Abbreviations used in this proposal

AD:	Alzheimer's Disease
ADAD:	Autosomal dominant AD, defined by family history and validated
ADCS:	Alzheimer's Disease Cooperative Study
ADNI:	Alzheimer's Disease Neuroimaging Initiative
ADRC:	Alzheimer Disease Research Center at Washington University (P50 AG05681).
AlzForum	Alzheimer Research Forum website
BIRN:	Biomedical Informatics Research Network
CCC:	DIAN Clinical Coordinating Center
CDR:	Clinical Dementia Rating, our staging instrument where CDR 0 indicates no dementia and CDR 0.5 and 1 indicate very mild and mild dementia.
CNDA:	Central Neuroimaging Data Archive
DAT:	Dementia of the Alzheimer type, the clinically diagnosed syndrome of AD by known mutations in PS1, PS2, or APP genes
DCA:	DIAN Central Archive
DIAN:	Dominantly Inherited Alzheimer Network
MRI:	Magnetic resonance imaging
NACC:	National Alzheimer's Coordinating Center
OASIS:	Open Access Series of Imaging Studies
PET:	Positron emission tomography
PIB:	Pittsburgh Compound-B, a [11C] benzothiazole amyloid imaging agent
QC:	Quality control
XNAT:	Extensible Neuroimaging Archive Toolkit

A. SPECIFIC AIMS

The Informatics Core will serve as the central data management and dissemination facility for the Dominantly Inherited Alzheimer Network (DIAN). The following specific aims are proposed for the informatics core:

1. Provide data management services to capture and archive all project data in a central integrated database.
2. Provide data distribution and sharing services to distribute project data to internal and external investigators in an efficient and timely manner.
3. Provide support services to each of the cores, including user interface tools and automation services.
4. Provide information hosting services to make documentation and related materials available to DIAN investigators, external scientific users, and the general public.

B. BACKGROUND AND SIGNIFICANCE

Autosomal dominant Alzheimer's Disease (ADAD) represents a small fraction (<1%) of all Alzheimer's Disease (AD) cases, but it presents a unique window into the disease. Because individuals possessing known ADAD-causing mutations are destined to develop the disease at an early and relatively predictable age, they can be studied from a presymptomatic stage and the progression of the disease can be observed. DIAN will, for the first time, study ADAD in a systematic and comprehensive manner, acquiring biochemical, magnetic resonance imaging (MRI), positron emission tomography (PET), cognitive, and clinical measures from 240 individuals from families with known ADAD mutations. Carefully deidentified releases of the DIAN data will be made publicly available to researchers at the earliest appropriate time.

The informatics core will serve as the scaffolding upon which the DIAN program operates. The core will provide data access and entry services for acquisition and quality control sites. The core will also provide informational and data sharing services to end users. Given this central role, we have worked with each core to develop an informatics architecture that meets their requirements, and we have worked with the program's leadership to ensure that it meets the overall requirements of the DIAN program. The resulting architecture follows three basic priorities: secure and streamlined upload and entry of program data, integration of all program data in a single unified database, and user-friendly data sharing methods. Considering these priorities, the informatics core will be built on existing data archiving tools developed primarily in Dr. Marcus' laboratory. This open source, standards-based software has been vetted through active use over a number of

years in a variety of environments. In particular, it has been used for 6 years by the Washington University Alzheimer's Disease Research Center (ADRC) to run an integrated database of its neuroimaging and related data. This proven, stable infrastructure will be leveraged by the Informatics Core to support the DIAN program.

C. PRELIMINARY STUDIES

C1. Previous experience

Dr. Daniel Marcus will serve as the leader of the informatics core. Dr. Marcus has been developing neuroinformatics tools and standards, with a focus on neuroimaging databases and analysis, since 2001. He directs the ongoing development of the Extensible Neuroimaging Archive Toolkit (XNAT) (www.xnat.org) (Marcus et al., 2007a), an open source data management platform described in more detail below. Dr. Marcus directs the operation of the Central Neuroimaging Data Archive (CNDA) at Washington University and runs several public data sharing projects to freely distribute neuroimaging data, including Brainscape (www.brainscape.org) (Fox et al., 2007) and OASIS (www.oasis-brains.org) (Marcus et al., 2007b). Dr. Marcus also serves on a number of committees in the Biomedical Informatics Research Network (BIRN) to develop and promote informatics standards and techniques. He is the lead architect of the Morphometry BIRN Data Repository and is a co-author of the XCEDE data format (www.xcede.org) for representing and exchanging clinical and imaging data (Marcus et al., 2007c). Dr. Marcus is an at-large member of the National Alzheimer's Coordinating Center (NACC) Steering Committee.

