently.

5.2.2.2 Planning and Strategy Development

The communications plan, developed as a working document, should present the background issues and provide direction for how the communications strategy will unfold. This plan should discuss the key audiences and delineate the specific information needs and interests of each target group, the ways in which each group typically obtains information, and the type of language most appropriate for each group.

At this stage, program planners also will build in evaluation steps to ensure that the various activities are having their desired effect and that resources are being used most effectively.

5.2.2.3 Selecting and Developing Channels and Materials

Once the preliminary communications plan is in place, planners can begin developing the education programs. Information needs that were identified during the analysis phase will be converted into specific messages. Planners will select the types of materials and the channels best suited to both the audience and messages to be communicated. If preliminary analysis indicates, for example, that researchers need training in the use of the NBN resources, planners will consider the best format available for conducting such training. They might use any of the following techniques:

- Formal continuing education courses, involving paper-based instruction, CD-ROM, or a Web-based system
- Informal training through articles in professional journals
- Training sessions at professional meetings

The NBN planners will select those channels that are most likely to have the desired outcome and reach the optimal number of end users at the lowest cost, and then develop the materials consistent with the research findings. The materials will be tested with the target audience to ensure that they have their desired effect. Based on the results of the testing, the materials can be modified to improve effectiveness.

5.2.2.4 Implementation

After the materials are ready for distribution, the education effort will be implemented. It will be important to monitor the program to ensure that it is developing as planned. Process monitoring may include, for example, assessing whether distributors of materials have received them, whether public service pieces were placed as planned, and whether posters appeared in correct locations.

5.2.2.5 Assessing Effectiveness and Providing Feedback

After the education efforts have been in place for some time, planners will evaluate whether the program is having the intended results. The evaluation method will depend, in part, upon the desired objective (for example, if the program were designed to increase the public's knowledge of the NBN, or if the program were designed to increase patient participation rates). The results of the evaluation efforts would be fed back into the communications planning process, so that refinements to the approach can be made.

5.2.3 Developing the Communications Plan

An effective way for the NBN to develop the communications plan is to obtain the services of a professional communications firm to assist with formalizing and implementing the NBN communications plan. With the completion of the preliminary work conducted in the planning phase, the NBN will have created a document that could be used in developing more specific tasking plans. There are a number of available contracting vehicles specifically designed to obtain communications-related services, which the NBN could use to procure the desired services.

5.2.4 Key Issues to Consider in the Communications Planning Effort

The NBN Design Team raised a number of issues that will need to be addressed in the development of the NBN. Many of these may be effectively resolved through the communications activities of the NBN.

5.2.4.1 Communicating Research Information to Patients

Patients are increasingly aware that test and research results from their tissues might be used to develop new therapies and identify risk factors that may be useful for their own disease. Returning research results to patients, however, raises major concerns. Patient/physician reliance

upon unvalidated results for clinical decisionmaking has caused harm to patients.[2] It is important that research results be validated and done in a reliable (e.g., Clinical Laboratories Improvement Act approved) laboratory before they are used for clinical decisionmaking. There is a potential risk of liability in providing information to patients at too early a stage in the research. The release of preliminary research results can lead to anxiety or requests for unnecessary procedures by uninformed patients. Donors will not have access to research results directly related to their specimen, but only to a class of data. This topic is further discussed in *2. Management of Ethical and Legal Issues*.

5.2.4.2 Communicating General Information on Research Results to the Public

The NBN will need to devise a strategy for communicating general research results in a way that satisfies participants' needs for information, researchers' desires to withhold their results until they are published or until patents are applied for, and physicians' desires to protect their patients from the harm that can come from the release of premature research results.

Consumer advocates should be engaged in the communications process. The challenge will be managing expectations from different constituencies and facilitating the balance between the responsible use of results by consumers and clinicians on the one hand, and the free availability of information demanded by scientists on the other. Useful models for making data available to patients include the approaches taken by First Genetic Trust (FGT) and IMPATH, Inc.[3]

5.2.4.3 Encouraging Equal Participation Among all Groups

The importance of having the NBN reflect the broad diversity of the U.S. population has been discussed elsewhere (see *2. Management of Ethical and Legal Issues* and *3. Biospecimen and Data Collection and Distribution*). Sociocultural values, economic disparities, and structural and institutional barriers all contribute to perpetuating significant health disparities among minority populations in the United States.

NBN outreach efforts will play a key role in reaching this objective. Special consideration must be given to the education and information needs of special populations, and to the challenges they face in participating in the system. A careful review of those repositories that have made an effort to increase minority representation, as illustrated in the RAND Report, will help the NBN plan this part of the outreach program.

---

[2] Eisen A. and Weber B.L. (2001). Prophylactic Mastectomy for Women with BRCA1 and BRCA2 Mutations – Facts and Controversy. *N. Engl. J. Med.*, Vol. 345, No. 3 (July 19): 207-208.

[3] IMPATH, Inc., is a private company formed in 1988 to improve outcomes for cancer patients by providing cancer information and analyses. IMPATH's mission is "to develop and offer innovative products and services that lead to more accurate diagnosis and more effective treatments for cancer patients." Two key components of IMPATH's integrated services to help accelerate the drug discovery and development process are its unique GeneBank tissue and serology archive, which is linked to longitudinal data, and the IMPATH Clinical Trials Network. IMPATH has a database of over 1 million patient profiles and outcomes data on over 2.3 million individuals. See *www.impath.com* for more information about the company. Additional information about FGT can be found in Appendix R and on the FGT Web site, *www.firstgenetic.net*.

5.3 Assessment of Stakeholders

This module has described communications planning as an information-based process in which each stage depends upon the information gathered in the previous stages. The very first step is to conduct a careful assessment of information needs of various NBN stakeholders. Each stakeholder group has specific concerns and needs that will require tailored educational approaches. Education of stakeholders will encourage overall participation in the system and, in particular, will facilitate the consent process as a precursor to tissue collection. This section presents a very preliminary needs assessment of those involved with the NBN. The NBN planners will want to reevaluate these needs before developing a communications plan. This preliminary assessment will, however, be useful in defining the general requirements for the NBN communications plan. The various components of the education program will need to be implemented using various methodologies in multiple settings. NDC TAWG members have suggested, as an intermediate goal, the use of educational forums at professional societies.[4]

5.3.1.1 The General Public

The public's concern with privacy and confidentiality may make some patients unwilling to donate tissue. Perceptions about tissue donation and research, and underlying sociocultural factors may affect the creation, implementation, and operation of a national biospecimen resource. For a variety of reasons, racial, ethnic, and cultural minorities have demonstrated lower rates of living and cadaveric organ and tissue donation for transplantation.[5] The lower donation rates have lead to disparities in health outcomes as some minority groups are over-represented on the organ waiting lists because of their higher rates of diseases such as end stage renal disease. These same factors may also affect the willingness of those groups to donate tissue for research. The consequent lack of adequate samples from subgroups of the population could impede the development of effective interventions or treatments targeted to these groups. Recent education efforts, such as the National Minority Organ and Tissue Transplant Education Program (MOTTEP), aimed at improving organ donation rates among some ethnic minority groups seem to have met with some success, and may also help promote a willingness to donate tissue for research among these group and may serve as models or collaborative opportunities for minority-targeted NBN education programs.[6]

An education program for the general public that helps to raise the overall level of understanding about biomedical and clinical research, addresses widely held misperceptions, and eases fears about privacy and confidentiality, could be very effective for promoting the NBN purpose.

---

[4] TAWG Meeting Summary, January 2003, p. 13. http://www.ndoc.org/Jan_2003_meeting_summary.pdf

[5] Verble M., et al. (2002). A multiethnic study of the relationship between fears and refusal rates. *Prog. Transplant.,* Vol. 12, No. 3 185-190.; Boulware L.E., et al. (2002). Understanding disparities in willingness to donate blood and cadaveric organs. *Medical Care,* Vol. 40:85-95; Boulware L.E., et al. (2002). Determinants of willingness to donate living related and cadaveric organs: identifying opportunities for intervention. *Transplantation,* Vol. 73:1683-1691.

[6] Callender C.O., et al. (2002). Increasing living donations: Expanding the national Mottep commnirt Grassroots model. *Transplantation Proceedings*, Vol. 34, No. 7 (November 2): 2563-2564; Callender C.O., et al. (2002). National MOTTEP: Educating to prevent the need for transplantation. *Ethnicity & Disease*, Vol. 12, No. 1 (Winter): SI34-SI37.

Increased knowledge about basic biomedical and clinical research and the benefits to research participation will form the backdrop for a specific program to encourage the public's willingness to participate. Even at this general level, in order to be effective, education efforts must be culturally sensitive and designed with an appreciation for the different attitudes toward biomedical research and how they color a group's perception of and attitudes toward involvement in research.

5.3.1.2 Patients (Potential Donors)

In addition to the general information developed for the public, the education program for patients should consider a number of additional concepts.

- Ethical concerns, the informed consent process, and what to expect
- Guarantee of privacy and confidentiality, and how this will be accomplished
- Future access to tissue and residual rights with tissue and information
- Future access to research results

Education for patients must be culturally sensitive. It must respect the basis of cultural barriers to participation, while providing enough accurate information to allow the potential participants to consider reasonable alternatives.

5.3.1.3 Family Physicians and Oncologists

The education program for the family physician and oncologist should focus on several types of information.

- Benefits of the NBN for research and therapies
- Information about how to counsel patients regarding tissue donation options
- Sociocultural issues that might influence their patients' willingness to discuss and participate in the system
- General information about the informed consent process and what to expect

5.3.1.4 Surgeons and Radiologists

The education program for surgeons/radiologists should focus on the following points.

- Benefits of the NBN for research and therapies
- Information about how to counsel patients regarding tissue donation options
- Specimen collection standards

Pathologists and surgeons may be reluctant to provide tissue for research because of liability concerns related to protection of privacy. An effective education program also will address these concerns.[7]

5.3.1.5 Pathologists, Biospecimen, and Clinical Data Banking Staff

The education program for the pathologists and staff responsible for banking biospecimens and clinical data should include specimen collection, preparation standards, and quality-control procedures.

5.3.1.6 Researchers

The education program for the researchers should focus on several issues:

- Benefits of using the system
- How to access and use the system
- Regulatory requirements governing access to and use of tissue samples and associated information, and what they need to do to comply

NBN will reach the research community through a variety of methods, including public advertisements, Web sites, and exhibits at national meetings. Additional outreach activities might include the following:

- Education forums at broad-based and discipline-specific professional societies involving pathologists, radiologists, and ethicists (e.g., American Association for Cancer Research, American Society for Clinical Oncology)
- NDC activities at, for example, the Public Responsibility in Medicine and Research meetings, where ethics groups gather
- Publications in high-impact media and biomedical research journals

5.3.1.7 Community-based Institutions

It will be important to create incentives for the healthcare community to participate in the NBN program. A communications program will need to answer the hospitals' question, "What is our role?" The education program for the community-based institutions will have to provide information on several topics:

- Benefits of participating in the system
- Impact on staff from participation in the system
- Kind of training required
- Role of consent counselors to educate patients/donors about patient issues

[7] NDC Forum II Meeting Summary, March 2003, p. 20.

- Training of staff about collection and preparation of biospecimens for storage
- Training of IRBs for new role in the consent process and infrastructure provided

5.3.1.7.1 Hospital Admissions Staff

The hospital admission staff is likely to be one of the earliest points of contact with potential donors. They must understand not only the benefits of research and participation in general, but also the critical role of hospitals in the process. The education program for the hospital admissions staff should focus on several areas:

- Benefits of participating in the system
- The vital role of the community-based institution in the acquisition process
- General information about the standardized and institutional informed consent process
- How to implement the standardized and institutional informed consent process, and what issues and questions to expect
- Training in how to answer questions about the use of specimens for research

5.3.1.7.2 Local Institution Review Boards

The education program for members of local IRBs should focus on a number of areas:

- Benefits of supporting the system
- Role in protecting the patient in a new national system
- The need to allow patients to make their own choices about participation in the research process
- How to address tissue donation options
- General information about the standardized and institutional informed consent process and what to expect, and why standardized consent is so valuable in genetic research

5.3.1.8 Patient Advocacy Groups

The education program for the advocacy groups should include:

- Benefits of the NBN to patient health
- Presentation of tissue donation options to their membership
- Sociocultural issues that might influence patients' willingness to participate
- Information about the informed consent process and the patient's right to retrospectively revoke consent
- Clarification that donors can expect to obtain generalized, not individualized, results deriving from their donation

5.3.1.9 Existing Repositories

The NBN should be clear in how it can serve as an additional resource that will help support research activities that current repositories were not designed to undertake. It will be important for the NBN to craft messages that reflect the collaborative spirit and open access inherent in its purpose.

5.4 Summary of Key Findings and Recommendations

In order for the NBN to succeed, it must be supported by many groups: Patients/potential donors, researchers, clinicians, academic institutions, hospitals, and commercial interests. This support will require a well-developed, comprehensive communications plan initiated at the earliest stage of the NBN project and sustained from development through implementation. Through a comprehensive program of education and training, outreach, and public relations, the NBN will promote a clear, accurate, and consistent set of messages about the resource. The information campaign will help to manage expectations and provide information about NBN policies that will encourage widespread confidence and trust in the new system. The following section provides more detail on specific recommendations.

Organizational Issues

- Effective communications among stakeholders should be considered a high priority for the NBN, and planning of the communications program should begin as early as possible.
- The NBN communications program should be broad and comprehensive, to meet a variety of communications needs and directed toward diverse stakeholders.
- Communications should be structured to ensure consistency of the NBN message throughout the planning, development, and implementation of the NBN program.
- The NBN should employ an evidence-based model in planning NBN communications. Under this model, planners identify the various audiences and determine their particular needs. Based on that information, the planners design a strategy to tailor messages to each specific audience, using appropriate media and effective channels of delivery.

Primary Communications Needs

- *Expectations Management*. The NBN should design a communication strategy that clarifies NBN's role and minimizes unrealistic expectations among different stakeholders (e.g., patients, researchers, potential financial sponsors, advocacy groups).
- *Minority Participation*. The NBN should design and implement communications strategies to encourage underrepresented ethnic and racial groups to contribute tissue samples. This will ensure a just system and a high-quality data bank that is capable of the highest level of research.
- *Longitudinal Data*. To develop high-quality longitudinal data, the NBN should develop a communication strategy that encourages patients to make a long-term commitment to providing the NBN with data (and possibly subsequent specimens).

- *Informed Consent and Communications*. The NBN should develop model methods for obtaining informed consent from patients, and model materials (including forms) explaining the consent process.
- *Clinician and Researcher Marketing*. The NBN is only successful if researchers use it to make new discoveries to accelerate progress against cancer. The NBN must place a strong emphasis on encouraging clinicians to participate in the system and encouraging researchers to use the resource as much as possible.
- *Crisis Management.* The NBN should design a communications strategy that, from the beginning, can handle unexpected crises—be they issues of technology, flawed human character, or political developments.

Module 6

Governance and Business Models

6.
Governance and Business Models

This module addresses options for National Biospecimen Network (NBN) governance and business and operations plans, as would be needed for any new corporation (for-profit or nonprofit), foundation, university, or consortium. It discusses governance models and associated authority and oversight; business models and options for funding the NBN, including rules of engagement, fees, and sources of funding; and initial organizational structures, including the requirements for operating a new enterprise with standard operating procedures (SOPs), quality assurance (QA), and communications plan. The governance and business models of organizations with analogous missions also are reviewed.

6.1 Introduction

The NBN is envisioned as a national scientific resource to support research on human biospecimens and associated data. It is anticipated that the NBN will start as a demonstration project and grow in size and scope as its user base increases to meet researchers' needs. Effectively governing and administering this enterprise will be an important challenge. The National Dialogue on Cancer (NDC) Tissue Access Working Group (TAWG) recognized the need for a sound governance structure and business plan to ensure that the goals of the NBN are met.

The NBN governance and business models should include strategies that permit the NBN to rapidly establish a firm foundation, and facilitate its growth into a self-sustaining organization. The strategies must be sufficiently defined to be attainable, yet flexible enough to accommodate the as yet undefined parameters of the NBN discussed elsewhere in this report.

The goals of this module are to:

- Identify an approach to continually incorporate best practices into NBN SOPs
- Propose options for a governance model
- Propose options for a business plan (define the market; explore costs and benefits and sources of public and private funding; recommend incentives for engaging participants and outline incentives needed to stock, maintain, and pay for the resource)
- Explore need for changes in existing rules and regulations, and/or new legislation
- Propose options for an operations plan, including an organizational model, SOPs for prioritizing requests for sample withdrawal, and QA

6.2 Background

6.2.1 Key Issues

The first step in developing a business plan is to decide how the NBN will be organized. For example, it could be a wholly new enterprise or a network of existing and/or new units where the collection of activities and organizations use shared data standards and business rules to carry out its purpose. If it is a network of existing and new organizations, it could be structured as a single agency, or managed as a confederation governed by a central body, or as a loose confederation of interested parties bound only by mutual agreements.

Once the organization and structure of the NBN is decided, the second step will be to design a sound governance model. This structure could involve differing levels of NDC and National Cancer Institute (NCI) participation, and would likely have a Board of Governors. The selection criteria for Board membership, for-profit or not-for-profit status of the entity, and the need for new rules, regulations, or legislation at the state or Federal level are all issues that must be addressed during this step.

The third step in launching the NBN will be to clarify the business plan to consider the market for biospecimens and associated data, anticipated costs, and potential revenue streams. Because the NBN is expected to be a joint effort between the private and public sectors, mechanisms should be identified to secure both private and public funding. In addition, user fees may cover some of the expenses.

The fourth step is to define an operations plan that describes the working structure and functions of the NBN. Networks are based on standards. The NBN will need to establish and enforce standards for biospecimen collection, data handling, and perhaps complex biological analysis. QA and systems for oversight will be vital to NBN activities. Specific issues include whether processing, storage, and distribution of samples will be centralized, regionalized, or housed locally; whether advanced analytic services will be provided and, if so, whether they will be centralized or decentralized; and what systems will be in place to ensure equitable tissue specimen distribution.

6.2.2 Existing Practices

It is helpful to review a broad array of existing governance and business models for a variety of multi-institutional cooperative research endeavors. These models include biospecimen repositories, research networks, and other research collaborations. They differ by organizational structure, funding streams, and their integration with bioinformatics functions (including the provision of advanced analytic capabilities). Some contain key centralized components, while others offer greater decentralization of specific functions such as tissue collection, processing, storage and dissemination. This section reviews the governance and business operations of a variety of systems involving biological materials, roughly classified by their funding and governance models, and then in alphabetical order (see Appendices N through T for more detailed descriptions of some of these systems).

6.2.2.1 Nonprofit, Government-Funded Models

Cooperative Human Tissue Network

The Cooperative Human Tissue Network (CHTN) is supported by the NCI to provide biomedical researchers with access to human tissues (see Appendix N). Since its establishment in 1987, the CHTN has provided more than 500,000 high-quality tissue specimens from a variety of organs to more than 1,000 investigators. The CHTN has an established, six-division infrastructure, with good quality control (QC), timely collection of specimens, and established relationships among surgeons and pathologists. Its primary focus thus far appears to be prospective (on-demand) collection of biospecimens.

Experienced personnel from all six divisions of the CHTN participate in a coordinating committee that formulates policies for the operation of the CHTN. The voting members of this committee include the Principal Investigator and an additional member from each division. In addition, the NCI has one voting member. The coordinating committee meets periodically to assess the operation of the CHTN and to change or modify operating policies. A chairman and secretary of the committee are elected yearly (*www-chtn.ims.nci.nih.gov/purpose.html*).

Early Detection and Research Network

NCI's Early Detection and Research Network (EDRN) of cancer investigators focuses on the development and testing of promising biomarkers or technologies aimed at the early detection of cancer. It aims to collect biospecimens and to promote collaboration and rapid dissemination of new information among academic and industrial leaders in molecular biology, molecular genetics, clinical oncology, computer science, public health, and clinical applications. Collaborators share resources for the development and testing of biomarkers, clinical and epidemiologic studies, and data management. It is a "closed" network for the two-dozen Principal Investigators, but collaborators may apply to participate in joint studies. It has been reported that the EDRN has deployed a national infrastructure for sharing data and biospecimens (*www3.cancer.gov/prevention/cbrg/edrn*).

Integrated Molecular Analysis of Genomes and Their Expression Consortium

The Integrated Molecular Analysis of Genomes and their Expression (IMAGE) Consortium was initiated in 1993 by four academic groups who shared a common vision of how to achieve an important goal in the study of the human genome. Specifically, the IMAGE Consortium shares high-quality, arrayed cDNA libraries and places sequence, map, and expression data on the clones in these arrays into the public domain. Using this information, unique clones then can be rearrayed to form a "master array," which ultimately may contain a representative cDNA from each gene in the genome under study. The human and mouse genomes were the first to be studied, and the collection now contains clones from rat, zebrafish, Xenopus, and rhesus macaque, with additional species being added as resources permit. All clones are available free of any royalties and may be used by anyone who ascribes to the guidelines, which specify identifying the source as IMAGE, and free redistribution of clones, progeny, and any derivatives thereof *(image.llnl.gov*).

United Kingdom National Cancer Tissue Resource

The UK has developed a comprehensive plan for the National Cancer Tissue Resource (NCTR) to meet the demands of its research communities for biological samples linked to clinical outcome information (Appendix O). The NCTR's primary aim, similar to that proposed for the United States, is to allow the routine collection, storage, processing, and distribution for research of malignant tissue with case-linked normal tissue linked to standardized, site-specific histopathological data and clinical outcome information. The UK National Cancer Research Institute partners and other governmental departments initially will fund the planned NCTR. In particular, the UK's Department of Trade and Industry will provide funding for informatics support. Eventually, private investment (as public-private partnerships) will be instituted in a regulated manner.

It is envisioned that the NCTR will comprise five key components:

- *Coordinating Unit*, located within the Coordinating Center, to implement and operate the NCTR.
- *Tissue Acquisition Resource Centers* (TARCs), created as a linked network, selected through a tendering process, and contracted to adhere to SOPs for prospective collection of biological samples and outcome clinical information (a fully operational NCTR will incorporate up to six geographically distributed TARCs. It is anticipated that each TARC will collect samples from up to 1,000 cases per year).
- *Tissue-processing Resource Centers*, created as a network, for production of DNA, RNA, and tissue microarray, selected through a tendering process and contracted to adhere to SOPs.
- The collection of samples from key established and future clinical trials, coordinated through Clinical Trials Offices and the National Cancer Research Network.
- *Bioinformatics Hub*, a central information system that will link (1) tracking of collection, processing, distribution, and analysis of samples to (2) histopathological datasets in line with recommendations by The Royal College of Pathologists, to (3) clinical/outcome information, to (4) results of research.

6.2.2.2 Nonprofit, No Government Funding

The Single Nucleotide Polymorphism Consortium

The Single Nucleotide Polymorphism (SNP) Consortium, a wholly privately funded collaboration of more than 10 pharmaceutical, information, and technology companies, academic centers, and the WellcomeTrust, is analogous to the data-sharing part of the NBN. The SNP Consortium was formed in 1999 to find and map 300,000 common SNPs (see Appendix P). Exceeding initial expectations, the Consortium has identified and mapped more than 1.8 million human SNPs to date. The maps, which are still being updated and refined, are publicly available to researchers worldwide.

The SNP Consortium has a 501(c)3 nonprofit governance core and is run by a Board of Directors, with representatives from each funding partner. The Consortium has only one full-time employee; the remainder of the work is contracted out to technical experts identified by the Board. Government involvement was deliberately avoided in order to expedite the SNP discovery process.

6.2.2.3 For-Profit, Minimal or No Government Funding

The Ardais Corporation and First Genetic Trust are only two examples that could be described under this model. Other examples, for which less detailed information was readily available about governance structure, include Genomics Collaborative, Inc., a diagnostic and therapeutic company providing access to DNA, serum, and tissue collected from patient populations around the globe (*www.genomicsinc.com*); Asterand, which operates a human tissue bank to help medical researchers discover new diagnostics and therapeutics for cancer and other diseases *(www.asterand.com*); and IMPATH, a private company that has a database of over 1 million patient profiles and outcomes data on over 2.3 million individuals (*www.impath.com*).

