**BEST PRACTICES FOR THE ALZHEIMER'S DISEASE RESEARCH CENTERS**

**INFORMATICS GUIDELINES**

The future of Alzheimer's disease research relies heavily upon research in the basic sciences, and flexible and robust informatics systems for biospecimen resources are vital to collaborative research efforts and progress. Best practices for biospecimen resource data system structure, function, and operational procedures are outlined below.

## I. Database Structure

A. ADRC local databases or biospecimen informatics systems should track, or have linkage capabilities to systems that track, a single biospecimen through all aspects of collection, processing, storage, dissemination, return, depletion, and disposal.

B. Biospecimen informatics systems should track associated clinical data and/or link to external sources of clinical data, where applicable.

C. At a minimum, it is recommended that biospecimen acquisition date and current availability status be tracked and linked to the Neuropathology Data Set, the Uniform Data Set, and the Biomarker and Imaging Database, where applicable.

a. Informatics systems should have linkage capabilities such that the physical tube or label of specimen containers or slides is linked to additional data on that specimen in the system.

b. It is recommended that each biospecimen be assigned a unique identifier in the form of a barcode and/or other identifying number.

c. Specimen ID format and database structure should be capable of tracking derivatives, aliquots, and mother/daughter sample relationships.

## II. Data Procedures

A. Informatics systems should be capable of generating a spreadsheet or CSV file representing current biospecimen availability that could be uploaded to NACC to keep the data current.

B. Informatics systems should be flexible and adaptable to add new biospecimen collection or processing protocols and data upload specifications as new specimen types are collected.

C. Informatics systems should be capable of performing the following functions: tracking, processing, data entry, data verification, querying, label printing/scanning, and audit trails.

## III. Sharing and dissemination of data

A. Center-specific guidelines that incorporate best practices for the dissemination of identifiable, de-identified and anonymous data, including genetic and biomarker data, are recommended to be established and adhered to for all data requests from academic and non-academic collaborators.

B. Policies and procedures for requesting ADRC resources should be published on each Center's website.

C. ADRCs should document and archive researcher requests for biospecimens and clinical data as well as the review outcome, and if possible, resulting publications with attribution to their grant. If possible, data should include characteristics of the individual researcher.

## IV. Quality Control, Security and Regulations

A. It is recommended that informatics systems document and monitor measures of biospecimen quality.

B. Database repositories are recommended to be installed on secured servers/network systems and should be backed up at least daily. For safety, additional copies of the data should be stored in a separate geographic location. The use of encrypted, cloud-based, multi-region, tiered storage is encouraged for long-term archival and disaster recovery as large providers offer automated backup, encryption, data redundancy, and data recovery for a fraction of what it would cost to implement these standards on the local level.

C. Network security may be established through consideration of (a) an institutional network firewall; (b) database password, user, group and role-based security; (c) application-level security with passwords and login required to access an application; (d) server-level access passwords. Multi-factor authentication is recommended for all administrator accounts.

D. Passwords should be required to contain, at a minimum, 8 characters including at least one number, capital letter, and special character. Users should be encouraged to use a secure and trusted password manager. This allows a user to auto-generate a unique, long, and complex password for each account while needing to remember only a single strong password. Additionally, two-factor authorization should be required wherever feasible, most notably on high-impact credentials such as root and administrator accounts.

E. Database write access is recommended to be limited to key authorized users and only from trusted Internet addresses, including trusted VPN address ranges.

F. Tiered-access should be specified to allow definition of “authority levels” for accessing and updating of data, particularly identifiable and genetic information These access definitions including user, database, and service account permissions should follow the Principle of Least Privilege.

G. Range checks and logical error checks are recommended for data, and as a quality control measure, errors should be flagged back to a user and disallowed entry into the database until repaired. All data entered into the database should be traceable back to a user through an automatic audit trail system. User metadata should be recorded entirely by the system, asking users to record their own name or identifier at the time of data entry is not a valid alternative.

H. Authorized data transfer is recommended to be protected via strong encryption capabilities through a secure Web or FTP site; data transfer via email is unacceptable. A minimum of a 128-bit encryption suite is recommended for web sites.

I. All databases must comply with HIPAA (Health Insurance Portability and Accountability Act of 1996) regulations to appropriately protect access to individually identifiable protected health information. All databases must comply with the Federal Information Processing Standards, if applicable.

J. All information on a participant should be linked within an ADRC by a common ID. This should be accomplished through an overall database design that allows creation of this ID at participant entry into the ADRC. Where possible, an NIA GUID should be generated and used as the participant ID or stored alongside the participant ID. See Best Practices GUID document.

## V. System Support and design

A. A designated team of institutional Information Systems personnel and system administrators are recommended to be in place for routine technical maintenance and trouble- shooting issues. Current CV and training records for Information Systems personnel should be kept on file.

B. Software development and data mining capabilities are recommended to evolve locally under the direction of a committee that may include database users and investigators, bioinformaticians, statisticians and software engineers.

C. The use of open-source platforms and software is heavily encouraged as it promotes sharing and collaboration, not only among ADRCs but with the larger research community.

D. When possible, systems should be designed with future integration and interoperability in mind. This includes designing modular and scalable architecture, considering secure ways to connect with new data sources, and planning for the potential for secure authorization of users and applications outside of the ADRC.

**References**:

1. National Cancer Institute, NCI best practices for biospecimen resources, 2011 (NCI Best Practices website: http://biospecimens.cancer.gov/practices/; PDF of the NCI Biospecimens Best Practice: http://biospecimens.cancer.gov/bestpractices/2011-NCIBestPractices.pdf