The Informatics Core will be staffed by personnel from the Neuroinformatics Research Group (NRG) at Washington University School of Medicine, which Dr. Marcus directs. The NRG staff includes a number of computer scientists, neuroscientists, and programmers with many years of experience in biomedical, imaging and neuro- informatics. The group has team-developed a number of open source applications that will be used by the Informatics Core, including tools for data archiving, automated image processing, image visualization, DICOM management, and image analysis (Marcus et al., 2007d). Dr. Kevin Archie will implement image processing pipelines and user interface tools. Timothy Olsen will implement data structures and database tools. Angela Jones will provide data management and user support services.

Dr. Ronald Thomas will lead data services at the CCC. Dr. Thomas is Chief of Biostatistics and Bioinformatics in the Department of Family and Preventive Medicine at UCSD and is Director of the Data Core within the NIA-funded Alzheimer's Disease Cooperative Study (ADCS), a unique cooperative agreement among 100 academic and commercial research centers to conduct AD drug trials. He also serves as Director of the Biostatistics Core of the Shiley-Marcos Alzheimer's Disease Research Center and leads the Informatics Group within the Alzheimer's Disease Neuroimaging Initiative (ADNI) Clinical Core, which oversees and manages the ADNI Clinical and Biomarker datasets.

Dr. Cynthia Csernansky will lead the development of the DIAN informational website. Dr. Csernansky has substantial scientific and informatics experience. She led the development team for the Antecedent Alzheimer Biomarkers website, which includes a database backend to manage longitudinal, repeated measures of relational data for AD biomarker studies.

C2. Extensible Neuroimaging Archive Toolkit (XNAT)

XNAT will be used as the primary data management environment for the Informatics Core. XNAT is a software platform designed to facilitate common management and productivity tasks for neuroimaging and associated data (Marcus et al., 2007a). It follows a three-tiered architecture that includes a data archive, user interface, and middleware engine. A number of XNAT features make it particularly attractive for the proposed DIAN project: 1) support for a range of image upload/download methods, including DICOM, FTP, web services, and web browsers; 2) an extensible XML-based data model that simplifies the incorporation of new data types by automatically generating the necessary database tables and relations, user interface components, and search engine plug-ins; 3) quality control features, including a virtual quarantine that houses uploaded data until authorized users have validated them and a complete history profile that tracks all changes made to the managed data; 4) a secure web-based user interface for data entry and access; 5) a search engine that builds queries across data types; 6) an online image viewer that supports a number of common neuroimaging formats, including DICOM and Analyze; and 7) an pipeline engine for automating image processing routines. XNAT is in used for similar projects at a number of institutions across the country. The National Database for

Autism Research (http://ndar.nih.gov), for example, uses XNAT to federate imaging, genetic, and clinical data collected by NIH-funded autism researchers. XNAT is a core component of the BIRN suite of tools.

C3. Central Neuroimaging Data Archive (CNDA)

The CNDA has provided data archiving and analysis services to the Washington University neuroimaging community since 2003. It currently stores over 20,000 individual scans obtained from some 2,500 subjects, representing studies from a dozen departments and centers at the University, including the ADRC. It also manages thousands of non-imaging experiments, including neuropsychological, clinical, biomarker, and behavioral data. The CNDA is built on the XNAT platform and has many of the same requirements as DIAN. The Informatics Core will therefore leverage much of the CNDA's existing technology and infrastructure. For example, image processing pipelines have been implemented for post-processing and segmentation of 3T T1- and T2-weighted images and for registration and time course analysis of PIB PET imaging. Similarly, many of the clinical and neuropsychological evaluations that are part of the proposed DIAN battery have already been implemented in the CNDA. Managing the daily operations of the CNDA has also provided rich experience for Dr. Marcus and the NRG team in delivering exceptional service to investigators. In particular, via his 6 years of service to the ADRC, Dr. Marcus has developed close working relationships with the leaders of each of the DIAN cores.