Ardais Corporation

Ardais Corporation is a privately held, 4-year-old, for-profit clinical genomics company that markets its products and services to pharmaceutical and biotechnology companies, academics, and government researchers (see Appendix Q). The company has invested $40 million establishing the infrastructure, capability, and bioinformatics systems to enable the collection, processing, and storage of tissue, assembling a library of tissue samples from four academic medical centers across the country: Beth Israel Deaconess Medical Center in Boston, MA; Duke University Medical Center in Durham, NC; Maine Medical Center in Portland, ME; and the University of Chicago in Chicago, IL. Ardais deploys its SOPs and manages and supports the collection process at each institution, with the goal of achieving consistently high standards for maximum sample quality and annotation while protecting patient donor safety and privacy.

Ardais' portfolio of clinical genomics resources includes formalin-fixed and frozen human tissue samples and associated clinical information; standard or custom tissue microarrays for high-throughput parallel analyses; and molecular derivatives, such as RNA, which are validated for research use through a battery of qualification procedures. Ardais' tissue samples and associated clinical information are accessible through a bioinformatics system called the BIGR™ (Biomaterials and Information for Genomic Research) Library, which currently provides access to over 170,000 human tissue samples collected under uniform, highly standardized conditions.

First Genetic Trust

First Genetic Trust is a business that develops information technology solutions to address data, privacy, confidentiality, and ethical challenges in genomics and proteomics (Appendix R). Although it does not distribute biospecimens, it is a relevant example for the NBN because it provides an innovative platform to store and analyze genetic and clinical data. It has created an information technology platform with three goals: (1) Enable large-scale genomic research and eventually clinical genomic research; (2) work with pharmaceutical companies to speed the development and use of new drugs; and (3) enable clinical adoption of genomics. Currently, First

Genetic Trust is involved in two sponsored research studies. The first is a large pharmaceutical protocol in the United States and Europe, which has received very favorable responses from 19 institutional review boards (IRBs). The second is a breast and ovarian cancer research program at the Memorial Sloan-Kettering Cancer Center, which enrolled its first patient in early 2003. Activities with other pharmaceutical companies and universities (including Johns Hopkins) and the International Genomic Consortium are in the formative stages.

Public access is the least-defined aspect of the system. Some data will be completely open, some will be conditionally open, and some will require reconsent or further aggregation to be open. Sponsors own the tissue and data, but the system will hold individual identifiers. Histopathological image data are not currently available, but it is technologically feasible to provide them. The design protocol can be written to automatically aggregate clinical updates and secular outcomes.

6.2.2.4 Public/Private, Mixed Funding

The Clinical Cooperative Group Tissue Bank is one of the many NCI-supported human specimen resources available.[1] The National Surgical Adjuvent Breast and Bowel Project, Cancer and Leukemia Group B, and others represent mixed funding banks. The NCI Cancer Centers and perhaps the Specialized Programs of Research Excellence receive outside funding that may go toward banking and specimen resource activities. The Children's Oncology Group (COG) provides yet another example that is described in more detail below because of its more unique features.

Children's Oncology Group

COG is an NCI-supported clinical trials cooperative program devoted exclusively to childhood and adolescent cancer research. It was formed by the amalgamation of several groups, including the Children's Cancer Group, the Pediatric Oncology Group, the Intergroup Rhabdomyosarcoma Study Group, and the National Wilms' Tumor Study Group. COG develops and coordinates cancer clinical trials conducted at the 238 member institutions, which include cancer centers of all the major universities and teaching hospitals throughout the United States and Canada, as well as at sites in Europe and Australia. There are 5,000 COG members.

All patients entering COG studies at member institutions are registered through a Web-based remote data entry system, through which patient responses to therapy are centrally collected, monitored, and analyzed. COG is managed under contract to NCI by the National Children's Cancer Foundation, which serves as the grantee and fundraising organization for COG; receives and manages the COG grant from NCI; provides administrative, fiscal, personnel, grants and contracts, and technical services to support the COG operations; and raises private funds for COG research, as well as for its own advocacy and education activities.

[1] Other resources include NCI AIDS Malignancy Bank, the Cooperative Family Registry for Breast Cancer Studies, and the Cooperative Family Registry for Colorectal Cancer Studies.

British Columbia Cancer Agency/Genyous Life Sciences, Inc.

The Tumor Tissue Repository (TTR), a public sector initiative begun by the British Columbia Cancer Agency in British Columbia, is a public/private model in Canada. The governance structure is nonprofit, with a Board of Directors and operational units staffed by employees of the TTR. Specimens will be obtained from the 4 million patients served by Canadian National Health system in British Columbia, and primary funding will come from the Canadian health system. Genyous Life Sciences, a commercial firm, provided funding to the TTR in return for commercialization rights to these research findings for the first 5 years of operation. Genyous does not have a seat on the TTR Board.

International Genomics Consortium

The International Genomics Consortium (IGC) is a private, nonprofit 501(c)3 medical research organization, operating to ensure full public benefit of the Human Genome Project (*www.intgen.org*/). IGC expression projects on human diseases will utilize a consortium model of enhanced and accelerated tissue collection under standardized conditions to produce detailed data on the molecular characterization of disease. The data will be available as a public, standardized, regulatory compliant, readily accessible and searchable series of genomic databases for use by the worldwide scientific community. The information from this unique resource, which joins together the public and private sectors, was developed to stimulate research on the underlying molecular mechanisms of diseases to accelerate development of diagnostic tests, improved treatments, novel therapies, and disease prevention.

Infrastructure and executive management positions in IGC are funded through multiyear committed support from public and private sources in Arizona. Funding of the expression projects is derived through donations from members, which include pharmaceutical and biotech companies, technology and informatics companies, foundations, governments, academic and research institutions (providing in-kind support), and private donations.

IGC is governed by a board of directors that oversees an executive management team composed of a chief executive officer (CEO), chief medical officer, chief financial officer, and chief information officer. A scientific advisory committee speaks to the vision, operations, and opportunity represented by IGC. Each of the different expression projects for human diseases is advised by an executive steering committee (ESC) of members representing the funding sources for the project. The ESC develops milestones for performance of IGC's projects, formulates metrics of deliverables, and advises operation of the execution of the projects. Subcommittees of the ESC present recommendations on technology platforms, information technology, clinical data, patient advisory affiliations, legal considerations, and public relations (see Appendix S).

United Network for Organ Sharing/Organ Procurement and Transplantation Network

The Organ Procurement and Transplantation Network (OPTN) is the unified transplant network established by the U.S. Congress under the National Organ Transplant Act of 1984 to be operated by a private, nonprofit organization under Federal contract (see Appendix T). Although the requirements for collection of organs for transplantation are very different from those for collection of diagnostic specimens for research, OPTN is included as an example of a

governance model that may be helpful in the development of the NBN. The United Network for Organ Sharing (UNOS) was awarded the first OPTN contract in 1986 and continues to administer the OPTN under contract with the Health Resources and Services Administration of the U.S. Department of Health and Human Services.

The OPTN contract with UNOS is a cost-share contract. The contractor is responsible for 92 percent of the costs. Section 372 of the Public Health Service Act requires that the OPTN be performed under contract by a private, nonprofit entity and limits appropriations to not more than $2 million in any fiscal year. The contractor collects patient registration fees (a listing fee of $450 in order to have each patient placed on the transplantation waiting list) to supplement the cost of performing the OPTN contract requirements.

6.3 NBN Models and Recommendations

6.3.1 Proposed Funding and Governance Model

Reflecting the idea that the NBN would be able to accept funds from both public and private sources, the Design Team proposed a funding and governance model with a not-for-profit organization at the center of the nexus (see Figure 6-1). It is recognized that a good portion of the chain of trust discussed in *2. Management of Ethical and Legal Issues* is embedded in the expectation that a nonpartisan and nonprofit-motivated organization have stewardship of the NBN. Biospecimen donors and the public in general would expect that the NBN owe primary allegiance not to stockholders, but to the pursuit of science and discovery. In addition to the advantages of enhanced credibility, a nonprofit entity might also attract a wider choice of possible leaders with experience and familiarity in the nonprofit world, tax preferences, competitive advantage in obtaining government grants and cooperative agreements, lower operating costs (since there is no profit incentive), and the ability to accept government and private funds. This model also may make it easier to keep the data "precompetitive." It is of course possible and recommended that the 501(c)3 organization contract various business functions to other organizations, for example to universities, research centers, or private companies.

In addition to suggesting a not-for-profit core operations center, the Design Team recommended the following model for NBN oversight (Figure 6-2):

I. Board of Governors — The oversight authority

II. NBN Operations Center — The "home office" or headquarters

III. The Business Units — The operation arm (discussed after the business plan).

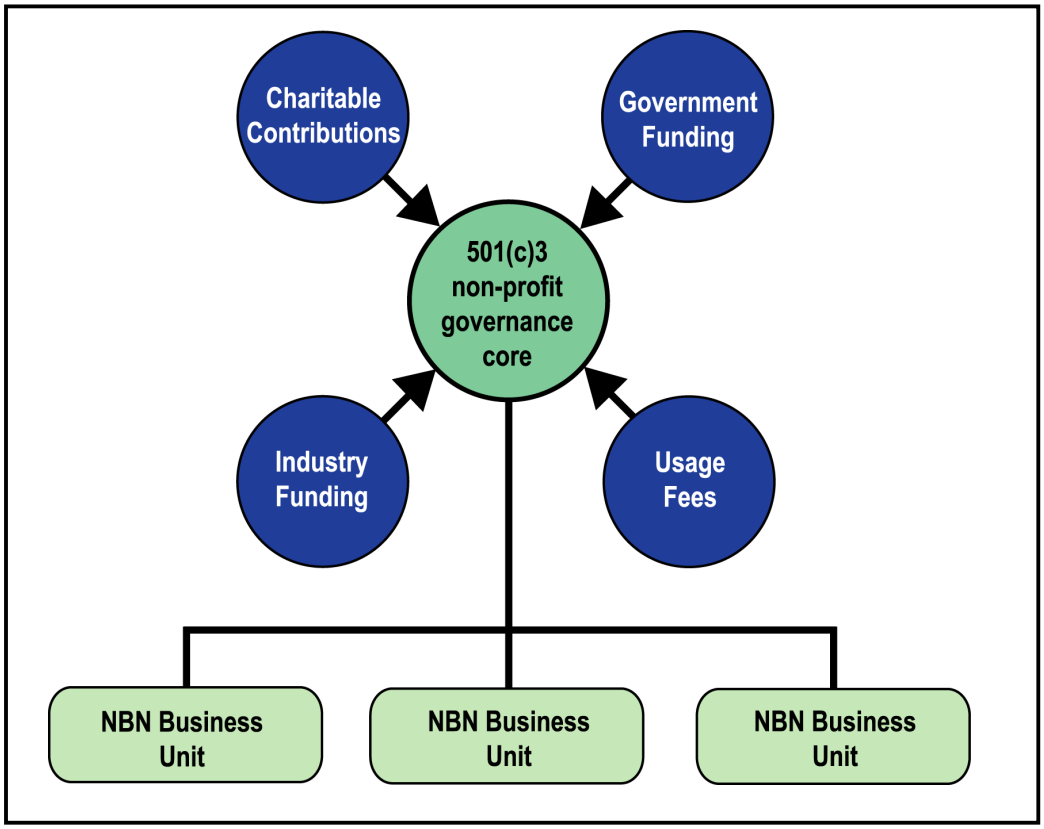

Figure 6-1. Proposed NBN Funding Model

The **Board of Governors** would be either the NDC itself or (more likely) a 7- to 10-member group formed by the NDC (and for which the NDC would have representation) to provide oversight. The needed skills of Board members include a mix of strong science and business experience. The Board would have to establish criteria for membership, draw up a charter and bylaws, and decide whether the CEO would be a member. Members would include individuals representing advocacy groups, major professional associations, and government agencies. After completing its own organization, the next job for the Board would be to decide the funding structure and the hiring of the CEO (and possibly other key officers).

The **Operations Center** would be the "home office" or headquarters for the NBN, under the leadership of the CEO. It would consist of a relatively small staff of senior executives and support staff, who would have overall responsibility for the NBN; and three in-house operational units: QA, Bioinformatics and Data Management, and Communications.

The main task of the Operations Center would be to manage the operations of the NBN. It would have basic management responsibilities, including strategic planning, budget, legal, and day-to-day administration.

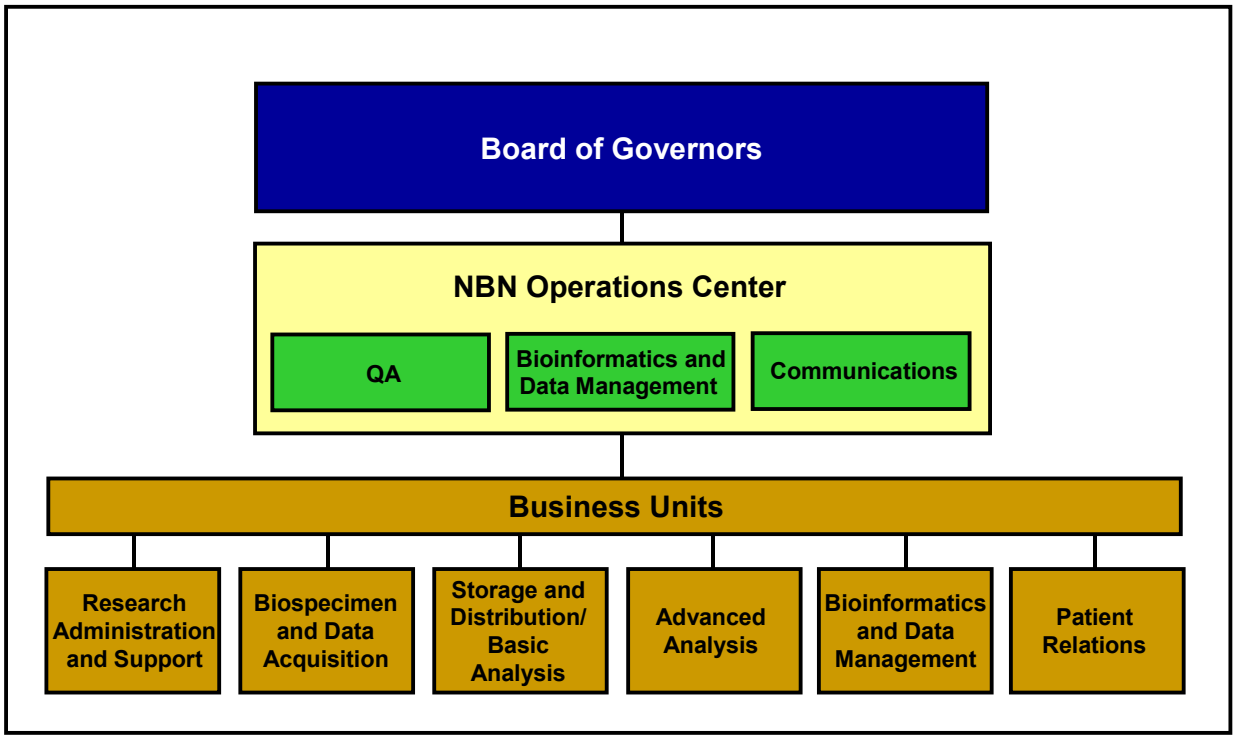

Figure 6-2. Possible Organizational Structure for the NBN

As discussed above, the Design Team recommended that the Operations Center be structured as a nonprofit entity. The Foundation for the National Institutes of Health (FNIH) may be an option to serve as an incubator for the NBN Operations Center. It could help with planning and promote the integration of nonprofit and private-sector entities, or serve as an agent for the NBN in accepting and distributing funding from Federal agencies and private donors. The FNIH has served in a similar capacity for the Mouse Sequencing Consortium. Specific milestone needed to trigger independence and how such a transition would be managed must be considered.

As an alternative, the NDC itself could play an organizing role. The real issue is whether or not an entity has conflicting roles, be it for-profit, nonprofit, or academic. In general, there ought to be a distance between the Operations Center and the business units, and between the Operations Center and sites. Appropriate charter and bylaws would remove many of the potential problems.

Quality Assurance (QA) is a critical and central management function within the Operations Center, devoted to ensuring that the NBN standards are set and documented; that business and operations goals are well defined (e.g., customer satisfaction and costs); and that the systems for meeting regulatory compliance such as those required by Health Insurance Portability and Accountability Act, and compliance with other applicable Federal and state regulations, are in place. QA should be distinguished from QC. QA is a set of processes to ensure quality. QC is the application of those processes to specific activities to ensure the quality of the work product. QA will functioin to work with business units to review and create all SOPs for all related business units and affiliated sites.

The QA function is a critical NBN function. In that spirit, the QA functions should take into account applicable best practices such as the Food and Drug Administration Good

Manufacturing/Laboratory Practices, other industry standards (College of American Pathologists [CAP], International Society for Biological and Environmental Repositories [ISBER], National Committee for Clinical Laboratory Standards [NCCLS], and other bodies) and document the NBN practices. The business units would be subject to inspection/audit and, as part of their QC functions, responsible for reporting required data to the QA Unit. The QA Unit would then analyze the QC data to monitor compliance. The NBN is envisioned as a *standards-based* organization with QA at its heart.[2] The QA oversight should be collaborative, proactive, and educative (and might have a research component for developing best practices and evaluating pilot projects). It is not authoritarian and inspection oriented. Resolution of concerns will need to be institutionalized and well documented. The QA Unit will have a staff of several highly trained individuals who report directly to the CEO.

The NBN would make every effort to maximize the value and use of existing resources, provided that they meet NBN standards. Moreover, depending on the research use, it may be possible to have ranges for compliance (e.g., it may be sufficient to know the age of the donor within 2 years); or to have different categories of standards (e.g., tissue was frozen within 30 minutes or can be paraffin embedded). The challenge will be to classify existing resources according to their utility for different analytic techniques, from hematoxylin and eosin staining to DNA microarrays.

The NBN Design Team recommends that QA be in-house because, like accounting, it is a central activity; also, this area is susceptible to conflicts of interest, especially in a for-profit business unit model. There was a strong consensus that the QA function is a vital corporate executive responsibility that ought to be closely monitored by the CEO, should report to the CEO, and for which the CEO should be held accountable.

Although there was some discussion that having the QA activity in-house may create the appearance of a conflict of interest, as the QA function ultimately measures the quality of the NBN Operations Center and its executives, this concern was outweighed by the importance of having direct oversight by the CEO. QA activities should be regularly reported to the Board.

The NBN will be a standards- and performance-based organization that will need to develop and enforce quality standards. The NBN would consult continually with other experienced organizations knowledgeable in acquisition, processing, storage, and distribution of high-quality biospecimens (e.g., CAP, ISBER, NCCLS, existing biospecimen resource providers) to determine and update appropriate processes. It is recommended that the NBN be neither an accrediting nor a certifying body, but call upon the experience of others in developing its standards.

Bioinformatics and Data Management will have an Operations Center function and an individual business unit function. Core staff, managed by a Chief Information Officer (or equivalent), will have a central architectural role and also manage the Bioinformatics and Data Management Business Unit. There will also be important bioinformatics activities needed for

[2] Possible liability implications or significant costs associated with being a standards-based organization still need to be examined.

business unit operations at every site. The core Operations Center Bioinformatics and Data Management staff will coordinate these cross-cutting activities. The scope of the bioinformatics activities at the operations center level will derive from needs (see *4. Bioinformatics and Data Management*) and must be carefully delineated to facilitate clear integration with the Bioinformatics and Data Management Business Unit and the bioinformatics needs of other business units.

Communications will have an Operations Center function and is also embedded within the Patient Relations and Research Administration Business Units. Core staff, managed by a Chief Communications officer (or equivalent), will have a central role in all communications internal and external to the NBN. The primary purpose of elevating Communications to the operations level is to ensure the appropriate coordination of a broad and comprehensive program of communication throughout the NBN and its various units, with responsibilities encompassing education and training, outreach, and public relations to ensure close coordination with associated responsibilities of order processing, distribution, and shipping. It is important to publicize the NBN to its key constituencies and to develop and maintain clear and effective communication with all participants in the system (see *5. Communications*).

Another extremely valuable function of Communications is to conduct an ongoing professional market assessment of users, including at pharmaceutical and biotechnology company researchers, to inform NBN design and engineering of evolving needs. This assessment could include questions to determine the different types of tissues sought by researchers, inquiries about what other resources are available (including those that might be NBN competitors), and what advanced analysis services researchers desire. Regular feedback about user needs can help to keep the NBN current and able to respond to emerging demands for more sophisticated analyses.

6.3.2 Business Plan

6.3.2.1 Business Plan: Issues

The NBN is a major new initiative that requires a sound business plan, with a careful assessment of costs and benefits. It is difficult to describe the NBN in classic business terms. The NBN has two "products." (1) The primary product is represented by the biospecimens and associated patient data; and (2) information derived by users, such as expression microarray results or the results of an *in silico* experiment, is the second. The "users" are researchers. The "producers" may include the patients, surgeons, endoscopists, dermatologists, radiologists, pathologists, and other clinicians. Finally, there is a long list of other interested parties (government, industry, academe, other nonprofits, and the survivor community).

Willingness to Pay. The cost of accessing the NBN resource should not be prohibitively high. The results of the NDC-NCI questionnaire administered at AACR showed that 85.3 percent of respondents would be willing to pay between $20 to $100 per specimen for well-characterized biospecimen samples with definitive associated patient clinical data. Additionally, a demand for research data provided along with specimens was noted, as stated above. Approximately 69

percent of respondents were willing to pay between $100 to $500 per specimen for well-characterized biospecimen samples with definitive associated patient data accompanied by standardized gene expression by DNA microarray data. (It should be noted that this survey targeted a convenience-based, nonrepresentative sample, but focused on the appropriate audience. Complete results are provided in Appendix E.)

Costs. The estimated cost of a sample varies by how much information accompanies it. There will also be economies of scale: A full freezer costs as much to operate as one that is one-quarter full. Experience has shown that collecting human specimens for research can cost between $70 and $2,000 per specimen shipped, depending on the degree of annotation.

Determining the precise cost per specimen is difficult since there are many complexities. A single case (patient specimen) can yield multiple research samples. Providing 250,000 samples at $200 each would have an annual cost of roughly $50 million (exclusive of start-up costs). The cost would of course scale if the marginal cost of specimens were greater than $200. By comparison, approximately $40 million per year is allocated currently by the NIH to extramural programs for "tissue banks."[3]

Charge for Use. Some of the costs for annotation, storage, and distribution could be recaptured through charges to researchers. Note that what is charged or paid by the user may be more or less than cost. Charges could also be used to provide incentives for certain behaviors. For example, researchers at academic institutions who provide specimens and data to the NBN could be charged a lower amount for usage. Similarly, users who provide their research data might be charged a lower fee than those who do not. Charges might be higher for more intensive users of data, or for those who use harder-to-obtain or more desirable samples. Whether the charges to industry and academia should be identical, cost-recovery considerations, and final pricing for general and subsets of NBN biospecimens, remain to be determined.

Funding Mechanisms. The NBN is designed to be a national resource, supported by a mix of public and private funding. It would be expected that the interested parties would be a mix of Federal, state, commercial, and private (philanthropic) organizations. Some will contribute in proportion to their perceived benefit, while others will provide funding purely to contribute to the advancement of research. Pharmaceutical companies may find it advantageous to support a national venture, either directly or via fees, as a way to share costs.

Incentives. To reap the enormous scientific benefits from the NBN, a sufficient number of tissue donors and researchers who see value in the resource must actively participate in the endeavor. Open communication and true partnerships with existing institutions are necessary to maximize utility, overcome barriers, and change attitudes. To develop a sound business plan, it will be critical that there be incentives to participation at multiple levels (see Table 6-1). Although it is difficult to quantify the value of these benefits, it is still useful to articulate the incentives or benefits that could accrue from the NBN.

---

[3] This figure may be an overestimate, since costs for projects with research components were included in the estimate. Personal communication, R. Aamodt, July 15, 2003.

Table 6-1. NBN Participants: Benefits and Incentives to Participation

Type of Participant	Benefits and Incentives to Participation
Government	• Efficiencies for research and operations, especially via promoting comparisons, economies of scale, and enforced QA/SOPs • Fulfilling government mission of creating resource that private sector may not be able to provide (may be question of scale and specifics) • If parts are "outsourced," then it promotes efficient use of government personnel for purely governmental functions, with NBN Operations and business units handling administration • Access to charitable monies for research
Academics, Research Institutes	• Efficiencies for research and operations, especially via economies of scale and enforced QA/SOPs • Access to previously "hidden" or simply unavailable resources • If independent organization is created, then it provides a very open process and may level competitive playing field
Pharmaceutical Manufacturers/ Contract Research Organizations	• Information for drug development, including improved information about clinical and molecular effects of drugs (desired and undesired effects) • Possible decrease in approval time if they are using NBN "national standards"
Device and Information Technology Companies	• New market for "national standards-based" products • Vastly more reliable substrate for product development and testing
Biospecimen Collection Sites	• Priority access to biospecimens • Institutional commitment to SOPs and QA developed by the NBN • Improvement of pathology and bioinformatics infrastructure • Fees • Possible Centers for Medicare and Medicaid Services reimbursement of selected NBN services (especially if Clinical Laboratories Improvement Act-like standards put in place)
Donors	• Sense of contributing to cure and prevention • Possible improved understanding of personal, familial, or ethnic group factors, especially long-term factors (but it must be understood that risks of using unvalidated research results may outweigh benefits)

Suggestions to foster increased cooperation and collaboration include the use of clinical cooperative groups to encourage the development of a culture of teamwork and cooperation among institutions that might participate in the NBN; and the use of demonstration projects, which can provide learning experiences, accelerate progress, and can be used to develop standards and determine good tissue practices for the full system. Federal funders can also encourage the use of data generated from demonstration projects to support grant proposals, and the use of the NBN by NCI grantees. Biospecimen contributors could receive priority access to tissues, or receive a number of samples for free in equal number to that which the institution contributes. Enhancing the NBN with strategic linkages to other data-gathering repositories, such as cancer registries, to enhance the database—or to clinical trials data—could further attract users.

6.3.2.2 Business Plan: Recommendations

The NBN Design Team recommended that the initial business plan assume a combination of direct government funding plus user fees. The government's role in funding will likely be most prominent at start-up, but it may be offset over time by user fees. As the NBN provides precompetitive access to specimens and data and takes no intellectual property position relative to findings from research using the specimens, users' ability to retain intellectual property rights should enhance participation. It is probably unrealistic to expect the NBN to be self-sustaining for several years, if at all. Institutions and individuals that benefit from this system would be expected to help support it financially, either directly or in-kind.

A detailed model that forecasts costs and charges will need to be developed. The model would take into account the various products, producers, users, buyers, and payers. This model ought to include fixed costs (governance, business infrastructure) and variable costs (the number of samples and sites). Developing a pricing strategy would be a major objective of this study.

6.3.3 Operations Plan

Detailed operations pertaining to building and maintaining of the NBN are discussed in this section.

6.3.3.1 Operations Plan: Issues

The operational issues include management of the NBN. Management covers activities related to staff, budget, administration, policies and procedures (including accreditation), facilities, and the scientific program itself. Key elements of the day-to-day operations include collection and storage of specimens and operation of an information system; and resource administration, including overseeing systems to allocate the specimens to researchers; compliance with applicable laws and regulations; communications with researchers and other interested parties; and QA. An organization of such complexity will require professional management and well-defined SOPs. It is anticipated that there will be intense interest in NBN operations by various organizations and agencies, such as government, academic, industry, and not-for-profit entities, including those representing survivor or patient groups, and standards bodies.

6.3.3.2 Operations Plan Recommendations: The Business Units

A series of business units would be supported and coordinated by the NBN Operations Center. Each business unit will focus on a particular part of the process, as discussed later. The business unit is a *functional entity* (e.g., biospecimen and data acquisition), and should not to be confused with a *site* (e.g., a community hospital). A business unit could have multiple sites, and one site might be part of one or more business units. For example, a community hospital might be part of the Biospecimen and Data Acquisition Business Unit and might also provide storage and distribution capabilities. The business arrangements between the business unit and its sites are mutually established and would not be managed by the NBN Operations Center. However, the sites would be subject to QA oversight from the NBN Operations Center. Each business unit would be responsible for developing local QC SOPs, and for reporting data to NBN Operations to allow for network-wide and business unit-specific QC. The six proposed business units are now described in more detail. The letting of contracts for individual business units should be flexible in that multiple business units may be contracted to a single entity under one contract to achieve synergy and consistency.

The Research Administration and Support Business Unit would administer the Biospecimen Utilization Review Committee, help set scientific priorities for access to specimens, document research results, and help define an ongoing scientific needs agenda for the NBN.[4] It would develop SOPs for each of these functions, coordinate support activities, work with standards organizations, and serve as a common meeting ground for the activities of the other business units. An important component of its work would be developing novel ways to track and allocate specimens; comply with regulations; and communicate with researchers, IRBs, and other interested parties. This business unit would have important communications responsibilities with Principal Investigators, other partners, contractors, and sponsors. It also would be responsible for understanding and meeting the needs of users.

The Specimen and Data Acquisition Business Unit would facilitate patient data collection by identifying patients, administering consent, acquiring specimens of specified quality, collecting clinical information, and performing a minimal set of analyses, as defined by SOPs (these activities are further detailed in *3. Biospecimen and Data Collection and Distribution*.) There could be multiple specimen acquisition organizations, and each might have multiple sites to ensure that they meet the range of tissue types, quantities, and qualities required. This business unit must have capabilities to collect associated pathology, demographic, social history, and clinical data as well as biospecimens.

The Storage and Distribution/Basic Analysis Business Unit would be responsible for specimen processing, storage, and retrieval, as detailed in *3. Biospecimen and Data Collection and Distribution*. Pathology review also could be based here, but it should be done in close proximity to the collection site.[5] Adherence to standards and SOPs (both current and future) is

[4] Although the Biospecimen Utilization Review Committee would represent a component of the Research Administration and Support business unit, it would be monitored at the Operations Center as part of the QA process.
[5] The issue of central review versus onsite review should be explored further.

important. This business unit might be partly or wholly decentralized, and could be a single entity or a consortium, as long as there is a single contracting entity. It could take advantage of existing facilities, build new ones, or use a mixed model. It can thus utilize existing facilities at universities or corporate or governmental laboratories. The challenge will be to develop fiscal and scientific incentives to make NBN regional facilities attractive to researchers, including the availability of precompetitive, highly standardized, well-annotated tissues with appropriate associated data in a national database The big attraction for researchers would be access to a far greater variety of biospecimens than could ever be acquired locally. Regionalization also would facilitate QC of research services such as generating DNA and RNA microarrays. Figure 6-3 illustrates a possible NBN configuration.

An initial snapshot of the nationwide NBN might include:

- Two to five regions in the United States
- Five to 20 collection centers per region
- One initial processing, storage, and distribution center per region (specimens would be distributed nationally from each region, and some specimens would be stored in other than the home region repository for disaster avoidance purposes)
- One advanced analysis center for the nation
- A virtual bioinformatics grid encompassing the entire country.

Expansion of the NBN would involve the identification of additional collection sites in each region, including Alaska and Hawaii. It is not anticipated that additional processing, storage, and distribution centers, or additional advanced analysis centers, would be added as the NBN grows.

The Advanced Analysis Business Unit would be charged with providing advanced analytics (e.g., gene expression microarrays, proteomics) to NBN users on a fee structure to be determined. The goal is to provide standardized tests at competitive prices, and promote comparability, to be specified by SOPs and monitored by sophisticated QA. It is not envisioned that individual researchers would be required to use this service in preference to those provided by other laboratories, including their own. In addition, this business unit would have a scientific function of keeping abreast of new technologies and bringing them, as appropriate, to bear on the NBN mission. This business unit may be a higher priority in later years. It is important to conduct market analysis to determine demand prior to establishing this business unit.

The Bioinformatics and Data Management Business Unit will work closely with the Operations Center's Bioinformatics and Data Management staff to define the system architecture, establish and maintain standards (and develop new ones as needed), and define policies for data exchange between the business units and sites as outlined in *4. Bioinformatics and Data Management*. It would manage data flows throughout the entire "life cycle" of a biospecimen and its associated information, from procurement, storage, distribution, and analyses to possible reposting of research results. It is expected that eventually most data will reside locally, with a central facility that will point to data, and thus provide users with access to information about specimens. SOPs and QA will focus on software quality, system availability and utility, security and backup, and data validation.

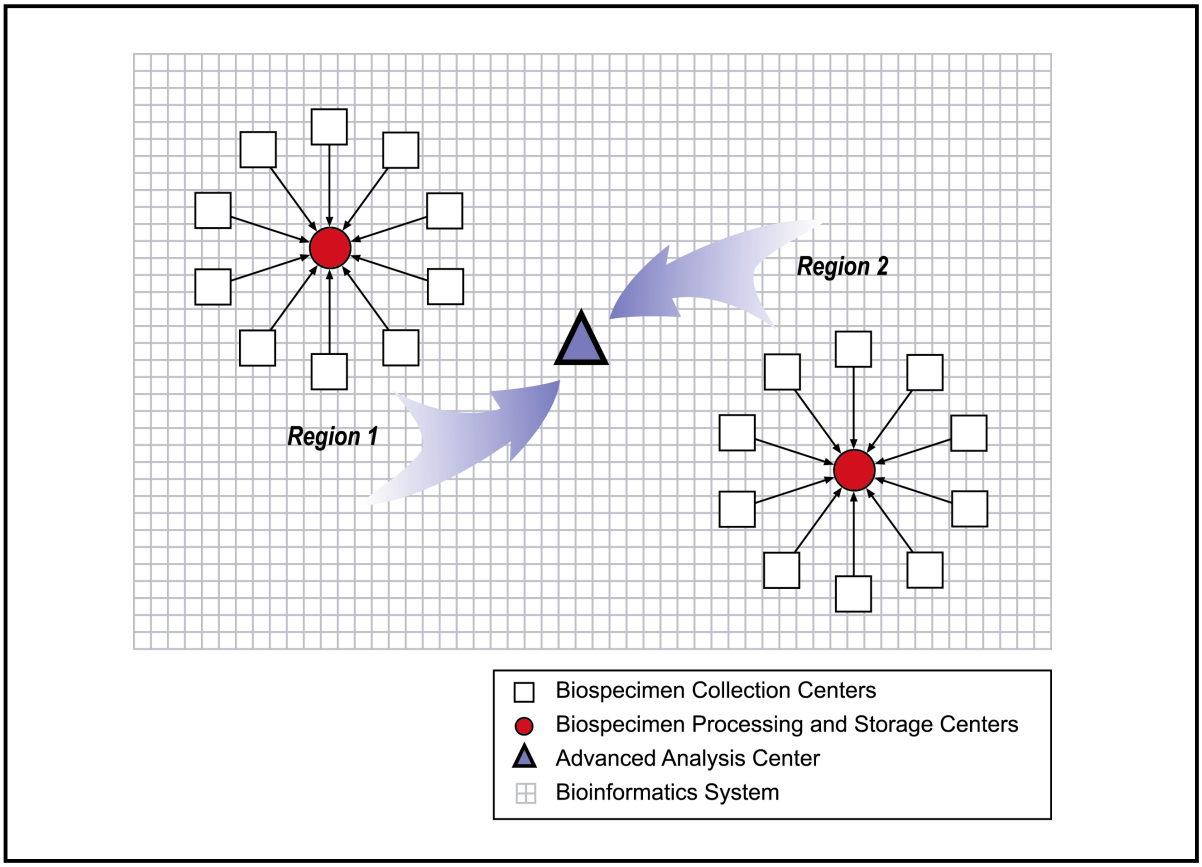

Figure 6-3. Hypothetical Model for NBN Biospecimen and Data Collection and Distribution

The funding and governance of the Bioinformatics system could be part and parcel of the overall NBN governance structure or an independent entity that has a contractual relationship with the NBN as a whole. As an independent entity, the NBN information system could be a commercial, for-profit enterprise, a not-for-profit, or a governmental concern. Even if it is governmental or not-for-profit, commercial firms could be contracted to do some or all of the work. For example, the NBN entity could design the system architecture and hire a firm to create the data dictionary or program the reports.

An alternative model is for the government to develop a base system and have another entity perform maintenance, administration, and enhancements. Alternatively, the roles could be reversed, in whole or in part. Finally, these arrangements could be changed after a set period (e.g., 3 years).

The Patient Relations Business Unit will be responsible for communicating with healthcare consumers to answer questions about the uses for their type of specimen, the results of research studies, and their eligibility for studies. See *5. Communications* for further discussion of its functions. Donors will not have access to research results directly related to *their* specimen, but only to a class of data (e.g., results of a particular study in women aged 50 to 69 with pancreatic cancer). This topic is further discussed in *2. Management of Ethical and Legal Issues.*

This business unit will handle other medical concerns directly relevant to the donor's participation in the NBN that their health providers would not be able to answer. It is likely that a variety of communications modalities will need to be developed, including direct telephone inquiry handling, referral to NBN constituent organizations (including the survivor organizations), Web sites and listservs, brochures, newsletters, etc.

This business unit may play a role in promoting appropriate consent processes and measuring overall satisfaction, and thus improving customer satisfaction, with the NBN. Marketing the NBN, both conceptually and practically, will be an important function. This business unit is likely to be a single entity with strong ties to many other organizations.

6.4 Summary of Key Findings and Recommendations

The NBN governance and business models must include strategies that permit the NBN to rapidly establish a firm foundation and facilitate its growth into a self-sustaining organization. The strategies must be sufficiently defined to be realizable, yet sufficiently flexible to meet the likely evolution of research opportunities and needs of the NBN discussed elsewhere in this report. It is unlikely that the NBN will be an entirely governmental project, in either management or funding.

The key recommendations include:

- NBN governance should be organized at three levels to include a Board of Governors, the NBN Operations Center, and its business units. The proposed business units would include the following: Research Administration and Support; Biospecimen and Data Acquisition; Storage and Distribution/Basic Analysis; Advanced Analysis; Bioinformatics and Data Management; and Patient Relations.
- Any private or public entity—including existing businesses or research consortia and tissue repositories—would be free to bid on business unit work (their own governance structures permitting).
- Because funding sources are expected to a be a mix of public and private (charity and fees), the overall funding structure must be able to accept funds from a variety of sources; hence the NBN Operations Center should be a not-for-profit entity, possibly incubated in the FNIH or within the NDC itself. Not-for-profit status of the NBN Operations Center brings the added advantage of enhanced credibility and maximizing public trust.
- QA, Bioinformatics and Data Management, and Communications are core efforts that will be critical and would be in the NBN Operations Center. Bioinformatics/Data Management and Communications will also be represented in the business units.
- A sound business plan requires that there be incentives to participation at multiple levels. Open communication and true partnerships with existing institutions are necessary to maximize utility, overcome barriers, and facilitate participation.

Module 7

National Biospecimen Network and Public Health

7.

National Biospecimen Network and Public Health*

Two essential elements contribute to the risk of all disease, including cancer: Genetic factors and environmental exposures. The completion of the first draft of the human genome will allow researchers to characterize genotypes to fine degrees of detail, and identify genetic traits associated with individuals. However, much of the genetic variation that is associated with cancer risk appears to modify risk only in the presence of environmental variability, both for exposures that increase risk and exposures that decrease risk. There are 10- to 200-fold differences in rates of disease when comparing different geographic locations around the world, and over 50 years, up to 10-fold when comparing the same geographic location over time. These differences cannot be explained by differences in genes, but only by differences in exposures, modified by the interaction between genetic variation and such exposures.

There exists a broad array of exposures that may influence the presence or absence of disease in humans. The accurate characterization and measurement of many of the environmental exposures is difficult, but there is extensive experience in the epidemiologic community. The human species has adapted to a wide variety of different environments, cultures, diets (both marginal and excessive), microorganisms and parasites, toxic exposures, and bad habits. There are thus a wide variety of susceptibilities to, and protections against, these exposures. Accordingly, an evaluation of variations in environmental exposures is necessary, along with measurements of genetic variation, to give a true picture of the causes of disease.

In addition, the characterization of the disease phenotypes is still problematic, with a myriad of classification schemes for different organs and systems that ranges from precise molecular characterization to vague syndromes. Exactness in description of disease phenotypes is necessary to identify the reasons for increased or decreased disease risk. The greater the degree of precision of phenotypic classification, the higher the likelihood of being able to reduce susceptibility or increase resistance (prevention), to detect early disease, and to treat at the earliest opportunity. The key to obtaining phenotypic precision is detailed outcome data. As more outcomes accumulate with time, and as classification schemes improve, the opportunities to define and redefine homogeneous phenotypic subsets will improve.

To attempt to establish the complete pattern of human disease susceptibility and resistance, and to identify more precise phenotypes, what is needed is a study of a very large number of ethnically diverse individuals who are well characterized genetically, whose exposures are well mapped, and whose illness pattern and mortality can be monitored. This cohort, labeled the "Last Cohort," would examine the impact of exposures on causes and rates of disease, and study the interaction of these exposures with genetic variation. Blood samples would be collected from healthy individuals, along with detailed information about each individual—including behaviors (e.g., diet, smoking, exercise), medical history, reproductive history, family history,

* Presented by Dr. John Potter at the July 28, 2003 NBN Blueprint Meeting, Bethesda, MD.

demographics, and geographical location. Over time, a certain proportion of these healthy individuals (see below) would develop diseases. With the passage of years, researchers would use the blood samples collected from the healthy donors—plus the additional exposure data on these donors—to go back and look for early markers of the disease. The Last Cohort, therefore, represents the opportunity to learn more about etiology (both genetic and environmental) of very tightly defined disease entities; about early detection (serial blood specimens will provide opportunities to establish proteomic marker profiles, cross-sectional and noting changes over time); and, ultimately, about prevention.

The size of the cohort is determined by the degree of human genetic variation, the degree of variation in exposures, and the size of the specific sets of outcomes to identify. It is estimated that approximately 500,000 Last Cohort participants would be needed to achieve the desired results. After 6 years of follow-up, for instance, a healthy cohort of this size aged 50 to 75 would experience about 40,000 cancers and 55,000 deaths.

The value, over the long term, would be a disease classification system (derived in particular from other work that the National Biospecimen Network [NBN] would facilitate) that would allow researchers to divide cancer types into subsets, based on molecularly defined homogeneous phenotypes. A large, long-term epidemiologic study such as this, tied to genomics and proteomics with linkage to prediagnostic serum and data and the capacity to follow up for outcomes of interest, would help researchers answer a number of questions:

- What is the association between specific exposures and molecularly defined disease risk? (Paradigm: Smoking and lung cancer)
- What is the association between specific allele variants and disease risk? (Paradigm: Adenomatous polyposis coli (APC) gene truncation mutation and colon cancer)
- What aspects of the interaction between exposure and genetic variants influence disease risk? (Paradigm: Folate/ 5,10 methylenetetrahydrofolate reductase (MTHFR) variants and colon polyps)
- What aspects of the proteome profile distinguish those with and without disease? (Paradigm: Prostate-specific antigen (PSA) and prostate cancer)

How does this relate to the NBN? The most difficult and expensive part of establishing such a cohort from scratch is collecting fresh specimens for proteomics, mRNA expression, etc., in order to define the outcomes as precisely as possible. The setting up of the NBN specifically to collect fresh tissues for these and other purposes markedly enhances the ability to ensure the collection of such specimens from the cohort members. Further, the establishment of the Last Cohort means that, in addition to providing specimens for drug development, sub-classifying outcomes, etc., the NBN will greatly increase both our understanding of causes and our capacity to develop serum markers for early detection.

If the right technology in put in place over the next 5 to 10 years to do high throughput genomic sequencing (for susceptibility and resistance) and proteomics (for screening) on very large populations, but an infrastructure is not in place to best exploit these gains, a significant opportunity will be lost. The NBN could provide this infrastructure to the Last Cohort study. The NBN would provide, by virtue of its biorepositories, extensive material capture, systematic

collection strategies, a centralized approach to ethical issues, and long-term follow-up. For the Last Cohort, the NBN could provide, in addition, at a subset of its collection sites, the linkage to prediagnostic exposure data and serum needed to undertake the studies such a cohort would allow.

One way to conceptualize this is to imagine that the specimens collected by the NBN would be derived from a virtual population cohort that will remain undefined. The Last Cohort proposal would establish a real cohort so that it represents a very well-defined subset of this virtual cohort —see Figure 7-1.

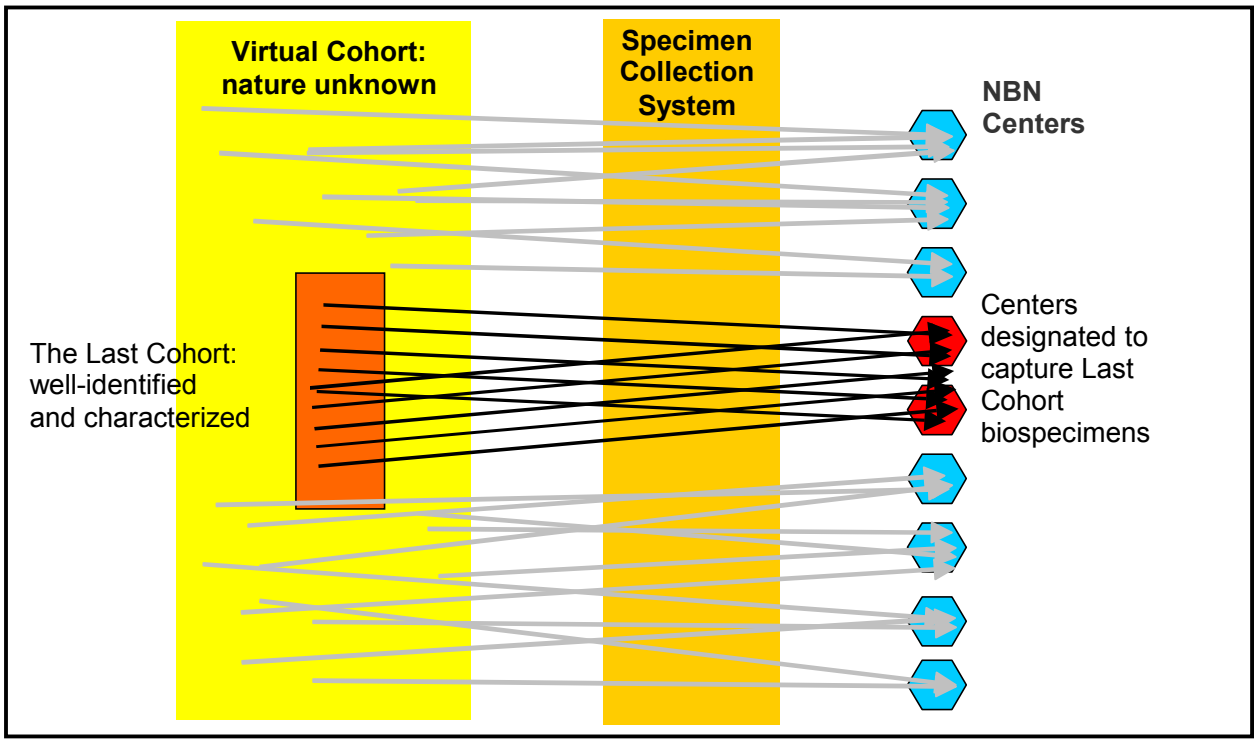

Figure 7-1. The Last Cohort

Admittedly, this would increase the cost of the NBN (estimates are a two- to three-fold increase), which plans to collect specimens from far fewer than 500,000 individuals. It is worth noting, however, that the number of end-point specimens collected for the Last Cohort will be modest (about 40,000 over the first 6 years, for instance); the additional costs arise from the recruitment of the cohort and the collection of baseline data and blood samples. The Last Cohort presents a unique opportunity to understand the causes of human disease, its prevention and early detection, and should be considered.

Module 8

Demonstration Project

8.

Demonstration Project

Participants at the July 28–29, 2003 Design Team meeting were asked to suggest a possible model for an NBN demonstration project. During the meeting, it became apparent that there were many possible approaches for initiating a demonstration project. The model described below provides a preliminary framework based on the discussion at that meeting. It is anticipated that this framework will require further refinement.

The ultimate implementation of the National Biospecimen Network (NBN) is understood to be a complex undertaking. It therefore will be important to evaluate the feasibility of integrating multiple aspects of a national network by beginning with a demonstration project. The NBN Design Team recommended establishing a 3-year demonstration project, supported by a combination of public and private funds, with several important characteristics outlined in Table 8-1. The demonstration project should be able to perform the basic functions of the NBN, address specific challenges discussed in this Blueprint report, and meet performance characteristics to evaluate success. In addition, the demonstration project should be capable of quickly expanding to operate on a national scale.