C4. Biomedical Informatics Research Network (BIRN)

The BIRN provides a number of services, standards, and applications to support multi-site human imaging studies (Keator et al., 2007). In addition to the MRI sequences and quality control methods that will be used by the imaging core, the Informatics Core will be using a number of additional BIRN resources (see attached letter). XNAT itself is a core tool in the BIRN suite; in building the informatics core around it, DIAN will have access to the BIRN data and computational grids. The data grid and associated BIRN Data Repository (BDR) (http://www.nbirn.net/bdr/) will be used for distributing public access imaging data sets, and the computational grid will be used for processing that exceeds the capacity of internal resources. BIRN standards for data exchange, including web services and XML structures, will be used to communicate with the clinical data repository at UCSD, with quality control sites at Michigan and Mayo, and when sharing data. Dr. Marcus is an investigator in the morphometry BIRN testbed and has been instrumental in crafting the XCEDE XML format and web services (Marcus et al., 2007c).

C5. Public data sharing

Under Dr. Marcus's direction, the NRG has developed a number of collaborative data sharing projects. The Open Access Series of Imaging Studies (OASIS) (www.oasis-brains.org) (Marcus et al, 2007b) has made openly available a large number of imaging and associated clinical and demographic data from participants in the ADRC. The initial set of studies includes T1-weighted MRI scans of 416 subject aged 18-96, including 100 who have been clinically diagnosed with very mild to moderate AD. A longitudinal set of 150 participants imaged on 2-3 separate occasions is currently in preparation. The data are distributed via a user-friendly XNAT-based website, which has attracted over 1000 visitors since opening to the public in May, 2007. As an example application of the data, Blezak and Miller (2007) used OASIS data to study methods for generating representative target atlases for specific study samples.

The Brainscape project (www.brainscape.org) (Fox et al., 2007a), a collaboration with the NeuroImaging Laboratories at Washington University, allows users to upload resting state functional MRI (fMRI) studies to a central archive, process and analyze them, and share them with select collaborators or with the general community. To our knowledge, Brainscape is the first and only resource to provide open access brain imaging analysis and data sharing services. Brainscape is built on the XNAT platform. It currently has 90 registered beta testers from 13 countries and is scheduled for full release in November, 2007. Examples of shared data include Fox et al. (2007b) and Fox et al. (2005).

The morphometry component of the BDR (mBDR) (http://iaclin2.wustl.edu/mbdr) makes publicly available curated published data sets of MRI and related measures. The mBDR utilizes a number of BIRN tools, including XNAT, the Storage Resource Broker (SRB, a tool for creating distributed files system), and BIRNLex (a method for defining terms and measurements across multiple studies). The mBDR currently holds over 1,700 MR studies collected from nearly 1,000 subjects, along with several hundred associated

neuropsychological and clinical assessments. Dr. Marcus is the lead architect of the mBDR, and its ongoing operation is managed by the NRG.

C6. Alzheimer's Disease Cooperative Study (ADCS) Coordinating Center and Data Core

The ADCS Coordinating Center provides support for a range of AD studies, including ADNI. The services they provide include the generation of case report forms, site training, data collection and quality control, study tracking, interim reporting to the Data and Safety Monitoring Board, progress reports to the NIH and ongoing daily support to the PI and project directors. These services will be utilized within DIAN to form the DIAN Clinical Coordinating Center (CCC), as described in great detail in the Clinical Core (Core B) application. Specific to the informatics core, the ADCS's data management services will be used to capture clinical and neuopsychological data from the participating sites and to perform quality assurance routines on these data. These efforts will be led by Dr. Ron Thomas, the director of the ADCS Data Core. The data captured and vetted by the CCC will be transferred from their database to the DIAN central database using data exchange services that will be co-developed by the CCC and the DIAN Informatics Core.

C7. Alzheimer Research Forum (AlzForum)

The Alzheimer Research Forum, founded in 1996, is dedicated to understanding Alzheimer's disease and related disorders. The web site reports on the latest scientific findings, creates and maintains public databases of essential research data and reagents, and hosts discussion forums to promote debate, speed the dissemination of new ideas, and break down barriers across the numerous disciplines that can contribute to the global effort to cure Alzheimer's disease. The AlzForum team includes professional science writers and editors, information technology experts and web developers. AlzForum is an independent nonprofit organization.

AlzForum has created a special section on its website to provide informational resources on ADAD. Articles are posted about diagnosis, treatment, genetic counseling and testing, and life issues connected with having this hereditary disorder. The section addresses the complex issues of drug research in AD and the possible role of early-onset AD families in clinical research studies. It includes directories to help families and care providers find clinical research studies of ADAD and centers with expertise in early-onset dementias around the world. AlzForum will organize and implement an outreach and informational campaign directing ADAD families to DIAN, and DIAN and AlzForum will coordinate to disseminate updated information to both project's websites. Dr. June Kinoshita, Executive Editor, will lead AlzForum's contributions to DIAN, which will include publicity and outreach to recruit ADAD families (see attached letter).