In considering the demonstration plan, the Design Team recognized the intrinsic value of incorporating the potential of existing biospecimen resource components as a backbone for the new system. As a starting point, the Design Team considered strengths in existing NCI resources in tissue collection and bioinformatics, and expertise in patient privacy issues. In addition, the expertise of private sector entities in establishing business operations and systems to collect, process, and distribute biospecimens to the research community is also highly valued. The Design Team considered and recommended a business model for the demonstration project that included the use of contracting mechanisms to build on these existing resources.

One of the major challenges in managing the demonstration project will be integrating and coordinating the diverse activities of the six proposed Business Units. Therefore, it will be critical to identify a leader for the demonstration project to coordinate these activities. The NBN demonstration project leader would be responsible for insuring that milestones are met consistently, and anticipating next steps. This individual could draw upon the resources of a biomedical incubator such as the Foundation for the NIH (FNIH) and might ultimately serve as the Chief Executive Officer of the NBN if the demonstration project were successful.

The NBN demonstration project should begin to disseminate biospecimens and associated data to researchers after one year of operation, and should aim to distribute 15,000–30,000 biospecimens annually after 3 years. The Design Team estimated that a demonstration project could cost approximately $15-20 million to develop and operate in the demonstration phase, and suggested that the NBN identify creative partnering opportunities with the public and private sectors to provide financial and organizational support for the initiative.

Table 8-1. Characteristics of NBN Demonstration Project

	DEMONSTRATION > > > > > IMPLEMENTATION		
Module	**Basic NBN Functions**	**Specific challenges**	**Performance Characteristics**
Governance and Business Models	• Identify Board of Governors • Establish Operations Center, hire key executive staff • Establish oversight-level QA, bioinformatics and communications functions • Coordinate establishment of all functional business units	• Develop detailed business models, examining role of reimbursements and other fees • Define milestones/performance characteristics for a formal evaluation of the NBN demonstration project by function and stage of development	• Evaluate cost recovery efforts • Evaluate contributions to publications, scientific discoveries, patents, and technology development
Management of Ethical and Legal Issues	• Establish bioethics and legal board • Establish baseline consent procedures	• Revise/expand/optimize consent procedures as needed based on experience and any new regulations	• Assess compliance with Federal, state, and local regulations
Biospecimen and Data Collection and Distribution	• Develop and implement Standard Operating Procedures (SOP) for collecting and annotating biospecimens • Define common data elements for biospecimen annotation • Develop criteria for distributing specimens to researchers that would eventually be adopted by the Biospecimen Utilization Committee	• Insure that SOP development is integrated with Bioinformatics and Quality Assurance efforts. • Investigate approaches (i.e. follow-up questionnaires and/or linking to tumor registries) to collect longitudinal data with biospecimens • Create incentives to encourage data submission from users • Perform genomic characterization on a subset of biospecimens and provide this information to the research community	• Assess longitudinal data collection (i.e. success rate of patient followup) • Document number of specimens collected and distributed • Evaluate the availability and utility of genomic and proteomic data • Test scalability of biospecimen collection system

	DEMONSTRATION > > > > > IMPLEMENTATION		
Module	**Basic NBN Functions**	**Specific challenges**	**Performance Characteristics**
Bioinformatics and Data Management	• Work with Biospecimen and Data Acquisition Business Unit to develop common data elements and informatics applications • Define minimal data set, associated data model, dictionary and baseline reports; deliver first functioning demonstration system	• Integrate work with Quality Assurance efforts • Integrate best functions of existing systems • Expand data model in context of advanced analysis needs	• Assess compliance with industry standards • Evaluate facilitation of all business unit operations (internal) • Assess facilitation of researcher access to clinical and specimen related info (external) • Review effectiveness of analytical tools to further the research goals of users
Communications	• Work with Specimen and Data Acquisition Business Unit to develop informed consent documents. • Develop an NBN communications strategy that addresses marketing, public relations and outreach issues	• Sponsor focused investigations of ethical and legal issues to develop innovative approaches for obtaining consent and protecting patient privacy within the NBN • Develop incentives to increase and retain participation • Recruit more researchers, tissue donors, and (possibly) funders	• Assess patient recruitment for specimen donation • Document number of researchers applying to use specimens through the Biospecimen Utilization Committee

In order to streamline the development and integration of disparate aspects of the NBN, the Design Team recommended focusing the demonstration project on approximately five tumor types. The Design Team suggested that the tumor types for the demonstration project should be selected by weighing multiple characteristics including: incidence, mortality, tissue availability and quantity, as well as the availability of diseased and normal (or normal adjacent) tissues. In addition, tumors included in the demonstration project should require specimen acquisition sites to overcome diverse technical procurement challenges, including the impact of technical advances (such as needle biopsies and neoadjuvant therapy) upon tissue availability. After

completing the demonstration project focused on five tumor types that meet all of these criteria, NBN specimen acquisition sites could readily expand to collect a broader range of tumors.

In addition to addressing challenges involved in biospecimen collection, the demonstration project must develop approaches to encourage use of the resource by basic and clinical researchers. The NBN demonstration project should initiate concerted communication efforts to identify scientists with ongoing scientific research projects that would benefit from analyses of biospecimens from the NBN. These researchers could provide timely feedback that could be included as part of a formal evaluation of the demonstration project, and which would provide an early indication of the promise for implementing the NBN on a national level.

References

References

Arias E. and Smith B. (2003). *Deaths: Preliminary Data for 2001.* National Vital Statistics Reports, Vol. 51, No. 5 (March 14), Centers for Disease Control and Prevention.

Asian Technology Information Program (2003). *ATIP03.042*: *The Biobank Japan Project*. July 23

Berman J.J., Edgerton M.E., and Friedman B.A. (2003). The tissue microarray data exchange specification: A community-based, open source tool for sharing tissue microarray data. *BMC Med. Inform. Decis. Mak.* 3:5.

Boulware L.E., et al. (2002). Understanding disparities in willingness to donate blood and cadaveric organs. *Medical Care,* Vol. 40:85-95.

Boulware L.E., et al. (2002). Determinants of willingness to donate living related and cadaveric organs: identifying opportunities for intervention. *Transplantation,* Vol. 73:1683-1691.

Brooks F.P., Jr. (1975). *The Mythical Man-Month.* New York: Addison-Wesley.

Callender C.O., et al. (2002). Increasing living donations: Expanding the national Mottep commnirt Grassroots model. *Transplantation Proceedings*, Vol. 34, No. 7 (November 2): 2563-2564.

Callender C.O., et al. (2002). National MOTTEP: Educating to prevent the need for transplantation. *Ethnicity & Disease*, Vol. 12, No. 1 (Winter): SI34-SI37.

Department of Health and Human Services. (2003). *HHS Launches New Efforts to Promote Paperless Health Care System.* Press Release (July 1).
www.hhs.gov/news/press/2003pres/20030701.html

Department of Health and Human Services. (2003). *Institutional Review Boards and the HIPAA Privacy Rule*. NIH Publication Number 03-5428.
privacyruleandresearch.nih.gov/IRBandprivacyrule.asp

Department of Health and Human Services, Office of the Secretary. (2002). 45 CFR Parts 160 and 164. *Standards for Privacy of Individually Identifiable Health Information; Final Rule.* Federal Register 67 (157):53182-53273. (August 14).

Department of Health and Human Services. (2003). *Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule*. NIH Publication Number 03-5388.
privacyruleandresearch.nih.gov/pr_02.asp

Eiseman E., Brower J., Olmsted S., Clancy N., and Bloom G. (2003). *Case Studies of Existing Human Tissue Repositories: "Best Practices" for a Biospecimen Resource for the Genomic and Proteomic Era.* RAND Science and Technology. (August 28).

Eiseman E. and Haga S.B. (1999). *Handbook of Selected Human Tissue Sources: A National Resource of Human Tissue Samples*. Santa Monica, CA: RAND.

Eisen A. and Weber B.L. (2001). Prophylactic Mastectomy for Women with BRCA1 and BRCA2 Mutations – Facts and Controversy. *N. Engl. J. Med.*, Vol. 345, No. 3 (July 19): 207-208.

Faden R.R. and Beauchamp T.L. (1986). *A History and Theory of Informed Consent,* New York, NY: Oxford University Press.

Feskanich D., Ziegler R.G., Michaud D.S., Giovannucci E.L., Speizer F.E., Willett W.C., et al. (2000). Prospective study of fruit and vegetable consumption and risk of lung cancer among men and women. *J. Nat. Cancer Inst.,* Vol. 92:1812-23.

Flood A., Velie E.M., Chaterjee N., Subar A.F., Thompson F.E., Lacey J.V., Jr., et al. (2002). Fruit and vegetable intakes and the risk of colorectal cancer in the Breast Cancer Detection Demonstration Project follow-up cohort. *Am. J. Clin. Nutr.,* Vol. 75, No. 5:936-43.

Goldbohm R.A., Van den Brandt P.A., and Dorant E. (1994). Estimation of the coverage of Dutch municipalities by cancer registries and PALGA based on hospital discharge data. *Tijdschr. Soc. Gezondheidsz*, Vol. 72: 80-84.

Hirota S., Isozaki K., Moriyama Y., et al. (1998). Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science* Vol. 279, No. 5350 (January 23): 577-580.

Jack A.L. and Womack C. (2003). Why surgical patients do not donate tissue for commercial research: A review of the records. *BMJ*, Vol. 327, No. 7409: 262.

Kincaid H., Crichton D., Winget M., et al. (2003). *A National Virtual Specimen Database for Early Cancer Detection.* Paper prepared for presentation to the Sixteenth IEEE Symposium on Computer-Based Medical Systems, June 25-27, 2003, New York.

Knox K. and Ratcliffe C. (2002). *A Strategic Framework for Establishing a National Cancer Tissue Resource for Cancer Biology and Treatment Development.* UK: National Translational Cancer Research Network Coordinating Centre (September).

Maurer S.M., Firestone R.B., and Scriver C.R. (2000). Science's neglected legacy: Large, sophisticated databases cannot be left to chance and improvisation. *Nature,* Vol. 405 (May 11): 117-120.

McCaffrey P. (2003). Iceland's database tussle. *Bio-IT World*. (April 1). *www.bio-itworld.com/news/040103_report2255.html*

McDonald C.J., Huff S.M., Suico J.G., et al. (2003). LOINC, A Universal Standard for Identifying Laboratory Observations: A 5-Year Update. *Clinical Chemistry,* Vol. 49:624-633.

Medical Research Council. (2002). The UK Biobank study gets funding go-ahead. January – June 2002 news archive. *www.mrc.ac.uk/txt/index/public-interest/ public-news-4/public-news_archive/public-news_archive_1_2002/public-biobank_uk.htm*

National Bioethics Advisory Commission. (1999). *Research Involving Human Biological Materials: Ethical Issues and Policy Guidance, Volume I: Report and Recommendations of the National Bioethics Advisory Commission.* Rockville, MD (August).

National Cancer Institute. (2002). *The Early Detection Research Network: Translational Research to Identify Early Cancer and Cancer Risk.* Second Report. Division of Cancer Prevention. (October). *www3.cancer.gov/prevention/cbrg/edrn/edrn_report2002.pdf*

Report of the Brain Tumor Progress Review Group (2000) to the National Cancer Institute and the National Institute of Neurological Disorders and Stroke. (November). *prg.nci.nih.gov/pdfprgreports/2000braintumor.pdf*

Report of the Gynecologic Cancers Progress Review Group (2001) to the National Cancer Institute. (November) *prg.nci.nih.gov/pdfprgreports/gynreport.pdf.*

Report of the Leukemia, Lymphoma, and Myeloma Progress Review Group (2001) to the National Cancer Institute. (May) *prg.nci.nih.gov/pdfprgreports/2001llm.pdf.*

Report of the Lung Cancer Progress Review Group (2001) to the National Cancer Institute. (August). *prg.nci.nih.gov/pdfprgreports/2001lung.pdf*

Singh P.N. and Fraser G.E. (1998). Dietary risk factors for colon cancer in a low-risk population. *Am. J. Epidemiol.,* Vol. 148: 761-774.

Spinney L. (2003). UK launches tumor bank to match maligned Biobank. *Nat. Med.*, Vol. 9, No. 5 (May): 491.

Triendl R. (2003). Japan launches controversial Biobank project. *Nat. Med.*, Vol. 9, No. 8 (August): 982.

U.S. Cancer Statistics Working Group. *United States Cancer Statistics: 1999 Incidence*. Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention *and* National Cancer Institute; 2002.

Verble M. et al. (2002). A multiethnic study of the relationship between fears and refusal rates. *Prog. Transplant.,* Vol. 12, No. 3 185-190.

Zitner A. (2003). Whose DNA is it, anyway? Many people, hoping for medical advances, give genetic material. But some researchers' refusal to share samples has donors up in arms. *Los Angeles Times* (July 18).

Glossary

Glossary

Annotation—explanatory or extra information associated with a particular biospecimen. Annotations may be added either by the pathologist or resource collector.

Biospecimen—human tissue including everything from subcellular structures like DNA to cells, tissue (bone, muscle, connective tissue, and skin), organs (liver, bladder, heart, and kidney), blood, gametes (sperm and ova), embryos, fetal tissue, and waste (urine, feces, sweat, hair and nail clippings, shed epithelial cells, and placenta).[1]

Central Institutional Review Board (CIRB)—a model that features a "facilitated review" process that streamlines local IRB review for national multicenter cancer treatment trials. Local IRBs can access CIRB reviews and decide whether or not to utilize the CIRB review for a particular protocol. If there are no concerns about local context, the Chair (or subcommittee) of the local IRB can decide to accept the CIRB review in lieu of a full board review.

CGxP—refers collectively to Current Good Laboratory Practices, Good Clinical Practices, and Good Manufacturing Practices, as defined by the Food and Drug Administration, to describe what records should be kept or what information is considered to be an official record.

Coded samples—also designated as "linked" or "identifiable," are supplied by repositories to investigators from identified specimens using a code rather than personal identifying information, such as a name or a Social Security number.

Confidentiality—a principle emergent from a relationship in which something about an individual, information, or material has been shared (with some degree of loss of privacy) in confidence.

Covered entity—a health plan, healthcare clearinghouse, and certain healthcare providers that transmit protected health information electronically for certain covered transactions, as defined in the Health Insurance Portability and Accountability Act of 1996.

Data—values derived from scientific experiments or diagnostic procedures organized especially for scientific analysis, in a numerical form suitable for processing by computer.

Human subject—a living individual about whom an investigator conducting research obtains (1) data through intervention or interaction with the individual or (2) identifiable private information.

Identifiable—linked to personal information in such a way that the person from whom the material is obtained could be identified by name, patient number, or clear pedigree location (i.e., his or her relationship to a family member whose identity is known). Identified samples are

[1] Definition adapted from Eiseman E. and Haga S.B. (1999). *Handbook of Human Tissue Sources: A National Resource of Human Tissue Samples*. Santa Monica, CA: RAND.

supplied by repositories from identified specimens and have a personal identifier (such as a name or patient number) that allows the researcher to link the biological information derived from the sample directly to the individual from whom the material was obtained.

Incidence—number of cases of disease having their onset during a prescribed period of time.

Informed consent—an educational process between the investigator and the prospective subject (or the subject's legally authorized representative) as a means to ensure respect for persons; mutual understanding of research procedures, risks, rights, and responsibilities; and continuous voluntary participation.

Matching (uninvolved) tissue—usually refers to the matching tissue type from which the tumor developed. Looks histologically normal but may demonstrate molecular changes (e.g., terminal ductal alveolar unit from patients with breast cancer).

Metastatic lesion—metastases to lymph nodes or distant organs; direct extensions do not satisfy such requests from most investigators.

Minimal risk—the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests.

Network—a collection of activities and organizations that use shared data standards and business rules to carry out a common purpose.

Normal tissue—used as the control for a specific type of tumor but comes from a patient without this type of cancer.

Patent—a property right granted by the Government to an inventor. In order to be patentable, an invention must contain an idea that serves some utility, is novel, and is "nonobvious" to an average person who is "skilled in the arts of the specified field."

Prevalence—number of cases of a disease, infected persons, or persons with some other attribute present during a particular interval of time.

Primary tumor—the first site at which a tumor arises. This may not be clear for diffuse neoplastic processes such as lymphoma.

Privacy—the state or condition of limited access to an individual and/or to information about that individual.

Protected health information—any health information that is collected by a covered entity and is individually identifiable.

Quality assurance (QA)—an integrated system of management activities involving planning, implementation, documentation, assessment, and improvement to ensure that a process or item is of the type and quality needed for the project.

Quality control (QC)—specific tests defined by the QA program to be performed to monitor procurement, processing, preservation, storage, and specimen quality and to test accuracy. These may include but are not limited to performance evaluations, testing, and controls used to determine accuracy and reliability of the repository's equipment and operational procedures, as well as monitoring of the supplies, reagents, equipment, and facilities.

Research—systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge.

Sample—a subset of a specimen. Researchers requesting access to tissues in the repository are usually provided with a sample of that specimen.

Specimen—a sample, as of tissue, blood, or urine, used for analysis and diagnosis. A single biopsy may generate several specimens, including a number of slides, paraffin blocks, and frozen specimens.

Tissue procurement—a fee paid for the right to acquire tissue specimens from a repository, as in a vendor relationship, with no further obligation to jointly publish with the tissue provider.

Tissue sharing—collaboration whereby no or nominal funds are exchanged for the privilege of analyzing tissue specimens with the purpose of publishing findings with the collector of the tissue specimens.

Unidentified or "anonymous" samples—samples supplied by repositories to investigators from a collection of unidentified human biological specimens and can never be traced to an individual. Unlinked or "anonymized" samples lack identifiers or codes that can link a particular sample to an identified specimen or a particular human being but may have been derived from an identified sample in the repository.

Appendices

Appendix A

Case Study:

The Role of Biospecimens for Discovery of a Targeted Cancer Therapy

Human epidermal growth factor receptor-2 (HER2; also called ErbB2 or Neu) is a cell-surface protein involved in cell development. In normal cells, HER2 controls aspects of cell growth and division. Activation of HER2 in cancer cells, however, accelerates many cellular processes associated with tumor formation, including cell proliferation, angiogenesis, adhesion, and resistance to chemotherapy. In 1987, Dennis Slamon and colleagues examined tumor samples from 189 breast cancer patients enrolled in an ongoing study and discovered that the gene that codes for HER2 is amplified in 20 to 30 percent of human breast cancers. Two years later, upon examining the gene and its RNA and protein products in more than 650 frozen and paraffin-embedded human breast cancer samples, Slamon and colleagues demonstrated that amplification of the HER2 gene correlates strongly with poor clinical prognosis. Slamon also observed that tissue samples subjected to freezing processes retained more protein activity for analysis than did samples prepared using paraffin-based or other fixative methods. As a result of these groundbreaking observations, which were made possible by the use of tissue specimens, HER2 has subsequently become a pivotal biomarker in breast cancer and has promoted the development of the revolutionary anti-cancer drug, trastuzumab.

The development of trastuzumab (Herceptin®; Genentech, San Francisco, CA) is a success story that demonstrates the potential of biomarkers in the rational design and development of cancer drugs. Trastuzumab is a recombinant monoclonal antibody directed against the extracellular domain of HER2. Designed specifically for those breast cancer tumors that overexpress HER2, trastuzumab has been highly effective as a single agent and in combination with other standard chemotherapy regimens for certain types of breast cancer. Based on trastuzumab's clinical trial success, the FDA approved the drug for treatment of HER2-positive metastatic breast cancer in 1998.

However, in all clinical trials with trastuzumab prior to approval, the antibody was studied only in patients whose tumors tested "HER2-positive." This classification was determined using a set of research-grade immunohistochemistry (IHC) assays, in which the tissue sample is treated with an antibody specific to the protein of interest. This antibody binds to the protein and is then reacted with a second, specific antibody that contains a fluorescent or chemiluminescent "tag," allowing the protein-antibody complex to be visualized as it naturally appears in the tissue. Access to banked tissue samples proved pivotal in translating these research assays into an FDA-approved assay (HercepTest) that could be used in a wide range of pathology laboratories. NCI's Cooperative Breast Cancer Tissue Resource provided 1,200 specimens that were used to demonstrate adequate concordance between the test methods. Following initial

approval of trastuzumab, the development of more sophisticated approaches to assess HER2 status has confirmed that patients who display high levels of HER2 overexpression (measured using IHC) or an amplification of the HER2 gene (measured using fluorescence *in situ* hybridization (FISH) techniques) receive the greatest clinical benefit from the drug. In one clinical trial evaluating trastuzumab as monotherapy in women with advanced metastatic breast cancer, all responding women who were moderately IHC-positive also tested FISH-positive, suggesting that the combination of IHC and FISH can further identify patients who may benefit from trastuzumab.

Access to tissue samples has proven crucial to the development of trastuzumab, from the initial observation that linked HER2 level with poor outcomes in metastatic breast cancer to the identification of target patient populations for clinical trials and response to therapy with the approved drug. The success of trastuzumab proves that biomarker-based patient selection at an early stage in the clinical trial process can optimize the development of successful cancer therapy. In retrospect, the clinical benefits of trastuzumab would almost certainly have been insufficient for FDA approval if the agent had been tested in unselected patient populations. Moreover, molecular insights gained from clinical trials with trastuzumab will promote research with new biomarkers, as well as the development of future therapies using monoclonal antibodies and other rationally designed drugs.

Appendix B

Case Study:

Biospecimens for Validation of a Biological Target and Discovery of Alternative Therapeutic Uses

Imatinib mesylate (STI-571, Gleevec®; Novartis, Basel, Switzerland), a small molecule developed originally for the treatment of chronic myeloid leukemia (CML), has become a model drug for the validation of molecular targets. Although designed to inhibit the tyrosine kinase (TK) activity associated with the CML-specific protein, BCR-ABL, imatinib mesylate demonstrated a similar effect on the activity of the proto-oncogenic TK, KIT (also called stem cell factor receptor). In 1998, Hirota and coworkers used immunohistochemistry to identify KIT expression in 49 gastrointestinal stromal tumors (GIST).[1] DNA analysis on a subset of these samples using reverse transcriptase-polymerase chain reaction (RT-PCR) revealed that KIT expression often correlated with specific mutations in *c-kit*, the gene that codes for KIT. Activation of KIT was then hypothesized as a critical step in the pathogenesis of GIST, suggesting another possible use for imatinib mesylate.

Clinical use of imatinib mesylate in patients with GIST proved dramatic, as 60-70 percent of appropriately selected GIST patients responded to the therapy. These results were all the more remarkable, given the difficulty of treating GIST, a cancer that is essentially completely resistant to other systemic therapies. Based on a rapid series of highly successful clinical trials, the FDA approved imatinib mesylate for GIST in 2001, less than 1 year after the agency approved the drug as therapy for CML. As with the case of trastuzumab (Herceptin®), successful treatment with imatinib mesylate is predicated upon identifying those patients whose tumors indicate the appropriate biologic aberrations.

Standardizing the collection and analysis of tissue samples is a key factor for selecting appropriate GIST patients for treatment. Samples are assayed for the presence of several specific mutations in *c-kit* that are linked with uncontrolled TK activity. Yet the percentage of GISTs reported to have the appropriate *c-kit* mutations varies widely. Differences in assay techniques and the type of tissues used for DNA extraction (e.g., paraffin-fixed samples versus frozen samples) affect the specificity and sensitivity of mutation detection, thereby affecting patient selection for therapy. The unqualified success of imatinib mesylate as therapy for GIST has been established. Delivering this therapy to every patient who will benefit requires a standardized

[1] Hirota S., Isozaki K., Moriyama Y., et al. (1998). Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science* Vol. 279, No. 5350 (January 23): 577-580.

methodology, the development of which will be greatly facilitated by a central repository of tissues.

The Gleevec story proves the concept of validation of a drug target and demonstrates that a cancer drug approved for one indication may find a "second career" as an agent for other cancers with similar etiologies. The success of imatinib mesylate is spurring the development of many novel agents with activity against cell lines that express KIT and BCR-ABL, many of which ultimately will be tested in combination with imatinib mesylate in clinical trials. However, as researchers unravel links between molecular pathways and specific cancers and treatments, thereby discovering yet untold uses for many existing therapies, standardization of tissue collection and analysis will become increasingly important for linking various independent observations. Imatinib mesylate has demonstrated that a therapeutic agent may have multiple specific applications, validating molecular targets along the causal pathway and leading to the development of more specific and more potent therapies in the future. As with the case of imatinib mesylate, the systematic use of a variety of tissue samples will serve a pivotal role in translating observations from the laboratory into benefits for the cancer patient.

Appendix C

Case Study:

Cancer Tissue Samples Key to Development of High-Precision Genomic Diagnostic Test

As scientists unlock the secrets of the human genome, tissues removed during surgery or biopsy can often indicate a patient's risk for cancer at increasingly early stages of disease progression. However, a tissue sample extracted from a breast cancer biopsy or a colon polyp is complex, as cancerous cells and healthy cells are often interspersed in intricate arrangements. Added to this challenge is the variety in the samples themselves; tissues are collected and preserved using a plethora of methods and protocols. Tissue samples thus have played a pivotal role in the development of laser capture microdissection (LCM), a breakthrough technique that facilitates the precise, reproducible, and accurate transfer of tissues for analysis using a wide variety of established analytical methods. The reproducibility and versatility of LCM will allow researchers to re-analyze tissue samples in light of new discoveries about regulatory pathways associated with various disease states.

In the future age of personalized medicine, it is envisioned that a person's genetic "fingerprint" will be the key to the diagnosis and treatment of illness. This fingerprint will be constructed using the analysis of a number of different components, which may include the patient's DNA or RNA profiles or the measurement of key proteins or other biomarkers. LCM facilitates the analysis of the "fingerprint" of diseased cells, thus providing insight into the ways these cells differ from normal cells. Selection of a pure cell population, such as a group of tumor cells, allows molecular "signals" from these diseased cells to be filtered from the background molecular "noise" of surrounding healthy tissue. This selectivity is accomplished by placing a thin transparent film over the tissue sample. The tissue is then visualized under a microscope, and the cells of interest are selected. A laser pulse applied to the selected cells transfers them to the film, which is then removed and placed directly into appropriate buffer for the analysis of DNA, RNA, or specific enzymes.

As a technique, LCM provides many advantages over conventional tissue microdissection. First, the one-step transfer process minimizes sample contamination. Following LCM, tissue samples retain integrity and can be analyzed using a variety of established genomic techniques and protein assays. Perhaps most importantly, however, is the true universality of the method; transfer is equally effective, regardless of tissue type. Frozen tissues, paraffin-embedded tissues, and cytology cell preparations are equally amenable to LCM transfer, as are all types of standard sample stains and fixatives.

LCM currently is being used in research on breast, prostate, and pancreatic cancers, diseases that traditionally have been diagnosed through histologic analysis of tissue samples. For example, recent research has shown that the genes *BRCA-1* and *BRCA-2* are associated with the familial

risk of breast cancer. Although these genes are present in healthy tissue, women who carry certain mutations in these genes have an elevated risk of developing breast cancer. Furthermore, changes in DNA are frequently observed at these genetic loci during tumor development. Understanding how changes in these genes relate to the development of breast cancer requires comparative analysis of samples from healthy and diseased tissues. LCM allows researchers to isolate the cancerous tissue for studies of *BRCA* mutations. The ability to highlight tumor cells provides a clearer, more accurate picture of the tumorigenic events. Although LCM is not an analytical tool per se, it facilitates a more accurate analysis using a battery of complementary techniques.

Tissue samples measure disease progression from multiple vantage points by relating onset to a variety of changes within the cell. With LCM, countless relationships may be elucidated from serial analysis of one set of tissue samples. LCM demonstrates that tissue samples are important not only as targets for analysis, but also for the development of universal diagnostic tools that will revolutionize the way diseases are studied.

Appendix D

Interview Instrument for RAND Evaluation of Selected Existing Tissue Resources

A. GENERAL

1. Can you give us some history/background about your repository?
2. What was the main purpose for developing the repository?
3. Do you belong to a professional organization that deals with specimen collection?

B. BIOSPECIMEN COLLECTION, PROCESSING, AND STORAGE

Biospecimen Collection

The following questions cover your tissue collection techniques and standards:

1. Who are the sources of the tissues (e.g., patients, volunteers)? How were donors recruited?
3. Do you collect samples from minority populations? From the aged? From children? From donors outside the United States?
    - Do the proportions of samples contained in your repository reflect the ethnic diversity in the general U.S. population?
4. Why were the tissue samples originally collected (e.g., diagnostic purposes, research)?
5. Who is responsible for collecting the samples for the repository (surgeons, pathologists, researchers, trained repository personnel)?
6. How many people do you employ to collect, store, process and distribute tissue?
7. Where are the samples collected (community hospitals, academic medical centers)?
8. How do the samples get transferred to the repository? Or are they stored locally?
9. What kind of quality control (QC), auditing, and/or standardization is performed during the collection of tissues?
    - How are these standards ensured at the participating institutions that contribute tissue?
    - Are certain standards required of the institutions you agree to participate by contributing tissue? Are they dismissed if they do not adhere to these standards?

Biospecimen Storage

10. Can you share with us the storage techniques and standards you use for tissues in your repository?
11. How many and what types of tissues do you have in storage (e.g., tissues from cancer patients, tissues from patients with rare diseases, etc.)? Are tissues collected from normal controls?
12. What types of tissues are collected, stored, available along with samples of diseased tissue (healthy adjacent tissue, blood, serum)?
13. How are tissues stored? In liquid nitrogen freezers, mechanical freezers, or both?
    - What maintenance and backup procedures do you have for the freezers?
    - What are your procedures/standards for storage of biospecimens in mechanical freezers (-80º C) and in liquid nitrogen?
14. In what form are tissues stored (e.g., fresh frozen, paraffin block)? For how long?
15. Do you attempt to keep your stock of tissues at a certain level, either numerically or as a distribution of various types of samples? If so, what techniques are used to maintain these levels (e.g., recruitment of additional medical facilities)?
16. What kind of QC, auditing, and/or standardization is performed during the storage of tissues?

Biospecimen Processing and Annotation

17. Can you share with us the processing and annotation techniques and standards you use for tissues stored in your repository?
18. Is there a basic set of tests conducted on each sample to characterize the tissues? Gene arrays, DNA/RNA studies, immunohistochemistry (histopathology), other?
19. What kind of data are available about each sample? Clinical, longitudinal, pathology report, diagnostic, demographic (gender, ethnicity), medical history, family history, genetic profile, environmental exposure, treatment outcomes data, recurrence, survival, etc.?
20. If you collect longitudinal data, how is this information tracked?
21. What kind of QC, auditing, and/or standardization is performed during the processing and annotation of tissues?

C. CONSUMER/USER NEEDS

1. Can you tell us about your tissue distribution policies?
    - How do you review and prioritize requests for tissue?
    - How do you prioritize the distribution of rare/precious tissues?

2. How many tissues do you distribute and to whom?
3. Who are your consumers/users? From what types of institutions are your consumers/users? Academic? Industry? Government? What is the relative proportion of types of users (e.g., 60-percent academic; 40-percent industry)?
4. Do researchers from certain institutions or researchers conducting certain types of research receive priority?
5. Are any of your samples distributed internationally?
6. For what purposes have the stored tissues been used (e.g., cancer research, gene mapping, genomics/proteomics)?
7. Do you believe there are unmet users' needs that are beyond your control to meet? If yes, what are they?
8. What kind of QC, auditing, and/or standardization is performed during the distribution of tissues?

D. BIOINFORMATICS AND DATA MANAGEMENT

The following questions concern your data storage, distribution, and analysis techniques and standards:

1. For what do you use your bioinformatics system (e.g., tracking of collection, processing, distribution and analysis of samples, histopathological datasets, clinical/outcome information, results of research)?
2. Does your bioinformatics system contain any genomics/proteomics data?
3. What kind of QC, auditing, and/or standardization is performed on data entered into your data repository/bioinformatics system? Standardized data reporting and data entry?
4. Is the data repository (bioinformatics system) aggregated? Searchable? For example, can a researcher query the database to determine whether you have *X* numbers of samples for a particular disease?
5. Is your data repository set up so that automated extraction of information (data mining) is possible (i.e., is the data repository [bioinformatics system] “minable”)? (NOTE: Data mining is part of a larger process called knowledge discovery; specifically, the step in which advanced statistical analysis and modeling techniques are applied to the data to find useful patterns and relationships.)
6. Does the extent of the access to the data in the repository differ for these different groups of people (physicians, researchers, employees, insurers, employers), and if so, what are the differences?
7. Are any of the data from your repository about the tissue publicly available? Are these data available on the World Wide Web? Who has access to the data (e.g., physicians, researchers, insurers, employers)?

8. Do any of the data from research performed on the tissues feed back into the repository data system (e.g., reentry of relevant outcomes or research results)? If so, how do you validate the results that come back into the database?
9. What kinds of network security do you employ (e.g., encryption algorithms, firewalls, intrusion detection, etc.)?
10. How does the medical informatics system located at the collection site (i.e., hospital, laboratory) interface with the tissue bank's bioinformatics system?
11. Who develops your informatics systems? How many people do you employ in bioinformatics?

E. BUSINESS PLAN AND OPERATIONS

1. With whom or how were the arrangements made between the institutions that provide the tissues and the repositories? Was this arranged through the medical facility's administration or through individual doctors?
2. Can you share with us any lessons learned from setting up the medical institution/repository relationships? The donor/repository relationships?
3. Are you primarily involved in tissue banking, or do you collect tissues prospectively for distribution to particular researchers or for specific studies?
4. Is the tissue repository centralized, decentralized, or a centralized resource deployed through a virtual network of geographically dispersed tissue centers?
5. How is your repository funded? Privately? Publicly? Some combination of the two?
6. How much does it cost your organization to collect, process, store, and distribute tissue? Per sample? Yearly?
7. How much does it cost a researcher to obtain tissues from your repository?
8. Do you regularly evaluate whether the resource is used effectively, and if so, how do you measure this?
9. What procedures are in place for improving specimen collection, storage, annotation, and distribution based on the development of new technologies? How do you plan for new technology?
10. Have you had to deal with any institutional barriers to collecting, storing, distributing, or using these tissues for research purposes? If so, how did you overcome these?
11. Can you share with us any "best practices" and standard operating procedures that are in place for your repository, including those that address sample and data handling; ethical, legal, and social issues; and intellectual property issues?
12. Can you identify procedures within your existing operating structure that you would change? For example, if you were starting over what would you do differently (e.g., eliminate procedures, add procedures, etc.)?

13. Do you have a reporting mechanism to track the use of your specimens, including the number of specimens that are distributed each year?

14. Do researchers who use specimens from your repository cite/acknowledge the use of repository specimens in a standardized manner?

F. PRIVACY, ETHICAL CONCERNS, AND CONSENT ISSUES

Privacy Issues

1. What kind of personal/identifying information about the patient/donor is stored with the tissues? Are tissues in the repository:
   - Unidentified – i.e., identifiable information was not collected or, if collected, was not maintained and cannot be retrieved by the repository?
   - Identified – i.e., the tissues are linked to personal information in such a way that the person from whom the material was obtained could be identified by name, patient number, or clear family relationship?
2. When tissues are distributed to researchers, what kind of personal information is sent with them? Are the samples given to the researchers:
   - Unidentified samples – i.e., samples and data are supplied by repositories to researchers from a collection of unidentified tissues (sometimes called "anonymous")?
   - Unlinked samples – i.e., samples and data lack identifiers or codes that can link particular samples to an identified specimen or a particular human being (sometimes called "anonymized")?
   - Coded samples – i.e., samples and data are supplied by repositories to researchers from identified tissues with a code rather than with personally identifying information such as a name or Social Security number (sometimes called "linked" or "identifiable")?
   - Identified samples – i.e., samples and data are supplied by repositories from identified tissues with a personal identifier (such as a name or patient number) that would allow the researcher to link the biological information derived from the research directly to the individual from whom the material was obtained?
3. What policies and procedures do you use to ensure that patient/donor privacy is protected? How do you ensure confidentiality of patient information?

4. How have repository processes changed as a result of the new Health Insurance Portablity and Accountability Act (HIPAA) regulations?

5. What impact do state privacy laws have on your repository?

Consent Issues

6. What type of informed consent is obtained from each patient/donor? General consent for any type of research versus explicit/specific consent for an individual research project?

7. Does consent for tissue donation occur with or separate from consent for the treatment/surgical procedure?

8. Do you use a consent interview?
   - If so, what type of individual conducts the interview?
   - For whom does this individual work? The hospital? The repository?

9. Can you share with us a copy of your informed consent form and policies and procedures?

Institutional Review Boards

10. Does your repository have an institutional review board (IRB) or have IRB approval? From whom?

11. What kind of IRB approval is needed for researchers to use the samples?

G. INTELLECTUAL PROPERTY AND OTHER LEGAL ISSUES

1. What kinds of policies do you (the repository) have in place regarding intellectual property rights?

2. What rights do the submitting institution/researcher have to the tissue once it is given to the repository?

3. What rights do the individuals donating the tissue have once it is given to the repository?
   - Do they have access to their own tissue once it is donated (e.g., for medical purposes, for research purposes)?

4. Do you (the repository) retain any rights to the tissue once it has been transferred to the user?
   - Do researchers have to sign an agreement/contract to obtain samples?
   - Do you use a materials transfer agreement (MTA)? If so, can you share with us a copy of your MTA?

5. Are donors compensated in any way for their tissue?

6. What kinds of policies, if any, are in place regarding publication review and approval, proper acknowledgement of the resource, and reporting of publications (i.e., to help the repository measure the impact of what has come out of the resource)?

7. How do you address liability issues associated with collection, distribution, and use of samples in your repository (e.g., insurance, researchers sign a release, etc.)?

8. Has there ever been an incident or legal action involving the collection, distribution, and/or use of samples in the repository? If so, how this was resolved?

9. What kind of security do you use to ensure that the persons requesting tissues are legitimate?

H. PUBLIC RELATIONS, MARKETING, AND EDUCATION

1. How do you market your tissue resource to researchers (e.g., booths at scientific meetings, advertising in journals word of mouth)?
2. What kind of postresearch communications do you have with patients who donated their tissues, if any?
    - Future discoveries and therapeutic advances
    - Results of research with their samples – general results of research versus individual patient results
    - Patient education
    - Contributions to patient care
3. Do you release any information back to the donors of the tissues?

Appendix E

National Cancer Institute and National Dialogue on Cancer Research Team Questionnaire to Assess Biospecimen Needs and Response Frequencies

This survey was conducted at the July 2003 American Association for Cancer Research (AACR) international conference in Washington, D.C. to help NCI and NDC better understand the needs of academic, government, and industrial researchers for human biospecimens, and results were provided to the National Biospecimen Network Design Team in their deliberations. The survey took respondents about 5-10 minutes to complete and could be returned to the Survey Collection Box at the NCI or Cooperative Human Tissue Network (CHTN) exhibit booths. Respondents were informed that individual responses would be kept confidential and would not be disclosed to anyone but the individuals conducting this survey, except as otherwise required by law. Participation was completely voluntary, and participants were told that they could skip any question they did not wish to answer. The response frequencies and percentages are shown below by question.

Question	Answers	Total	Percent
Question 1 - Does your research involve the use of human biologic specimens, or are you planning to use such specimens?	Yes, my research *currently* involves the use of human biologic specimens	188	75.2
	Yes, I plan on using human biologic specimens in *future* research	51	20.4
	No, my research does *not* involve the use of human biologic specimens (If no, the survey is completed, thank you for your time!)	11	4.4
Question 2 - What type of research are you doing? (check all that apply)	Basic and developmental research	175	69.7
	Translational research	87	34.7
	Clinical research	57	22.7
	Other (please specify)	7	2.8

Question	Answers	Total	Percent
Question 3 - What area of research are you working in? (check all that apply)	Cancer prevention	74	29.5
	Cancer epidemiology	34	13.5
	Cancer etiology	47	18.7
	Cancer diagnosis and detection	80	31.9
	Cancer immunology	35	13.9
	Cancer treatment	106	42.2
	Other (please specify)	11	4.4
Question 4 - Which of the following cancers does your research include? (check all that apply)	Bladder	20	8
	Breast	126	50.6
	Cervix	26	10.4
	Colorectal	58	23.3
	Endometrial	17	6.8
	Esophagus	19	7.6
	Head and Neck	36	14.5
	Kidney	18	7.2
	Lung	49	19.7
	Lymphoma	34	13.7
	Ovary	43	17.3
	Pancreas	24	9.6
	Prostate	79	31.7
	Stomach	21	8.4
	Other (please specify)	45	18.1
Question 5 - Which of the following best describes your research position?	Principal Investigator	107	44.2
	Postdoctoral fellow/Ph.D. student	100	41.3
	Technician	13	5.4
	Other (please specify)	22	9.1

Question	Answers	Total	Percent
Question 6 - Is your research funded by NIH?	Yes	117	47.6
	No	129	52.4
Question 7 - Is your research institution located in the United States?	Yes	180	73.5
	No (If no, where is your institution located?)	64	26.1
Question 8 - At what type of institution are you located?	Academic	193	79.1
	Government	19	7.8
	Industrial	32	13.1
Question 8A - If industrial, which of the following best describes your industrial research institution?	Start-up or smaller biologically related institution	8	32
	Medium biologically related institution	9	36
	Large biologically related institution	8	32
Question 9 - Is there a hospital associated with your research institution?	Yes	183	75.3
	No	60	24.7
Question 9A - If yes, approximately how many beds are in the hospital?	Fewer than 200 beds	8	5.8
	201-500 beds	47	33.8
	More than 500 beds	84	60.4
Question 10 - What sources do you use to obtain specimens? (check all that apply)	Your institution/hospital	167	69
	NCI-supported resources	35	14.5
	Commercial sources	43	17.8
	Other researchers/collaborators	81	33.5
	Other (please specify)	9	3.7

Question	Answers	Total	Percent
Question 11 - Which of the following types of tissue do you typically use or do you plan to use? (check all that apply)	Frozen tumor (In OCT? _____)	157	66.8
	Frozen uninvolved or normal (In OCT? _____)	58	24.7
	Paraffin blocks/slides	130	55.3
	Serum	85	36.2
	Plasma	54	23
	Buffy coat	30	12.8
	Urine	17	7.2
	Saliva	10	4.3
	Other (please specify)	15	6.4
Question 12 - How many patient specimens do you typically require per year?	0 to 25 patient specimens	61	27
	26 to 100 patient specimens	90	39.8
	101 to 500 patient specimens	54	23.9
	501 to 1,000 patient specimens	11	4.9
	Greater than 1,000 patient specimens	10	4.4
Question 13 - What volume of tissue do you typically require per specimen?	Less than 0.1 g	50	22.8
	0.1-1.0 g	120	54.8
	1.1-5.0 g	42	19.2
	More than 5.0 g	7	3.2
Question 14 - Does the tissue you obtain currently meet your needs for:			
Number of specimens?	Yes	123	58.9
	No	78	37.3
Volume of specimens?	Yes	116	55.5
	No	61	29.2

Question	Answers	Total	Percent
Quality of specimens?	Yes	135	64.6
	No	44	21.1
Question 15 - If a tissue bank outside of your institution provided the following derivatives and/or services, which would you use? (check all that apply)	DNA	133	61.9
	RNA	121	56.3
	Protein	107	49.8
	Tissue microarrays	101	47
	Other (please specify)	6	2.8
For the remaining questions, "*associated patient data*" refers to demographic, clinical, and outcomes data related to the specimen. Examples of this type of data are listed as responses in Question 17.			
Question 16 - How often do you require extensive *associated patient data* (i.e., demographic, clinical, outcomes, etc.) related to the specimen?	Always	102	46.2
	Sometimes	93	42.1
	Never (SKIP TO QUESTION 18)	26	11.8
Question 17 - What *associated patient data* do you use or plan to use in your research? (check all that apply)	Demographic	71	35
	Family history	83	40.9
	Histopathologic (i.e., diagnosis, stage, grade)	130	64
	Time to progression	90	44.3
	Treatment and response	137	67.5
	Recurrence	93	45.8
	Survival	118	58.1
	Other (please specify)	1	0.5
Question 18 - What is the maximum you would pay for well-characterized biospecimen samples with definitive *associated patient data*?	$20 to $100 per specimen	174	85.3
	$101 to $300 per specimen	24	11.8
	Greater than $300 per specimen	6	2.9

Question	Answers	Total	Percent
Question 19 - What research data would you like to have provided along with specimens? (check all that apply)	Gene expression by DNA microarray (mRNA transcript expression profiles of approx. 30,000 genes)	148	70.1
	Marker expression by fluorescence *in situ* hybridization	56	26.5
	Marker expression by immunohistochemistry	101	47.9
	Comprehensive protein expression data	82	38.9
	Other (please specify)	4	1.9
Question 20 - What is the maximum you would pay for well-characterized biospecimen samples with definitive *associated patient data* **and** standardized gene expression by DNA microarray data?	$100 to $500 per specimen	144	69.6
	$501 to $1000 per specimen	32	15.5
	Greater than $1000 per specimen	1	0.5
	Don't need	30	14.5
Total Surveys Returned: 254			

Appendix F

National Biospecimen Network Blueprint
Design Team and Advisors Meeting Participant List

July 28–29, 2003

Hyatt Regency Bethesda • One Bethesda Metro Center • Bethesda, MD

Anna D. Barker, Ph.D., National Cancer Institute and NDC Research Team Chair
Jeffrey Trent, Ph.D., Translational Genomics Research Institute (TGen)
and TAWG Co-Director
Paula Kim, Pancreatic Cancer Action Network and TAWG Co-Director

Roger Aamodt, Ph.D., National Cancer Institute
Michael E. Berens, Ph.D., The International Genomics Consortium
Kerry L. Blanchard, M.D., Ph.D., Eli Lilly and Company
Alan Buckler, Ph.D., Ardais Corporation
Kenneth H. Buetow, Ph.D., National Cancer Institute
Robert L. Cohen, M.D., Genentech
Carolyn Compton, M.D., Ph.D., McGill University
Peter Covitz, Ph.D., National Cancer Institute
Dan Crichton, National Aeronautics and Space Administration
Mary E. Edgerton, M.D., Ph.D., Vanderbilt University Medical Center
Elisa Eiseman, Ph.D., RAND Corporation
William Grizzle, M.D., Ph.D., University of Alabama at Birmingham
Rina Hakimian, J.D., M.P.H., Association of American Medical Colleges
Stanley R. Hamilton, M.D., University of Texas M.D. Anderson Cancer Center
Stephen Hewitt, M.D., Ph.D., National Cancer Institute
Arthur L. Holden, M.B.A., Ph.D., First Genetic Trust – SNP Consortium
Valerie Hurt, J.D., National Institutes of Health
Julie Kaneshiro, DHHS Office for Human Research Protections
Kirstine Knox, Ph.D., National Translational Cancer Research Network
Thane Kreiner, Ph.D., Affymetrix, Inc.
Annette Levey, J.D., National Institutes of Health
Amy McGuire, Foundation for the National Institutes of Health
John D. Potter, M.D., Ph.D., Fred Hutchinson Cancer Research Center
Wendy R. Sanhai, Ph.D., Foundation for the National Institutes of Health
Howard Schulman, Ph.D., SurroMed, Inc.
Mary Lou Smith, J.D., Research Advocacy Network
Janet Warrington, Ph.D., Affymetrix, Inc.
Victor Weedn, M.D., J.D., Carnegie Mellon University

Appendix G

Examples of Potential Needs and Uses for Biospecimens by Constituency

User Type	Needs	Uses
Academic Researchers	20-100 samples or more per experiment < 100 mg to 100-200mg Standard clinical quality RNA grade Protein grade	Biology (genetic basis of human disease) Aberrations of signal transduction Genomic analysis, i.e., mutation screening, loss-of-heterzygosity studies, etc.
Pharmaceutical Companies	20-100 samples per experiment Multiple experiments (~ 1,000 samples per year) < 100 mg to 100-200 mg Standard clinical quality RNA grade Protein grade	Drug and diagnostics development Target credentialing and validation
Biotechnology Companies	20-100 samples per experiment Multiple experiments (~ 1,000 samples per year) < 100 mg to 100-200 mg Standard clinical quality RNA grade Protein grade	Biomarker discovery and validation Target identification

Appendix H

Guidance on Informed Consent for Tissue Repositories

Office for Human Research Protections (OHRP)

In an August 1996 memo included as part of a 1997 guidance document, "Issues to Consider in the Research Use of Stored Data or Tissues," the Acting Director of the Office for Human Research Protections (OHRP; then Office for Protection from Research Risks, or OPRR) delineated the following points to be considered in developing informed consent as part of the operation of human cell repositories:

- Written informed consent should be obtained from each donor-subject in accordance with HHS regulations 45 CFR 46.116, (*ohrp.osophs.dhhs.gov/humansubjects/guidance/45cfr46.htm#46.116)*. Included among the basic elements of informed consent should be a clear description of (1) the operation of the cell repository; (2) the specific types of research to be conducted; (3) the conditions under which data and specimens will be released to recipient-investigators; and (4) procedures for protecting the privacy of subjects and maintaining the confidentiality of data.
- Informed consent information describing the nature and purposes of the research should be as specific as possible.
- Where human genetic research is anticipated, informed consent information should include information about the consequences of DNA typing (e.g., regarding possible paternity determinations).
- Informed consent documents may not include any exculpatory language through which subjects are made to waive or appear to waive any legal rights.
- OPRR recommends that the cell repository develop a sample collection protocol and informed consent document for distribution to collector-investigators and their local institutional review boards (IRBs).
- A written submittal agreement for collector-investigators should require written informed consent of the donor-subjects utilizing an informed consent document approved by the local IRB. It should also contain an acknowledgment that collector-investigators are prohibited from providing recipient-investigators with access to the identities of donor-subjects or to information through which the identities of donor-subjects may readily be ascertained.