D. RESEARCH DESIGN AND METHODS

The informatics core will provide four primary services: data management, data distribution and sharing, support services to the other cores, and information hosting. Data management will include oversight of the data flow from each acquisition site, integration of the data into a single data archive, and distribution of the data to and from quality control sites and the other cores. The core's data distribution and sharing services will include operation of a web-based portal that will provide a comprehensive and unified interface to the integrated database. The portal will include a role-based security system, allowing tiered access to the data; a customized search engine for querying the data within and across data types; a built-in image viewer; and custom tools for downloading raw and post-processed image data. The core's information hosting services will include a website with content for participating investigators (e.g. operations manuals, forms), external researchers (e.g. methods, data and specimen request forms), and the general public (e.g. recruitment, advocacy). Each of the core's primary services is described in more detail below.

D.1. Informatics Core Aim 1. Provide data management services to capture and archive all project data in a central integrated database.

D.1.1.The DIAN Central Archive (DCA)

The Informatics Core will develop the DIAN Central Archive (DCA), an XNAT-based database for storing all DIAN data. The DCA will include a flat-file repository of raw and processed image files and a relational database for storing non-imaging data (e.g. biomarkers, genotyping), derived image measures (e.g.

regional volumes, PIB time courses), imaging meta-data (e.g. scanner model, acquisition parameters), and links to image files. All measures and assessments in the DCA will be linked into a data hierarchy that captures the acquisition site, subject, and visit. A web-based user interface to the DCA will provide a user friendly view of the DIAN data.

The DCA will extend methods developed by the ADNI Coordinating Center and Informatics Core by co-locating *all* of the experimental data, including derived MRI and PET measures, in a single searchable database. This adds tremendous power to the database by allowing users to search and join across data types. As an example, the CNDA, which integrates the WU ADRC data in a manner similar to the proposed DCA, allows users to search for subjects with a specific Clinical Dementia Rating (CDR) score and PIB binding potential range. The users can then retrieve demographic, clinical, brain volume, and PIB binding potential measures for these subjects and download the associated raw and processed images. The integrated database also allows data mining and knowledge discovery methods to be implemented to identify unexpected patterns within the data.

D.1.2. Data capture

While the DCA will serve as the final data storage resource for all DIAN data, the method of initial data entry will be determined by the specific type of data to be entered. Imaging data and meta-data, for instance, will be uploaded and entered into the DCA itself, while initial subject registration, clinical assessments, and neuropsychological tests will be supported by the CCC web-based data entry system. Anonymized subject identifiers will be generated by the CCC consisting of a site number and random 4 digit number at the time of subject registration. Anonymized accession numbers will be generated by either the CCC or DCA to label individual acquisitions and samples. Data entered into the CCC system will be synchronized with the DCA in real-time via a set of secure web services. Data capture methods for each data type are described in more detail in the following sections.

D.1.2.1. MRI and PET

MRI and PET acquisitions will be uploaded from each site directly to the DCA. Secure uploads will be enabled using a standard SSL-enabled FTP client. A custom FTP server built into XNAT will capture the uploaded image files, parse the files (including header and binary content) to test for validity, and add the images to a “prearchive” holding area. The acquisition site coordinator or technician will then login to the DCA, locate the uploaded study in the prearchive, and enter study meta-data into an online form. The study images will then be stored into the permanent file repository. After the files are stored, a confirmation email will be sent to the contributing site study coordinator, the relevant quality control site, the Imaging Core leader, the Informatics Core leader, and the Informatics Core data manager. Archived images will be marked as “in quarantine” and made available to limited users until they have passed all quality control procedures.

As the data are stored to the permanent archive, all identifying header content (e.g. patient name, date of birth) will be removed from the image files. A non-altered set of the study data will be maintained in a separate repository, to which access will be restricted to the minimally required personnel. These files will be maintained as a safeguard against accidental removal of critical header information. At no time will they be distributed to users outside the informatics core.

Once the raw image files have been archived and notification emails delivered, the PET and MR quality control (QC) sites will be download the files via secure web services. XML meta-data and human readable “README” documents, detailing the scans and subject profile (e.g. age, gender) will be included with the downloaded image files. As post-processed image files are generated by the QC procedures, provenance and descriptive details will be appended to the delivered XML document. The XCEDE XML format developed by the BIRN will be used for these purposes (Marcus et al., 2007c).