National Bioethics Advisory Commission (NBAC)

In 1999, following careful consideration of the ethical issues involved, the NBAC made a series of recommendations for the use of human biological materials in research. In their recommendations for informed consent in the collection, storage, and use of human biological materials, they noted that consent forms should be developed to provide potential donors with a

sufficient number of consent options from which to choose, to help them understand clearly the nature of the decision they are about to make. They suggested that such options might include, for example:

- Refusing the use of their biological materials in research
- Permitting only unidentified or unlinked use of their biological materials in research
- Permitting coded or identified use of their biological materials for one particular study only, with no further contact permitted to ask for permission to do further studies
- Permitting coded or identified use of their biological materials for one particular study only, with further contact permitted to ask for permission to do further studies
- Permitting coded or identified use of their biological materials for any study relating to the condition for which the sample was collected originally, with further contact allowed to seek permission for other types of studies
- Permitting coded use of their biological materials for any kind of future study

The NBAC pointed out that the last option, "permitting coded use of biological materials for any kind of future study," is problematic because it is impossible to foresee what kind of research the biological materials may be needed for in the future. This is particularly the case when a patient has identified a certain kind of study as objectionable.

The primary concern of the NBAC is that informed consent is truly informed. In cases in which potential donors cannot be given a complete picture of what might happen to their tissues, the consent given for tissue use is not considered to be truly informed.

Appendix I

The National Cancer Institute Tiered Consent Form

The National Cancer Institute (NCI), in cooperation with the National Action Plan for Breast Cancer, has developed a model informed consent document for future unspecified research use of specimens collected during routine medical care. The NCI databases currently use this consent form and an accompanying educational brochure for patients. A great deal of thought and testing went into formulating the form and accompanying brochure to ensure their utility and effectiveness.

The form contains three statements to which patients must respond affirmatively:

1. My tissue may be kept for use in research to learn about, prevent, or treat cancer.
2. My tissue may be kept for use in research to learn about, prevent, or treat other health problems (i.e., diabetes, Alzheimer's disease, or heart disease).
3. Someone from XYZ may contact me in the future to ask me to take part in more research.

Deleting the third statement would seem to realize the broadest type of consent. However, most institutional review boards (IRBs) require that permission be obtained to recontact a research subject. The third statement effectively protects researchers and presumably allows them more options, given changing guidelines and policies. Tiered consent, although presenting some complexities for the private sector, provides the most flexibility for patients. It is important to note that there are many examples of the successful use of the tiered consent process with IRB approval.

Appendix J

Examples of Ethics Oversight Groups

Genomics Collaborative, Inc., Bioethics Advisory Board

Genomics Collaborative, Inc. (GCI) is a commercial venture that provides consented DNA, RNA sera, and snap-frozen tissues linked to clinical information from donors worldwide in its GCI global repository. GCI empanelled a bioethics advisory board at the outset to develop solutions to the ethical issues inherent in creating a repository of this nature. The board played a significant role in creating a consent process for the repository, developing a proprietary system for making collections anonymous while still collecting longitudinal clinical data without compromising patient confidentiality.

Department of Veterans Affairs Cooperative Study DNA Bank Ethics Oversight Committee

The Department of Veterans Affairs (VA), through the Cooperative Studies Program, supports a genetic tissue bank (Study 478, "Genetic Tissue Banking in VA Clinical Research") for the collection of samples in ongoing clinical trials and the provision of samples for future study. The bank is overseen by three appointed committees that guide the ethical use of materials and provide scientific review of policies and procedures. The Ethics Oversight Committee (EOC), the Scientific Advisory Committee, and the Veterans Advisory Committee are both appointed by the VA Research and Development Office and serve overlapping terms. The EOC is made up of experts in the legal and ethical implications of genetic research, and experts in relevant scientific disciplines.

Ardais Bioethics Advisory Board

Ardais Corporation, a commercial clinical genomics company that maintains a sizable repository of annotated tissue samples, created a bioethics advisory board to oversee its human subjects activities. The board, comprised of legal, ethical, and technical experts, helps set policies for the ethical distribution of tissues but is not involved in the day-to-day operational aspects of decisionmaking or approval. The board also supports the evaluation of new opportunities for providing new clinical genomics products and services to the research community, each of which comes with unique sets of bioethical issues.

Appendix K

Best Practices for Repositories
International Society for Biological and Environmental Repositories

These Best Practices are reviewed periodically and revised to incorporate improved application and research findings that would affect repository work. The reader is referred to the International Society for Biological and Environmental Repositories (ISBER) Web site (*www.isber.org*) for the most current guidelines.

INTRODUCTION

The desire to preserve biological and environmental specimens for research purposes and to ensure species biodiversity requires the development of methods for long-term storage that will enable their effective future use. Sharing successful strategies for accomplishing this goal was one of the early driving forces for ISBER. In addition, ISBER fosters education and research and promotes quality and safety in all activities relating to specimen collection, storage, and dissemination.

ISBER's Best Practices for Repositories (Best Practices) reflect the collective experience of its members to provide repository professionals with a comprehensive foundation for the guidance of repository activities. These practices reflect input from individuals within and outside of ad hoc committees. Best Practices will be reviewed periodically and will be revised to reflect advances in research and technology. All revisions are subject to approval by the ISBER Board of Directors. These practices reflect the most effective approaches to the establishment and running of specimen collection facilities and are not intended as required practices.

TABLE OF CONTENTS

SECTION A: GENERAL INFORMATION
A1.000 ADHERENCE TO ISBER BEST PRACTICES
A2.000 DEFINITION OF TERMS

SECTION B: GENERAL ORGANIZATIONAL REQUIREMENTS OF A REPOSITORY
B1.000 GENERAL INSTITUTIONAL REQUIREMENTS
B1.100 Institutional Identity and Affiliations
B1.200 Contracted Laboratory Services
B2.000 FUNCTIONAL COMPONENTS OF A REPOSITORY
B2.100 Director
B2.200 Technical Staff

SECTION C: RECORDS MANAGEMENT
C1.000 GENERAL
C1.100 Availability for Inspection
C1.200 Record Retention
C1.300 Security
C1.400 Archival System
C1.500 Corrections and/or Changes

SECTION D: FACILITIES
D1.000 GENERAL
D2.000 HEATING, VENTILATION, AND AIR CONDITIONING
D2.100 Temperature
D2.200 Air Flow and Circulation
D3.000 LIGHTING
D3.100 General Lighting
D3.200 Task Lighting
D4.000 SECURITY SYSTEMS
D4.100 General
D4.200 Fire
D4.300 Access
D5.000 BACKUP POWER
D5.100 Uninterruptable Power Supplies
D5.200 Generators

SECTION E: OPERATIONS
E1.000 STANDARD OPERATING PROCEDURES (SOPs) MANUAL
E1.100 Purpose and Design
E1.200 Contents
E1.300 Implementation
E1.400 Modifications
E1.500 Review
E1.600 Staff Access and Review

E2.000 FREEZER AND REFRIGERATOR MONITORING
E3.000 INVENTORY VERIFICATION
E3.000 EMERGENCY PREPAREDNESS
E3.000 STORAGE ENVIRONMENTS
- E3.100 Backup Storage Capacity
- E3.200 Cryogenic Freezers
- E3.300 Mechanical Freezers
- E3.400 Refrigerators
- E3.500 Safety Features for Walk-in Freezers and Refrigerators
- E3.600 General Use of Dry Ice

E4.000 MAINTENANCE
- E4.100 General
- E4.200 Calibration
- E4.300 Validation

SECTION F: QUALITY ASSURANCE AND QUALITY CONTROL

F1.000 DEFINITIONS
- F1.100 Quality Assurance
- F1.200 Quality Control

F2.000 QUALITY MANAGEMENT PLAN
- F2.100 Staff Responsibilities
- F2.200 Quality Standards
- F2.300 Audits

SECTION G: SAFETY

G1.000 GENERAL
G2.000 REGULATIONS
G3.000 CONSIDERATIONS
G4.000 SAFETY INFRASTRUCTURE
G5.000 TRAINING
G6.000 SPECIFIC AREAS OF CONCERN
- G6.100 Biological Safety
- G6.200 Chemical Safety
- G6.300 Electrical Safety
- G6.400 Fire Safety
- G6.500 Physical Safety
- G6.600 Radiologic Safety

G7.000 KEY STEPS IN DEVELOPING A BIOSAFETY PROGRAM
G8.000 MANDATED ELEMENTS OF A CHEMICAL HYGIENE PLAN
- G8.100 Safety and Health SOPs

SECTION H: TRAINING

H1.000 GENERAL
H2.000 TRAINING INFRASTRUCTURE
- H2.100 Training Program

H2.200 Administrator Responsible for Training
H2.300 Trainers
H2.400 Training Documentation
H2.600 Cross-training
H2.700 Training Records

SECTION I: SPECIMEN TRACKING

I1.000 GENERAL
I2.000 LABELS
I2.100 Barcoding
I2.200 Labels for Human Specimens
I2.300 Labels for Animal Specimens, Specimens from *in Vitro* Models of Human Diseases, and Environmental Specimens
I3.000 INVENTORY SYSTEMS
I3.100 Specimen Location
I3.200 Other Specimen Descriptors
I3.300 Additional Information for Human Tissues
I3.400 Additional Information for Environmental Specimens
I3.500 Validation
I4.000 SHIPPING LOG

SECTION J: PACKAGING AND SHIPPING

J1.000 GENERAL
J2.000 TRANSPORT SPECIFICATIONS
J2.100 Regulatory Requirements
J2.200 Temperature Requirements
J2.300 Humidity Requirements
J2.400 Light Sensitivity Requirements
J2.500 Arrival Time Requirements
J2.600 Sample Quantities
J3.000 VALIDATION OF SHIPPING CONDITIONS
J3.100 Review of Packaging Test Report
J3.200 Validation of Packaging
J3.300 Test Shipments
J3.400 International Shipments
J4.000 TRACKING SHIPMENTS DURING TRANSPORT
J4.100 General
J4.200 Notification of Shipment
J4.300 Shipping Manifest
J4.400 Confirmation of Receipt

SECTION K: SPECIMEN COLLECTION, PROCESSING, AND RETRIEVAL

K1.000 GENERAL
- K1.100 Pilot Studies and Proof of Performance Studies

K2.000 SPECIMEN TYPES
- K2.100 Human Specimens

K3.000 COLLECTION PROCEDURES
- K3.100 Special Considerations for All Specimen Collection Protocols
- K3.200 Blood
- K3.300 Urine
- K3.500 Tissues
- K3.600 Nail Clippings
- K3.700 Saliva
- K3.800 Breast Milk

K5.000 ALIQUOTING FROZEN SPECIMENS
- K5.100 Plasma, Serum, and Urine

K6.000 SPECIMEN RETRIEVAL
- K6.100 Locating Specimens in Storage
- K6.200 Specimen Retrieval
- K6.300 Documentation of Retrieval
- K6.400 Special Safety Considerations

SECTION L: HUMAN SUBJECTS

L1.000 BACKGROUND

L2.000 APPLICABLE REGULATIONS

L3.000 UNRESOLVED ISSUES
- L3.100 Need for Harmonization

APPENDIX A: SPECIMEN COLLECTION AND PROCESSING PROTOCOLS AND REFERENCES

A. COLLECTING BUCCAL CELLS

B. PROCESSING BUCCAL CELLS

C. PREPARATION OF SERUM AND CLOTS FROM RED TOP VACUTAINERS

D. CRYOPRESERVATION OF WHOLE BLOOD

E. COLLECTION OF FECAL MATERIAL

F. PREPARATION OF PLASMA, BUFFY COAT, RED BLOOD CELL/GRANULOCYTE FRACTIONS

G. ESTABLISHMENT OF CELL LINES

H. SKIN BIOPSY COLLECTION FOR THE ESTABLISHMENT OF FIBROBLAST CELL LINES

I. PROCESSING URINE SAMPLES

APPENDIX B: INTERNET RESOURCES ON SAFETY

APPENDIX C: REFERENCES

Appendix L

Vignette: Life Cycle of a Biospecimen Through the National Biospecimen Network

The salient considerations—from tissue donation consent by the patient, through researcher access to specimens and data, and finally to efficacious treatment products—are described in this vignette.

The year is 2008. You have just been diagnosed with lung cancer.

Your oncologist sends you to your well-respected and familiar community hospital for a biopsy. During the consent process, after the biopsy procedure is thoroughly explained, you are asked if you will donate tissue obtained during the biopsy that is not used for your diagnosis to the National Biospecimen Network (NBN). You vaguely remember hearing public service announcements describing this resource, as well as a pamphlet about it from the National Cancer Institute "cancer cope kit" that your oncologist reviewed with you, and you ask for more information. A hospital staff member, whose ID badge identifies him as a trained consent counselor employed by the NBN, explains that cancerous lung tissue is on a list of biological samples needed for research. You learn that many doctors and hospitals across the nation have agreed to collect needed tissues from cancer patients to donate to the NBN. Research with tissue can help researchers better understand what causes cancer, how to prevent it, and how to treat it.

If you agree to donate your tissue, only leftover tissue from the biopsy will be saved for the tissue resource, and whether or not you donate the tissue will not have any impact on your treatment. The counselor explains a detailed consent form to you. You are concerned that your name, social security number, and other identifying information will travel with the tissue, to be recorded in a nationwide database. Who will have access to this information? The counselor assures you that the database is designed to protect the privacy of your medical information. In fact, you will not be recontacted concerning your medical condition unless you choose this option. After the emotional shock of the cancer diagnosis, you feel empowered that in some way you might be able to help contribute to a cure. You agree to donate your tissue.

It is the day of the surgery. You are asleep, under the influence of anesthesia, but the operating room staff has the situation well in hand. The NBN surgical liaison, knowing that you have consented to donate tissue and your preliminary diagnosis, arrives in the operating room well ahead of schedule to confirm that all standard operating protocols for the tissue collection are in place. After the tissue has been collected using defined time and condition requirements, a pathologist who has collaborated with the resource since its inception removes sufficient tissue to obtain your diagnosis and assesses a section of the remainder for the tissue bank using internationally accepted terminology, staging and grading criteria, and classification. After quality controls have been completed and documentation has been entered into the resource

database according to strict protocols, the sample is transported to a nonprofit regional tissue processing center located near an academic medical center about 50 miles away.

A researcher several states away is conducting a study to determine whether there are measurable differences between gene expression samples in normal and lung cancer tissues. His study has been approved by the NBN peer review process and accorded high priority for tissue distribution. The researcher wants to make use of a new proteomic assay for his study, which the NBN Oversight Committee has recently approved for addition to its database system. He turns to his computer, accesses the NBN Web site, and determines that, with the addition of your sample, he now has a sufficient number of tissue samples to proceed with the study. He knows that he will receive, in addition to a standard protocol of preliminary testing results performed by the tissue processing center, your age, sex, race, diagnosis, and possibly treatment and family history. He will not know your name or any other information that might affect your privacy.

It is 4 years later. You have survived your bout with cancer, and you now volunteer as an advocate for cancer research. You moved to another state 6 months ago and were able to access your file on record with the NBN and update your address and treatment outcome. The scientist who used your tissue for his research has submitted his raw data into the database and it is easily accessible by other researchers. From your work with the advocacy community, you know that this expanding body of information on lung cancer has produced several opportunities for early detection, therapy, and prevention. In fact, several pharmaceutical companies have produced new products that are in the pipeline or have been fast-tracked by the Food and Drug Administration and are already benefiting patients. You are glad that you have made a contribution to this process. You are even happier that when you talk to newly diagnosed lung cancer patients, you can tell them that there is hope.

Appendix M

Advanced Analysis Techniques[1]

DNA

DNA analysis requires relatively little tissue (polymerase chain reaction [PCR] amplification techniques require only a few nanograms). A milligram of tissue left over from diagnostic procedures would be sufficient to enable almost infinite amplification and distribution. Since DNA sequences are 99.9-percent identical, it is possible to consider assaying for just differences, such as single nucleotide polymorphisms (SNPs). There are a variety of technologies that enable high-throughput serial analysis of SNPs, or even parallel analysis of a limited number of SNPs. The challenge with scaling these technologies to whole-genome SNP analysis is their requirement for specific primers for each SNP, a costly and complex proposal. It is possible to employ a generic approach for reducing the complexity of the whole genome and to perform a parallel analysis of 10,000 SNPs on a DNA microarray. This level of genetic information could be useful for analyses like loss-of-heterozygosity studies in cancer. As the DNA resource is close to inexhaustible, tissue still would be available when more powerful whole-genome SNP analysis is developed, likely in the near future.

Protein

Other than using two-dimensional gels, there is no way to do whole-proteome analysis. Well-justified specimen use for specific proteomic studies makes the most sense today. In the future, technologies may allow researchers to look at transcriptomes (noncoding regions that are transcribed), which may provide new targets for therapeutic intervention. Therefore, it would be wise to preserve specimens for proteomic analysis when more procedures become available.

RNA

Reverse transcription-PCR and Northern blot analyses are useful for identification of a limited number of RNA markers but do not scale effectively for whole-genome analysis and should be reserved for hypothesis testing. Microarrays for whole-genome expression analysis are achievable and align with the goal of acquiring the maximum amount of information possible from the specimens. Whole-genome expression analysis using microarrays is now a standard approach in both the pharmaceutical and academic sectors. Microarrays can be used to determine if a tumor is benign or malignant, guide therapeutic choices, identify new classes of tumors, and predict patient outcome; they also have many other potential uses. With currently available technologies, 0.5 to 1 mg of tissue is needed to generate total RNA for whole-genome microarray analysis. It also is possible to recover labeled cRNA with certain technologies and save this tissue for future analysis.

[1] Source: *Technology Platforms* presentation by Dr. Thane Kreiner, Affymetrix, Inc., NDC Forum II, March 6, 2003.

Appendix N

Nonprofit, Government-Funded Example: The Cooperative Human Tissue Network*

The Cooperative Human Tissue Network (CHTN) provides biomedical researchers with access to human tissues. Six member institutions coordinate the collection and distribution of tissues across the United States and Canada in six regional divisions. The CHTN specializes in the prospective procurement, preservation, and distribution of human tissues for research. In addition to normal, benign, and malignant tissues, tissues from patients with specific diseases such as ulcerative colitis, a premalignant state, are provided. Trained personnel coordinate the retrieval, preservation, and delivery of specimens obtained from surgical resections and from autopsies.

Regional Divisions

The CHTN has six regional divisions:

- The Eastern Division is responsible for the area of the Northeast bounded by the western and southern borders of Pennsylvania, as well as Delaware, Alaska, and Hawaii.
- The Mid-Atlantic Division is responsible for Maryland, Virginia, and the District of Columbia.
- The Midwestern Division includes West Virginia and states west of Pennsylvania, north to Minnesota and south through Missouri, as well as Canada.
- The Southern Division encompasses Kentucky and all states south and west from the Carolinas to Texas.
- The Western Division covers all states north of Oklahoma and west of Texas.
- The Pediatric Division provides childhood tumors and diseased and normal tissues nationwide.

Tissue Processing and Storage

The processing of specimens varies according to the protocol of each individual investigator. Specimens can be snap-frozen and stored for limited periods of time at ultralow temperatures, according to the investigator's needs. Alternatively, investigators may require fresh tissues that can be collected using an aseptic technique, if necessary. Fresh tissues can be minced, placed in transfer media, and shipped to investigators for next-day arrival. Similarly, for research involving substances stable after fixation, fixed or paraffin-processed tissues can be provided. In addition, some divisions offer limited histological processing by special arrangement. This includes hematoxylin-and-eosin-stained sections and stained or unstained frozen sections.

* Information obtained from the CHTN Web site.

The institutions and Principal Investigators comprising the CHTN make every effort to ensure that there is an equitable distribution of tissues into and by the CHTN. In addition, several factors act to prevent CHTN tissue resources from benefiting only a limited number of investigators. Foremost, except for rare tumors, the number of surgical specimens is usually adequate for the recovery of research material.

Investigator Costs To Use Resources

Investigators pay a nominal processing fee for specimens received from the CHTN: $20 per sample for researchers at academic institutions and $60 per sample for researchers at nonacademic institutions. Slides and blocks accompanying tissue samples may be available at costs ranging from $4.50 to $8.50. The charges are not for the tissues, but to partially offset the costs of collecting, handling, and preparing the specimens in accordance with the detailed requirements of the investigators. Investigators must also pay the cost of shipping specimens to their laboratory.

Access Requirements and Priority

Access to the CHTN is provided to any investigator who signs the agreements regarding biohazards and commercial use and who provides a summary of the project for which the tissues are requested. Patient identity or other identifying information cannot be provided to investigators. This ensures complete confidentiality regarding medical information of patients. In addition, a copy of the approval of the research, which is obtained from the investigator's local institutional review board (IRB) for human use, is required for all projects.

Canada is the only foreign country for which tissues are provided to investigators (via the Midwestern Division). Requests from other countries will not be filled.

The CHTN seeks to provide tissues to the widest group of investigators practicable, and tissues are not restricted to investigators with National Institutes of Health (NIH) grants. In this regard, no investigator should expect to have access to unlimited numbers of malignant specimens of one type. If a large number of specimens and/or extensive patient data are required, it may be necessary to contact the National Cancer Institute (NCI) Tissue Expediter, consult the NCI Specimen Resource Locator, or identify local sources for the specimens and data. The CHTN attempts to provide each investigator with as many specimens as is equitable. Investigator access to increased numbers of specimens is improved by requesting as small a specimen as possible. For example, investigators requesting a minimum of 10 grams of one tumor would have access to only a very limited number of specimens, while investigators requesting a minimum of 0.5 grams of tumor would have access to many more samples. Access to specimens varies according to the surgical schedules and autopsy rates; thus, access is not predictable. Therefore, investigators requesting fresh specimens should make every effort to use these specimens when they are made available. Investigators may be charged a processing fee for specimens collected according to their protocol that are subsequently refused.

Coordinating Committee

The experience of personnel from all divisions of the CHTN is utilized in a coordinating committee that formulates policies for the operation of the CHTN. Voting members of this committee include the Principal Investigator and an additional member from each of the six divisions. In addition, NCI has one voting member. The coordinating committee meets periodically to assess the operation of the CHTN and to change or modify operating policies. A chairman and secretary of the committee are elected yearly.

History of the Cooperative Human Tissue Network

The CHTN was formed in 1987 by three organizations, along with a subcontract with the Children's Cancer Group. In 1991, the Pediatric Division became an independent group, and an additional adult collection center was added to the CHTN. The following six organizations, which currently comprise the CHTN, have extensive experience in providing human tissues for research:

1. **Ohio State University** has had an internal tissue procurement service, sponsored by the Ohio State University Comprehensive Cancer Center, since 1975. In 1987, it joined the CHTN. Daniel Sedmak, M.D., is the Principal Investigator for this Midwestern Division of the CHTN.

2. **The Pediatric Division**, located at Children's Hospital of Columbus, under the direction of Stephen J. Qualman, M.D., obtains most of its specimens through Children's Oncology Group institutions.

3. **The University of Alabama at Birmingham**'s Tissue Procurement Facility has supplied fresh and snap-frozen human tissues to investigators within and outside of their Comprehensive Cancer Center since 1978. The University of Alabama at Birmingham joined the CHTN in 1987. William E. Grizzle, M.D., Ph.D., the Principal Investigator for this Southern Division of the CHTN, is a board-certified clinical and anatomic pathologist as well as a funded medical researcher.

4. **The University of Pennsylvania** serves as the Eastern Division of the CHTN, under the direction of Principal Investigator, Dr. Virginia LiVolsi, Professor of Pathology and Laboratory Medicine, who has served as the Eastern Division Principal Investigator since the inception of the CHTN. During that time, the Eastern Division has expanded its tissue procurement network to include numerous hospitals in the Eastern region.

5. **The University of Virginia** has had a tissue procurement service since 1993. It served NCI as a major tissue supplier for the Cancer Genome Anatomy Project. Dr. Christopher Moskaluk, Assistant Professor of Pathology, serves as the Principal Investigator of the Mid-Atlantic Division.

6. **Vanderbilt University** had a Tissue Procurement Facility that operated as a subcontract tissue procurement site for the CHTN for 6 years before becoming the Western Division of the CHTN in 2001. Dr. Mary Kay Washington, Professor of Pathology, is currently the Principal Investigator of the Western Division.

Specimen Types

The CHTN obtains tissues from routine surgical resections and autopsies. In addition, remnant body fluids (e.g., serum, leukapheresis products, pleural effusions) are available on some patients, and limited histological services (e.g., frozen sections) can be obtained by special arrangement. Normal tissues (i.e., microscopically uninvolved by disease) commonly available from surgical resections include skeletal muscle, skin, bone, cartilage, connective tissue, breast, adipose tissue, small and large intestine, arteries, veins, thyroid, lung, gynecological tissues, kidney, spleen, tonsils, testes, salivary glands, and oral mucosa. Various pathological tissues also are obtained from surgical resections, including head and neck malignancies and carcinomas of the lung, gastrointestinal tract, breast, genitourinary system, prostate, kidney, and skin. In addition to the above more common specimens that are provided to investigators from surgical resections on a regional basis, access to rare tumors, e.g., neuroendocrine tumors and sarcomas, is provided from all divisions via networking. Pediatric tumors and diseased and normal tissues are available nationwide.

Additional normal tissues can be obtained from autopsies. These include liver, pancreas, adrenals, brain, pituitary, lymph nodes, heart, and thymus.

Quality Control

All specimens are subject to an immediate gross examination by a pathologist. The diagnosis then is verified through frozen section, touch preparations, or subsequent evaluation of permanent histopathology. In most cases, the diagnosis can be assured before the specimen is released to the investigators. However, in some cases in which fresh specimens are needed, the tissues are provided to the investigators with a preliminary diagnosis, and the investigators are informed of the final diagnosis as soon as this information is available. In all divisions and institutions affiliated with the CHTN, pathological diagnosis and patient care have priority over the use of any specimen in research.

CHTN Policy for IRB Approval

The CHTN requires researchers to (1) agree to follow the provisions of the "Common Rule" (45CFR46) Federal human subjects regulations and (2) obtain IRB approval before receiving specimens for their research. While these regulations currently do not apply to institutions that do not receive Federal support, the CHTN policy requires all researchers using CHTN tissues to follow the Common Rule. The CHTN does not provide patient identity or other identifiers to investigators. This ensures complete confidentiality regarding patients' medical information.

Commercial Use of CHTN Tissues

Tissues are provided by the CHTN only for research purposes. Before requests are accepted by the CHTN, investigators must sign a statement that they "shall not sell any portion of the tissues provided by the CHTN, or products directly extracted from these tissues (e.g., protein, mRNA, or DNA)." The recipients also must agree that they shall not transfer tissues (or any portion thereof) supplied by the CHTN to third parties without prior written permission from the CHTN. The intent of the CHTN is to encourage research using human tissues for the good of the public rather than for private gain. The commercial development of products from tissues could raise questions of ownership for patentability of products so developed. Disputes over ownership and royalties could disrupt significantly the ability of the CHTN to distribute tissues based on scientific merit and need. Legal experts at each division of the CHTN and experts at NIH continue to study this area of commercial use of research tissues.

Use by Third Parties

The CHTN application is intended for the use and processing of samples utilized by the laboratory and/or personnel that fall under the supervision of the Principal Investigator listed in the application. Any transfer of samples or aliquots to personnel or laboratories that are not under the supervision of the indicated Principal Investigator requires the following:

- An explanation of the need to transfer the materials and benefit to the investigator's research
- A copy of the enclosed CHTN agreement page signed by the collaborator
- A copy of the collaborator's IRB approval, unless the collaborator is covered under the IRB approval granted for the project proposed in the application.

The CHTN does not supply samples to tissue banks solely for distribution to third parties.

Appendix O

Nonprofit, Eventual Private/Public Funding Example: United Kingdom National Cancer Tissue Resource

Background

In 2000, the United Kingdom (UK) developed a National Cancer Plan with the primary aim of improving the quality of cancer care delivered by its National Health System (NHS) and positively impacting the country's cancer survival rates. One of the plan's most important initiatives was the establishment of two national cancer research networks, embedded in the NHS and distributed around the country. The first network, the National Cancer Research Network (NCRN), would be an infrastructure for large, multicenter clinical trials. The second network, the National Translational Research Cancer Network (NTRAC), would take the lead on translational research. These research networks are funded by the Department of Health (DH) and are under the strategic oversight of the National Cancer Research Institute (NCRI), whose key partners include DH, the Medical Research Council, and Cancer Research UK (a consortium of medical charities).

NTRAC, the translational research arm, hopes to improve the quality of cancer care by creating a national network of cancer research centers that integrates scientific and clinical expertise and ultimately shares knowledge and resources for the benefit of cancer organizations and patients. These aims will be achieved through building research infrastructure and workforce capability within the centers, both to support the development of novel anticancer therapeutics and diagnostics and to test their promise in early clinical trials. Currently, full NTRAC Network Center status and funding have been awarded through a competitive process to 10 centers of scientific and clinical excellence. The funding support may be used with complete flexibility by each of the centers to actualize fast-track research for cancer patients.

Like the United States, the UK found that limited access by academic research communities and industry to large-scale, high-quality biological samples linked to clinical information was a primary barrier to the maximization of recent advances in genomic, proteomic, and molecular research. Historically, there has been difficulty in obtaining funding in the UK through existing peer review mechanisms to support local sample acquisition. A shortage of pathologists and an increased demand for their services have negatively affected the acquisition and annotation of high-quality biological samples. Limited personnel and funding for sample collection and a lack of nationally standardized quality control (QC) mechanisms and standard operating procedures (SOPs) for the collection and storage of those samples have resulted in varied tissue collections, which are of limited potential use to emerging technologies. A lack of high-quality clinical outcome data associated with local biological sample collections also limits the ability to support research aimed at identifying novel prognostic markers. Recent misuse of human tissues, widely

publicized by the popular press, has created ethical challenges, necessitating the legislative design of an ethical framework for the acquisition and use of human tissue. Faced with these challenges, many researchers are reluctant to share the limited tissue resources they might have with the wider research community.

In order to overcome these substantial economic, logistical, and ethical barriers and to meet the current and future needs of the research communities, it was determined that a unified national approach to tissue collection was necessary. In January 2002, NCRI requested that NTRAC produce a strategic framework and blueprint for a National Cancer Tissue Resource (NCTR), including infrastructure requirements and resource implications, for the collection of high-quality biological tissue samples, well annotated with pathological and clinical data. It was necessary that the blueprint be built on principles of inclusion, accessibility, and flexibility and that it be established within a clear ethical framework that balances the interests of society and the rights of patients. Cathy Ratcliffe and Dr. Kirstine Knox of NTRAC assumed this responsibility and, following a series of workshops and consultations with experts (including from the U.S. NCI) as well as stakeholders throughout the country, produced a draft strategic plan in September 2002 for an NCTR in the UK.

Goals

The proposed strategy for the NCTR includes three key components:

1) The NCTR must maximize benefits from existing high-quality collections of biological samples linked to clinical outcome information, with the aim of realizing the potential of the retrospective NHS paraffin-embedded tissue archive. Paraffin samples are valuable at both the hypothesis validation and clinical stages of the research process. This aim is particularly focused around gathering specimens from patients already enrolled in key clinical trials and will involve central collection of samples, the production of tissue microarrays, distribution to researchers, and collation of research results. The inclusion of this component will allow some work to go forward quickly with little infusion of new resources.

2) The NCTR must provide infrastructure and workforce capability for prospective collection of population-based, high-quality samples linked to clinical information that would meet the needs of the academic research community and industry in a timely fashion. Frozen samples collected prospectively are the key to innovation at the hypothesis-generation stage of research. This requires dedicated personnel in surgical and pathology environments to collect, process, annotate, and store samples in a standardized, routine manner and to enter information into a clinical information system.

3) The NCTR must ensure the compatibility of approach with international partners (Europe, the United States, Canada, and Australia) to maximize benefits for cancer researchers and patients. Standardization of systems that communicate globally could benefit cancer research around the world. As a result of its work in the UK, NTRAC has been charged with developing the SOPs for the collection of biological samples in Europe (ETRAC).

Structural Overview

NCTR will take the form of a managed, distributed network. NTRAC will be responsible for the implementation and operation of the resource. A competitive application process will be used to select a network of tissue acquisition resource centers for prospective collection of frozen tissue. Key existing clinical trials will provide paraffin-archived samples, with the process coordinated through NCRN and the Clinical Trials Office. Both frozen and paraffin-embedded tissues will feed into a second network of tissue processing centers. It is possible that some tissue processing centers and some tissue acquisition centers will reside in the same locale. The processing centers will process raw tissue into bioproducts including RNA, DNA, and tissue microarrays.

A central bioinformatics hub will form the core of the NCTR. Initial plans are underway to use components developed with the UK e-Science program to develop a powerful, scalable, and secure infrastructure within 2 years. The hub will receive the results of the genomic and proteomic research in addition to standardized, site-specific histopathological data, and it will be responsible for tracking collection, storage, and analysis of deidentified samples. New procedures will be developed to access clinical and outcome information that currently is being collected by clinical trials. Additionally, clinical and outcome information from across the care pathway, including cancer registries, eventually will feed into the bioinformatics hub. Finally, the results of the analysis of samples will be integrated into the system. The ultimate aim is to build an information grid that automatically and seamlessly can incorporate all relevant data from each new patient into the appropriate database, input that patient's data into existing predictive models, and transmit the resulting information back to the patient's physician in the clinical environment. As envisioned, this bioinformatics platform also has the potential to serve as a model for other diseases and may facilitate international collaboration.

Access to both tissue and data will be available to the academic research community and to industry through an equitable and transparent review mechanism. Initially, the NCTR plans to collect malignant and normal tissue, together with blood, from the five most common cancers in the UK: Breast, colorectal, lung, ovarian, and prostate. This initial limitation of scope increased the ease of obtaining funding for the resource. Expansion to collecting other (especially rarer) forms of cancer will occur as soon as is reasonably possible.

Existing Systems

The UK NCTR will incorporate existing systems, and in fact, base their new system on the contributions of various tissue collection protocols in existing clinical trials programs. In the UK, participation in the NCTR is linked to payment for services (National Health). Additionally, standards can be mandated by the Royal College of Pathologists and can be a requirement to receive National Health compensation. However, it is critical to the success of the UK system that it be established in a way that correlates with and is complementary to related national tissue bank initiatives.

Governance/Organization

The NCTR governance/organizational structure will comprise five key component parts:

- A Coordinating Unit, located within the NTRAC Coordinating Centre, will implement and operate the NCTR.
- A linked network of tissue acquisition resource centers (TARCs) will be selected through a tendering process and will be contracted to adhere to standard operating protocols for prospective collection of biological samples and outcome clinical information.
- A linked network of tissue processing resource centers (PRCs) for production of DNA, RNA, and tissue microarrays also will be selected through a tendering process and contracted to adhere to standard operating protocols.
- Tissue samples will be collected from key established and future clinical trials and will be coordinated through clinical trials offices and the NCRN.
- Finally, a central information system and bioinformatics hub will link: (1) Tracking of collection, processing, distribution, and analysis of samples to (2) histopathological datasets in line with The Royal College of Pathologists recommendations to (3) clinical/outcome information to (4) results of research.

Oversight

Oversight of the strategic direction and management of and access to the NCTR will be implemented and operated by NTRAC. Because the governance of the NCTR exists within the ethical framework mandated by the legal structure for England and Wales, the system of governance for the NCTR is the responsibility of the DH. Locating the NCTR within NTRAC means that the NCTR Chief Operating Officer is accountable to the Director of NTRAC and, through established accountability lines, to the DH as well as to the NCRI Board. The NCRI Board will provide strategic oversight, possibly through establishment of an NCRI Management Committee made up of representatives of the contributing funding bodies. An NCTR Steering Committee comprised of independent experts from the UK and other countries, together with representatives from the contributing funding bodies, will oversee access to tissue/information. It has been suggested that this committee be chaired by an international expert, possibly Dr. Stephen Hewitt, Director of the Tissue Array Research and Production Laboratory, NCI, or Dr. Peter Riegman, head of the European Union Framework V initiative on establishing a European human frozen tumor bank.

Operational policies and procedures for the NCTR will ensure that:

- The NCTR operates within an ethical framework
- Standard contractual agreement that tissue acquisition and processing resource centers contribute to the activities of the NCTR
- Appropriate guidance and support are provided to all NCTR personnel through a training, education, and communications program as well as through ongoing advice

- NCTR equipment is adequately maintained
- A disaster management program is in place
- A succession planning program is in place

Standards and Quality Control

NCTR operational policies will include delineated SOPs for the TARCs. These SOPs will have guidelines for (1) obtaining explicit, written consent from well-informed candidate patients, using a standard national consent form and (2) collecting, handling, storing, and processing frozen tissues (including minimizing risk of tissue hypoxia and ensuring that patient diagnosis is never compromised by the sample harvesting process). Tissue manipulation SOPs will cover:

- Fixation, preparation, storage, and processing of paraffin-embedded tissue
- Collection, processing, and storage of blood samples
- Coding of samples and mapping storage locations at each TARC
- Obtaining complete histopathological datasets and systematic clinical and outcome information
- Distribution of NCTR samples to processing centers and/or directly to research communities
- Tissue sample processing for the production of bioproducts

A QC policy will be implemented by the NCTR Coordinating Centre to ensure optimal standards of:

- Tissue sample quality and associated histopathological description—pathological review of representative hematoxylin and eosin sections
- Frozen tissues and bioproduct quality—extract and assess RNA
- Paraffin-embedded samples—assess fixation quality
- Sample location mapping and tracking
- Interrogation/validation of the information technology (IT) tracking system

Sample Acquisition

A fully operational NCTR will incorporate up to six geographically distributed TARCs. National Health Service units will be selected and contracted as TARCs by a tender selection process (i.e., through Requests for Proposals [RFPs]). A pilot phase will involve establishment of two or three TARCs functioning under SOPs and linked by an NCTR information system. Each TARC will guarantee Royal College of Pathologists-recommended minimum dataset histopathological assessment of every sample collected for the NCTR. Quality assurance (QA) mechanisms will be enforced by the NCTR Coordinating Centre.

Tendering submissions will outline predicted sample accrual during Year 1 for breast, colorectal, lung, bladder, and ovarian tumor types. It is anticipated that each TARC will collect samples from up to 1,000 cases per year. Each TARC will be contractually obliged to fulfill sample and data collection criteria. The NCTR will implement a four-site pilot phase, which will allow assessment of sample collection rates across two or three TARCs and will inform the necessary scale of NCTR to meet the needs of the research community.

It is envisioned that the approach will be scaled up during full operation of the NCTR, to include the acquisition, storage, and processing of rarer cancers such as pancreatic, head and neck, bladder, melanoma, and renal cell carcinoma.

Sample Processing

One or more PRCs will be selected to provide tissue microarray facilities for clinical trial-associated paraffin-embedded sample collections. Up to four PRCs will be selected to provide tissue processing facilities for frozen and paraffin-embedded, population-based TARC collections. It is possible that some PRCs may be synonymous with the TARCs. PRCs will be selected by a tendering process and will be contracted to process samples stored at the TARCs, as well as those associated with key clinical trials, according to SOP and subject to QA assessment.

Funding

NCRI partners and other government departments initially will fund NCTR. In particular, the Department of Trade and Industry will provide funding for informatics support. Eventually, private investment (as public/private partnerships) will be instituted in a regulated manner.

Informed Consent

The DH currently is reviewing the UK law on the removal, retention, and use of human organs and tissues. NTRAC is working to convey the feelings and wishes of researchers to the department. A new bill is being written for introduction in Parliament in late 2003 and for enactment in 2004. The key principle of the legislation is valid consent in a standard paragraph that is nationally agreed upon. Patients would consent (opt-in) to the use of their tissue and associated data for research. Eventually, as the level of public awareness is raised through national educational campaigns, an opt-out system (consent would be assumed unless patients opt-out of donating their tissues) might be considered.

Access to Tissue and Data

The ultimate goal of the NCTR is to automatically incorporate all relevant data from each new patient into the appropriate database, input that patient's data into existing predictive models, and transmit that information to the clinician in the clinical environment to assist in treatment decision making.

Access to both tissue and data will be available to the academic research community and to industry through an equitable and transparent review mechanism. It is envisaged that academic research communities will have access to tissue on a cost-recovery basis only and to bioproducts at a subsidized cost. It also is envisaged that industry will have access to bioproducts only at full cost of processing and delivery. Arguments for allowing access to industry include the following:

- Industry is a key player in drug development—on its own and increasingly in partnership with the academic research community; patients and society benefit from drug development advances
- Access to the NCTR will provide incentives for industry investment in the UK and thereby will help to improve the competitiveness of the UK cancer research base
- NCRI is a partnership among government, charities, and industry

Expanded Services

The NCTR expects to collect 6,000 samples per year. Initial testing will include DNA; RNA, supported by laser-capture microdissection; cryostat sections; microarray preparations for high-throughput screening; and blood components. A future rollout will encompass a variety of testing options including:

- Longitudinal—the NCTR plans to maximize benefits from existing high-quality collections of biological samples linked to clinical outcome information using the retrospective NHS paraffin-embedded tissue archive (which may limit its usefulness). This aim is focused particularly on gathering specimens from patients who are already enrolled in key clinical trials, where procedures for collection of outcome information are in place.
- Digital image analysis—molecular imaging; image acquisition; image analysis and archiving; quantitation tissue/cell microarray; and expression profile analysis
- Genomics—custom cDNA/genomic library services; DNA isolation/purification/ characterization; RNA isolation/purification/characterization; recombinant DNA plasmid construction; probe synthesis/labeling; mutagenesis services; transfection services; DNA sequencing; custom oligonucleotide synthesis; peptide synthesis
- Proteomics—target identification (cDNA cloning technologies); target validation (i.e., cell cycle, apoptosis, tumor suppressors); expression profiling using microarray technologies; protein expression target validation services (i.e., immunocytochemistry interpretation in relation to underlying pathology)
- Telepathology and reference laboratory services—special, molecular-based classification of human cancer (database linked with above—i.e., localization of gene and gene products in a range of normal and diseased tissues by *in situ* hybridization and immunocytochemistry); molecular therapies during clinical trials; contract services (i.e., therapeutic monoclonal antibody [cross reactivity]), other Food and Drug Administration-regulated, therapeutic-related studies (i.e., Herceptin, etc.); and clinical trials (i.e., automated DNA ploidy clinical trial; Good Laboratory Practice-compliant kinetic imaging Komet system) designed for use in regulated drug approval studies.

Demonstration Project

The NTRAC Strategic Plan was met with widespread acceptance and has been approved by the NCRI board. Final funding approval is expected shortly and will include expenditures of one million pounds per year over the next 5 years, with a review at 3 years. The NCTR will be implemented in a two-stage process that includes clear milestones and measures of performance. The first, a demonstration project stage, will involve four linked pilot programs to evaluate and refine operational models; the demonstration project will inform the second stage—widespread implementation. An oversight committee will ensure that the NCTR meets the ongoing needs of all its stakeholders.

Appendix P

Nonprofit, No Government Funding Example: The SNP Consortium

History

In April 1999, 10 large pharmaceutical companies and the Wellcome Trust philanthropy formed a consortium to find and map 300,000 common single nucleotide polymorphisms (SNPs). The goal was to generate a widely accepted, high-quality, extensive, publicly available map using SNPs as markers evenly distributed throughout the human genome. The SNP Consortium viewed its map as a way to make available an important, precompetitive, high-quality research tool to spark innovative work throughout the research and industrial communities. By 2003, a total of 1.8 million SNPs had been discovered—many more than originally anticipated.

The international member companies, which together committed at least $30 million, were APBiotech; AstraZeneca Group PLC; Aventis; Bayer Group AG; Bristol-Myers Squibb Co.; F. Hoffmann-La Roche; Glaxo Wellcome PLC; IBM, Motorola; Novartis AG; Pfizer, Inc.; Searle; and SmithKline Beecham PLC, and the Wellcome Trust contributed at least $14 million.

Operations

The SNP Consortium's goal was achieved by providing research funding to existing academic centers currently involved in genome sequencing and/or mapping of genetic markers. The SNP identification was divided among the three research centers, including the Whitehead Institute, Washington University, and the Sanger Center (at the Wellcome Genome Research Campus). Direct radiation hybrid mapping was conducted at the Stanford Genome Center and the Sanger Center. Data handling, in-silico mapping, and bioinformatics were provided by Cold Spring Harbor Laboratory.

The SNP Consortium used DNA resources from a pool of samples obtained from 24 individuals representing several racial groups. This is a subset of the DNA reference panel for SNP identification collected by the NIH National Human Genome Research Institute. The anonymous, voluntary DNA contributions were made with informed consent specifically for this use.

SNP identification was performed by each center through sequencing of reduced representation fractions of the genome generated by selecting certain-size fractions of restriction digests of genomic DNA pooled from 24 unrelated individuals. The genome centers refined this technique through pilot studies completed in the first quarter of 1999.

Intellectual Property Management

The overall intellectual property (IP) objective was to maximize the number of SNPs that (1) enter the public domain at the earliest possible date and (2) are free of third-party encumbrances, such that the map can be used by all without financial or other IP obligations. To meet the second objective, the SNP Consortium withheld public release of identified SNPs until mapping was achieved, to prevent facilitating the patenting of the same SNPs by third parties. Mapped SNPs were publicly released quarterly, approximately one quarter after they were identified. The IP plan was intended to maintain the priority dates of discovery of the unmapped SNPs during the period between identification and release, for use as "prior art."

All parties received access to the SNPs at the same time. The SNP Consortium members and the genome sequencing centers were not granted advanced access to unpublished SNPs.

Appendix Q

For-Profit, Minimal Government Funding Example: Ardais Corporation

Ardais Corporation is a privately held clinical genomics company whose goal is to accelerate biomedical research by applying actual human disease, in the form of human tissue samples, as the discovery model in pharmaceutical research. The Ardais Biomaterials and Information for Genomic Research (BIGR™) System encompasses a unique repository, called the BIGR™ Library, with more than 120,000 research-quality tissue samples representing a broad diversity of disease. The samples are collected through the National Clinical Genomics Initiative, a strategic collaboration between Ardais and four leading U.S. medical centers.

The formation of the National Clinical Genomics Initiative was based on the identification of the broad need for tissues and data to support genomics-based biomedical research by researchers in academe and industry. The goals of the Initiative are to develop systematic, large-scale procedures to comprehensively collect, process, and store research-quality clinical materials and associated information; to provide these critical resources in highly optimized formats for efficient and robust design of biomedical research studies; and to support the research and clinical programs at each participating medical institution.

Ardais began the Initiative in January 2000 with Duke University Medical Center (DUMC) and Beth Israel Deaconess Medical Center, a Harvard University Medical School affiliate. The initiative was launched with these two participants in September 2000 and since has expanded to include Maine Medical Center and the University of Chicago. To date, the Initiative has collected, processed, and stored over 160,000 biospecimens from approximately 10,000 patients at its partner institutions. Approximately 50,000 samples per year are collected from a total of about 4,000 patients seen at the four sites.

The Initiative presented many challenges in tissue and data collection. Broad-based tissue collection by a for-profit company would require iron-clad patient protection. To address this issue, Ardais designed a comprehensive bioethics strategy that considered legal requirements, the recommendations of the National Bioethics Advisory Commission (NBAC), and suggestions from experts in government, industry, and academe to protect the privacy of patient donors.

Maintaining the quality of tissue samples and clinical information was a primary focus of the Initiative. Ardais developed standard operating protocols covering tissue sample access and archiving patient data collection and patient identification protection and dissemination to qualified researchers. These protocols were designed to maximize the molecular and histological quality of the samples. Ardais emphasizes the importance of high standards to disease research because:

- Only if one complies fully with all existing Federal, State, and institutional requirements regarding patient privacy can clinical materials be appropriately licensed to a researcher.
- Only if samples and data are collected under a unified set of standards can a researcher focus on biological variations between diseased samples rather than on variations due to poor sample handling and processing.
- Only if the relevant clinical and pathological information has been captured consistently in structured formats can a researcher select the appropriate samples easily and cost effectively.
- Only if a rigorous quality assurance/quality control (QC) process of sample validation and pathology verification has been applied can a researcher be sure that the sample is consistent with the hospital diagnosis and is of sufficient molecular integrity.

The need for a comprehensive information technology (IT)/bioinformatics system was also identified. Specific needs included:

- Sample tracking
- Case linking
- Web-based access
- Experimental design
- Browsing
- Remote deployment
- Multisite coordination

To satisfy IT and bioinformatics requirements, Ardais created the BIGR™ System as a discovery platform for application to drug discovery and development. The system includes a centralized, shared clinical genomics repository that encompasses tissue samples, molecular derivatives, and associated clinical information accessed by an array of bioinformatics tools.

Finally, Ardais developed standardized protocols to address the complicated logistics requirements (training, supplies, barcoding, shipping, storage, inventory control, infrastructure, and equipment) of a multisite repository system.

Program Structure

The Genomics Initiative pays rigorous attention to legal and bioethical issues. These steps include:

- Active dialogue with participants
- Formalized policies
- Formation of a Bioethics Advisory Board
- Required institutional review board (IRB) approval of all protocols (linked and unlinked)
- IRB oversight of distribution of samples to commercial entities

- A fully informed consent process
- Deidentified data
- Title for research use
- A stated policy of no interference with necessary medical procedures
- Procedures to obtain longitudinal clinical data followup
- Fully Health Insurance Portability and Accountability Act-compliant patient identity coding processes
- In line with NBAC report and the Code of Federal Regulations.

Ardais developed stringent collaboration operating procedures to minimize conflict with the participating institution. There is no interference with independent, ongoing tissue collection procedures. There is a stated desire to efficiently address the tissue collection needs at each site. Surgeon approval is required before patients can be approached. Consenting activities are coordinated during preoperative screening procedures to ensure that patients are not unnecessarily burdened with multiple consent requests.

The program structure at each site includes a dedicated staff and infrastructure, to ensure full focus on appropriate data and sample collection. The institution hires and pays the staff, with salary and benefit support provided by Ardais. Uniform protocols are implemented across all sites. All supplies and equipment for each site are obtained from the same vendors and are identical. Specific technical training procedures are utilized across sites. There is ongoing in-service training and QC monitoring throughout the Initiative.

The program management team at each site consists of representatives from Ardais Executive Management and the home medical center. Day-to-day management is the responsibility of the Banking Staff Management team (pathologists, bank manager, research nurse, project leaders) and the Ardais Alliance Management team. Joint steering/operating committee meetings occur at prescribed intervals (monthly via telephone or onsite). Open communication has been found to be critical to management success. A formalized structure has been developed for identifying and deploying best practices.

Consent

Ardais consent and confidentiality procedures are standardized across the four participating institutions. An informed consent interview (10 to 30 minutes in length) takes place with each patient. Appropriate patients are selected by research nurses, who have a success rate of over 95 percent in obtaining consent from those patients who are approached. A unique Ardais identifier and case identification are assigned to each consented donor for use in the BIGR™ Library. The link to the patient is maintained at the home medical center in a secure database (only banking staff have access). A time-stamped consent tracking form has been developed, which includes:

- Verification of every consent after sample banking (required prior to case release)
- Validation of time stamp on initial informed consent

- Indication if specimen was successfully banked

Sample Processing

Sample collection and banking procedures follow set guidelines at all institutions. After samples are retrieved from the patients, they are evaluated at a grossing station. A pathologist determines which surplus tissues can be banked, and both normal and diseased tissues are dissected. At a processing station staffed by trained, project-dedicated personnel, samples are cut and processed into modules. Samples are weighed, the time of each procedure is recorded, and all information is entered in the informatics system. Samples are linked to barcodes and stored in Ardais-designed protective containers (cryosettes). Samples are:

- Optimally frozen to enhance preservation of histologic quality and enable molecular analysis with histological specificity
- Matched to relevant, structured, and in-depth clinical data
- Available in frozen/optimal cutting temperature-embedded, and formalin-fixed/paraffin-embedded formats
- Of uniform size and shape
- Pathology-verified by board-certified pathologists and accompanied by verification data
- Preserved within 1 hour, postresection—actual time is tracked

QC and verification data associated with each sample are available through the BIGR™ Library. Medical center IRB-approved protocols are in place, with linked samples accompanied by patient informed consent that permits broad and ongoing data collection.

All samples are held at the home institution for an embargo period—usually 30 days. This permits the attending pathologist to retrieve additional tissue specimens for patient care, if needed. Once the pathologist has signed off on the samples, the embargo period ends and the samples are shipped to the tissue repository in Lexington, Massachusetts.

Shipping procedures follow similar uniform procedures and are tracked by the informatics system. Shipping containers consist of unified kits designed specifically for these purposes. There is strict adherence to Government regulations (Occupational Safety and Health Administration, U.S. Department of Transportation, and International Air Transport Association) for shipment of biomaterials. Personnel involved in shipping operations are highly trained. The informatics inventory management system can track where all samples are at any time and at any site—a "virtual" repository. The specific location of samples in a freezer can be ascertained.

It is necessary to maintain the link to the donor for ongoing data collection efforts, while adhering to all donor confidentiality precepts. At DUMC, the Medical Assistant on the World Wide Web (MAW3™) system was created to collect longitudinal data. Ardais is linked to this system as a central hub, with MAW3™ as a donor site. Each system collects and stores structured, comparable data. Systemized Nomenclature of Medicine—Clinical Terms is used to encode data elements in a common dictionary for communication between the systems. At

DUMC, data extraction from the database is needed. At Ardais, the necessary data are abstracted and put into an electronic, structured, deidentified form that does not allow the capture of patient-identifying information. Case report forms include 250 to 400 disease-specific fields for structured data as well as fields for demographics, history of the illness, surgical pathology findings, postoperative notes, discharge summary, and follow-up information. Image capture capability allows for display of representative images and translational detail of samples.

Clinical Genomics

Ardais provides quality samples for research use, with tissue collection based on identified research needs. All major cancers are represented in the repository as well as some of the rarer cancers. Ardais has board-certified pathologists on staff or serving as consultants and can provide sample qualification through pathology verification. All samples are pathology-verified prior to distribution to researchers. In addition, the company offers consultation for product development and human molecular pathology services. Company histotechnology capabilities include sample QC, slide production, and immunohistochemistry.

Pathology data are disconnected from the repository samples because pathological evaluation is conducted on the tissues retained for diagnosis, not on the samples donated to the repository. Therefore, pathology data are verified for the samples to make sure that the findings are consistent with the diagnosis and tissue composition of the original samples. A mechanism exists for reporting back major discordants in diagnosis. The event chain of exactly what happens to the samples after their removal from the patient is meticulously tracked. It is possible to tell who requests a sample, which personnel performed various analyses, and what specific processes were used.

Samples are either frozen or formalin-fixed (for 8 to 15 hours before processing)/paraffin-embedded. Ardais offers a variety of clinical genomics applications for sample processing:

- DNA extraction
- RNA extraction (90 percent of samples)
- Molecular derivatives
- Laser-capture microdissection
- Protein extraction
- Cell cultures
- Qualification data
- Microarrays
- Macroarrays (for angiogenesis studies)
- Tissue sections

Available frozen tissue arrays have 2-mm core dimensions and up to 40 samples per array, and are qualified for immunohistochemistry analysis. RNA is qualified for research use in:

- Gel-based analysis
- Bioanalyzer analysis
- Polymerase chain reaction-based analysis
- Array-based analysis
- Quantitative measurement of gene expression
- QC

In order to prioritize which samples will be collected at each site, target scores were created to determine what types of samples are of interest to the institution's researchers. Monthly reports are generated to identify how well the targets are met, if demands are satisfied, and if corrective measures are needed. There is structured quality feedback to the medical institutions.

Bioinformatics Library

The BIGR™ Library is a scalable, Web-deployable, Java-based system architecture that is managed from a central database. The Library is multi-institutional and will allow Web-based access to the repository on a researcher's desktop computer.

Each medical institution establishes a "Research Strategy Committee" to review requests for samples and ensure IRB compliance. It is possible to designate samples for restricted allocation to the home institution. Ardais has instituted an education protocol (involving beta system utilization) to inform researchers about the benefits of using BIGR™. To date, DUMC has trained 8 investigators, Beth Israel has trained 6 investigators, Maine has trained 4 investigators, and Chicago has trained 12 investigators to use the BIGR™ Library. The idea is to train the investigators so that they can/will apply to the committee for access to the live system. Once the distribution program is operational, there will be an open enrollment period for researchers.

It is possible to search for samples using:

- Diagnosis
- Identification (patient identifier)
- Attributes
- Tissue
- Appearance
- All of the above

It is also possible to click on a microscope icon and see a digitalized pathology image of the samples on the screen. When an RNA derivative is generated, all of the images and data are available and immediately viewable.

Appendix R

For-Profit, Primarily Private Funding Example: First Genetic Trust

First Genetic Trust (FGT) is a business that develops information technology (IT) solutions to address data, privacy, confidentiality, and ethical challenges in genomics and proteomics. It has created an IT platform with three goals: First, to enable large-scale genomic research and eventually clinical genomic research; second, to work with pharmaceutical companies to speed the development and use of new drugs; and third, to enable clinical adoption of genomics.

FGT is focused on supporting genetic research as a trusted third party, by providing a highly secure, Web-based IT infrastructure for genetic banking, as a cornerstone of an integrated research solution for patient recruitment and informed consent, and medical and genetic data acquisition, transfer, storage, and analysis. The patient, physician, investigators, administrators, and laboratory personnel can dynamically interface via the Web for patient education, information regarding the scope of the proposed research, and the consent process. The physician has similar access to aggregating phenotypical clinical data and obtaining clinical samples.

To address privacy and confidentiality protection, FGT has developed the enTRUST Genetic Banking System, using Web-based architecture and a highly secure, distributed genetic banking system. The FGT enTRUST Genetic Banking System consists of:

- An integrated genetic banking system designed for recruiting large-scale cohorts and developing and managing biospecimen and data resources on a worldwide basis
- A scalable, secure IT infrastructure, which includes encrypted data transfer and encrypted identity, phenotypic and genotypic databases, and high security access control
- A system compliant with legal, ethical, and regulatory aspects of collecting samples and data in the United States and Europe. FGT has designed its system to be fully compliant with the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the Data Protection Act and Data Protection Authority of the European Commission.[1]
- Regulatory compliance of computer systems (21 Code of Federal Regulations, part 11) and clinical trials (cGXP)[2] and quality assurance standards. As the trusted third party, FGT and enTRUST have been successfully audited for cGXP compliance by two major pharmaceutical companies, and they have been reviewed and approved for research use by 20-plus institutional review boards (IRBs).

This system uses "Virtual Vault," Hewlett Packard's military-grade operating system, which leverages standard security technology for encryption and intrusion detection and exceeds both

[1] See *europa.eu.int/comm/internal_market/privacy/index_en.htm* for information about the Data Protection Act.
[2] cGXP refers collectively to current Good Laboratory Practices, Good Clinical Practices, and Good Manufacturing Practices, as defined by the Food and Drug Administration to describe which records should be kept or what information is considered to be an official record.

HIPAA and European Directive requirements for data collection, consent, data accuracy, and adequate data security. Patients are assigned an encrypted electronic identifier, which serves as a virtual private identity and is stored in one dataset; phenotypic or clinical information is stored in a second dataset; and genotypic data are stored in a third dataset. The three datasets are linked through the patient's virtual private identity. Information is accessed through role-based access control mechanisms. Role-based access authorizes individuals to view specific information based on their role (e.g., research associate, Principal Investigator, statistician, IRB member) and is controlled via login identification.

FGT aggregates data via a Web-based architecture that interfaces with existing datasets. Data are accumulated, cleaned, aggregated, and stored in a repository. A common architecture in the system provides for distributed, centralized sample banking. There are four basic classes of access:

1. Patients or study participants have access to an electronic record of their consents.
2. Physicians and investigators have access to phenotypic and clinical data.
3. Research assistants who are characterizing the data can enter data online.
4. Sponsors/collaborators have access to the data they need to manage the study.

The FGT research management tools are all Web based. They include consent and reconsent modules (including information feedback to the patient, such as genetic counseling), clinical and genomic data capture, the ability to configure specific studies, sample logistics and banking, remote clinical data capture, study contract storage, and bioinformatics. Data representation standards support data exchange and mining, including aggregation of complex studies.

Multiple privacy categorization of tissues is possible, and FGT consent forms should be examined as possible models. The dynamic consent and reconsent mechanism allows for sharing of patient resources and permits information to be tailored to a particular study. The consent process uses an e-signature with a proxy option. This method also enables acquisition of current data for longitudinal studies.

FGT is involved in numerous sponsored research studies and biobanking/translational genomics initiatives with both academic and commercial entities. For example, it is working on a large pharmaceutical protocol in the United States and Europe, which received favorable review from 19 IRBs. Patient enrollment and sample data aggregation across multiple sites and countries are nearly complete. A second example is a breast and ovarian cancer research program at Memorial Sloan-Kettering Cancer Center that enrolled its first patient in early 2003. The company is working with a series of major academic institutions (e.g., Howard University), Pfizer, and the International Genomics Consortium on complex cohort and banking projects covering most of the functionality specified in this document.

Since FGT is a trusted third-party banking technology provider, data access rights and policies are determined by the sponsor of the banking initiative. Public and "managed data access" models both can be accommodated. Histopathological image data are not currently available, but it is technologically feasible to provide them. The design protocol can be written to automatically aggregate clinical updates and secular outcomes.

Appendix S

Private/Public Funding Example: International Genomics Consortium Expression Project for Oncology (exp*O*)

The International Genomics Consortium (IGC) is a private, nonprofit 501(c)3 medical research organization, operating to ensure full public benefit of the Human Genome Project. IGC's expression projects on human diseases will utilize a consortium model of enhanced and accelerated tissue collection under standardized conditions to produce detailed data on the molecular characterization of disease. The data will be available as a public, standardized, regulatory-compliant, readily accessible, and searchable series of genomic databases for use by the worldwide scientific community. The information emanating from this unique resource, which joins together the public and private sectors, will stimulate research on the underlying molecular mechanisms of diseases; the likely outcomes are the accelerated development of diagnostic tests, improved treatments, novel therapies, and disease prevention.

Infrastructure and executive management positions in IGC are funded through multiyear, committed support from entities in Arizona. Funding of the expression projects is derived through donations from members, which include pharmaceutical and biotech companies, technology and informatics companies, foundations, governments, academic and research institutions (providing in-kind support), and private donors.

Governance

IGC is governed by a board of directors, which oversees an executive management team comprised of a Chief Executive Officer, Chief Medical Officer, Chief Financial Officer, and Chief Information Officer. A scientific advisory committee speaks to the vision, operations, and opportunity represented by IGC. Each of the different expression projects for human diseases is advised by an executive steering committee (ESC), which is comprised of members representing the funding sources for the project. The ESC develops milestones for performance of IGC's projects, formulates metrics of deliverables, and offers advice regarding the execution of the projects. Subcommittees of the ESC present recommendations on technology platforms, information technology, clinical data, patient advisory affiliations, legal considerations, and public relations.

Common Ground for Success of IGC

The research direction of the IGC was developed through a series of open meetings held in multiple venues, which provided input from physicians, scientists, patient advocates, and leaders in the fields of pharmaceuticals, biotechnology, business and government to define the benefit from a national effort. The key leadership of IGC includes recognized leaders in cancer treatment, genomics, and medical research.

exp*O*

The goal of IGC's Expression Project for Oncology (exp*O*) is to standardize patient consenting, tissue collection, genomic analysis, and data reporting in a "precompetitive" and regulatory-compliant fashion. This international effort will obtain a significant number of both normal and cancerous tissue samples. Gene expression analysis will be performed on these samples at a central facility. Specific patient information associated with these tissues will be deidentified, and the data will be released into the public domain.

IGC has the opportunity to work with 19 medical centers throughout the United States; sites from the United Kingdom will also be included in the study, making exp*O* an effort with international support. These medical centers have agreed to provide samples for analysis and to collect patient outcomes under the direction of a designated Principal Investigator. IGC intends to utilize multiple host medical centers as hubs in the collection of annotated tissue samples. ICG also intends to follow these patients and provide outcome information. exp*O* uses a consortium model to bring together public and private sector involvement in the acquisition and analysis of human clinical tissue samples.

exp*O* as a Public Utility

IGC will release all exp*O* gene expression information into the public domain. The "consortium" strategy has been critical to the rapid pace and success of the Human Genome Project. Availability of exp*O* data to cancer researchers worldwide will foster discussion and comparison of existing computational and experimental approaches. The exp*O* data will encourage development of new bioinformatics approaches, leading to commensurately large datasets and significant breakthroughs in the area of cancer genetics, epidemiology, and therapeutics.

exp*O* Objectives

Database: The goal of exp*O* is to collaboratively obtain and directly perform gene expression analyses on a highly annotated set of normal tissue and tumor samples, with the release of all data into the public domain. Information from the gene expression analysis, along with the clinical information obtained at the point of acquisition of the tumor, will populate the exp*O* database, with regularly scheduled releases of the data. The data will be presented in a standardized, computationally traceable fashion, allowing integration with data from other allied scientific or clinical databases. The implementation of these informatics standards not only will maximize the utility of the data in the database but also will allow researchers to compare data independently generated, using the same methodologies and represented in the same standardized format.

Repository of residual tumor tissue, tissue adjacent to tumor, normal tissue, and blood samples from exp*O* patients for follow-on studies: IGC plans to obtain patient consent to collect clinical annotations, gather annual follow-up data, attain tissue and blood to support gene expression profiling, and enable follow-on studies that will be managed through a tissue redistribution

process. Such follow-on studies could include haplotyping, single nucleotide polymorphism (SNP) analysis, comparative genetic hybridization, DNA methylation, proteomics, kinomics, etc.

Intellectual Property (IP): IGC believes that the high-quality, standardized gene expression profile data, combined with standardized, clinical annotation resulting from the exp*O,* will best help accelerate oncology research and clinical efforts if the data are deposited in public databases and made available for the widest possible use by industry and academic researchers and clinicians. The goal of the IGC will be to make all exp*O* data publicly available and to impose no IP restrictions.

Of particular value to industry is the fact that all participating academic members have agreed to comply with the adopted IP strategy required by the IGC. Although the final decisions regarding the need to patent any information emanating from this project will be made by the exp*O* ESC, the model used for the data generated by the SNP Consortium will be considered as a solution.

Appendix T

Private/Public Funding Example: Organ Procurement and Transplantation Network*

History and Responsibilities

The National Organ Transplant Act (NOTA) of 1984 created the Organ Procurement and Transplantation Network (OPTN). Congress envisioned this to be an equitable national system that would be operated by the transplant community, including physicians and officials of transplant facilities as well as other specialists and individuals representing transplant patients, their families, and the general public.

NOTA gave the Secretary of the Department of Health and Human Services (HHS) oversight of the OPTN and responsibility for ensuring public benefit. Amendments to the Social Security Act in 1986 underscored the Secretary's role. Working in partnership with the transplant community, the Secretary has final authority over OPTN policies and procedures.

Until the enactment of the Omnibus Budget Reconciliation Act of 1986 (Public Law 99-509), membership in the OPTN was voluntary. Section 9318 of Public Law 99-509 added a new section (1138) to the Social Security Act. Section 1138(a)(1)(B) required hospitals that perform organ transplants to be members of and abide by the rules and requirements of the OPTN as a condition for participation in the Medicare and Medicaid programs. This requirement places at risk the transplant hospitals' participation in these programs, not just payments for transplantation, and as a practical matter makes the hospitals' survival dependent on following such rules and requirements. Section 1138(b)(1)(D) required that to be eligible for reimbursement of organ procurement costs by Medicare or Medicaid, an Organ Procurement Organization (OPO) must be a member of and abide by the rules and requirements of the OPTN.

The OPTN has responsibility for developing medical criteria for patient listing, medical urgency criteria ("status" definitions), organ allocation policies, other policies governing organ transplantation, and policies for the day-to-day operation of the OPTN. The Secretary has responsibility for (1) oversight of the OPTN, (2) establishment of performance goals and indicators to guide the national system for organ distribution, and (3) final approval of those OPTN policies that are to be enforceable. Both the OPTN and the Secretary have responsibility for dissemination of information to the public, including patients, physicians, payers, and researchers.

* Information obtained from the OPTN and UNOS Web sites (See: *www.optn.org* and *www.unos.org*).

HHS and OPTN Relationships

The United Network for Organ Sharing, a private corporation, operates the OPTN under contract with HHS. The contract is subject to the competitive bidding process. Under recent requests for proposals, there have been no effective competitors to the current contractor.

When the OPTN develops policies, or when complaints are raised concerning OPTN policies, there are a number of options. The Secretary may approve an OPTN-proposed policy or find that the complaint has no merit. The Secretary also may take another approach, depending on the issues presented. For example, the Secretary may: Seek broader public input on the issue; determine whether violations of OPTN-proposed policies should carry any of a range of consequences (no consequence, loss of membership in the OPTN, or loss of a hospital's ability to participate in Medicare and Medicaid); provide comments for the OPTN's consideration; direct the OPTN to adopt a policy; or develop a policy that the OPTN must follow.

Questions also have been raised about the relationship of OPTN policies to other standards and requirements. A number of Federal statutes, including those relating to Medicare and Medicaid, civil rights, fraud and abuse, clinical laboratories, organ procurement, control of infectious disease, and regulation of blood and blood products, have provisions that may affect or be affected by the policies of the OPTN.

In order to prevent such problems, a system was created in which the OPTN has three options whenever it identifies a policy that it believes will contribute to high performance. The OPTN can: (1) Recommend its use by members; (2) request that HHS make it enforceable; or (3) petition HHS to modify other regulations (such as clinical laboratory or blood regulations) to adopt that policy. What the OPTN cannot do is unilaterally impose a policy that has the effect of, or changes the terms of, a national policy already subject to the oversight of a cognizant Federal agency. The Secretary reviews the OPTN policies that may interact with other statutes or with rules promulgated through other Federal programs.

The OPTN Board

The 30-member Board of Directors is determined as follows: First, at least eight of the board members are to be transplant candidates, transplant recipients, organ donors, or family members, and none of these members or general public members may have an employment or similar relationship with the OPTN or with the categories of members listed in Section 121.3(a)(1)(I) or (iii)—OPOs, transplant hospitals, etc. Second, at least six members of the Board of Directors are to represent the general public; these members must be free of an employment or similar relationship to the OPTN or to institutions or individuals involved in transplantation. Third, not more than 50 percent of the board members (and of the Executive Committee) may be transplant physicians or transplant surgeons. Fourth, at least 25 percent of the board members must be transplant candidates, transplant recipients, organ donors, or family members of individuals in any of these categories.

Institute of Medicine Recommendations

In 1999, the Institute of Medicine issued a report, *Organ Procurement and Transplantation*, with five general recommendations for the OPTN. These included: Establishing organ allocation areas for livers; changing the waiting-time requirements for some liver transplantation patients; implementing Federal (HHS) oversight in the form of greater use of patient-centered, outcome-oriented performance measures for OPOs, transplant centers, and the OPTN; establishing independent scientific review; improving the collection of standardized and useful data regarding the system of organ procurement and transplantation; and making these data widely available to independent investigators and scientific reviewers in a timely manner.