When QC is complete, the derived image files, along with the XML record will be uploaded to the DCA via secure web services. Once uploaded, the images will be added to the file repository, and the XML record will be parsed and written to the database. The full dataset – raw acquired images and post-processed images – will be released from quarantine and made available to DIAN investigators via the DCA portal and web services according to the data access rules described in Section D.2.1 below.

D.1.2.2. Clinical, demographic, and neuropsychological data

All clinical, demographic, and neuropsychological data will be entered into web-based forms

implemented and deployed by the CCC. These forms and data representations for these data will be developed in close consultation between the Clinical Core, Informatics Core, and CCC. The QC methods of the CCC will be extensive. They will include development of clear and complete documentation of procedures and databases, cleaning of data and locking of completed databases prior to analysis, clinical monitoring of data both in-house and at sites, as well as computerized data editing. All procedures for performing QC checks will be fully documented and updated as needed. Data will be cleaned on an on-going basis, during the data collection phase of the protocol as well as after closure of the protocol. During each protocol, computerized data checks will be used to confirm that subjects meet inclusion and exclusion criteria at the time of entry, identify missing or out-of-range items, identify missing forms from visit packets, identify any duplicate entries into the database, evaluate longitudinal consistency between visits, and track subject status in the protocol (active vs. discontinued). Requests for corrections will be sent to sites and site monitors. The sites will be asked to provide missing information or to clarify contradictory responses.

Changes to the database record after initial entry will be automatically recorded in an audit trail. Each audit trail record will include the name of the field being changed, the reason for the change, the old and new values, change date, and the name of the person making the change. Only selected staff members will be permitted to make changes. The audit trail will allow full change histories to be constructed for any record in the database.

The CCC database and the DCA will be synchronized daily, and subject-specific updates will be requested from the CCC whenever image data are uploaded to the DCA or the subject's information is requested from the DCA website. Updates will be requested and delivered over web services, and data will be exchanged using the XCEDE XML format. Data collected by the CCC will be made available to the DCA immediately after entry, with clear indication whether or not it is has passed QC procedures. The DCA will maintain data in quarantine until it has passed QC at the CCC. This "real-time" synchronizing of the databases will provide data managers and site coordinators with a single interface via the DCA to view the status of subject data.

D.1.2.3. Genetics and biomarkers

Biospecimen kits for the genetics and biomarker protocols will be constructed and distributed to the acquisition sites by the CCC. Specimens will be labeled at the acquisition site with the subject ID, gender, year of birth, and date of acquisition and then shipped directly to the appropriate core for permanent storage and analysis. The site coordinator will then enter into the DCA details about the sample and shipping tracking information. A sample accession number will be generated by the DCA, and notification emails will be sent to the appropriate core leaders containing the shipping information. When the specimens are received, core personnel will log in to the DCA to retrieve the sample's accession number by subject ID and date of acquisition and indicate the condition of the received sample. Each sample will then be entered into the core's commercial laboratory management software by accession number for tracking inventory. Barcoded labels with sample accession number, date of acquisition, and additional protocol-specific information will be generated by this software and affixed to the sample tubes. The samples will then be processed and stored according to protocol. Finally, core personnel will enter meta-data (e.g. storage location, date of processing) and protocol-specific measures into the DCA. For the Genetics Core, these measures will include APOE genotype, mutation status and novel risk factor genotypes. For the Biomarkers Core, these measures will include cerebrospinal fluid levels of Aβ1-42, Aβ1-40, tau, and ptau181 and plasma levels of Aβ1-40, Aβ x-40, Aβ 1-42, and Aβ x-42. In addition, ongoing entry of inventory status will be entered into the DCA. Validation and quality control measures will be built into the entry forms, preventing mis-entry of the data. Email reminders will be sent to each core director if sample receipt and data entry are not complete within an acceptable window after a subject's visit.

D.1.2.4. Biostatistics

The Biostatistics Core will be responsible for preparing analytical data sets for addressing the core aims of DIAN and on an ad hoc basis to meet requests for data. Once papers using these data sets have been accepted for publication, the prepared data sets will be made publicly available via the DCA web portal. In addition, derived statistical measures (e.g. summary statistics, population statistics) generated by Biostatistics will be incorporated into the DCA to facilitate searches and use of the overall data.

D.1.2.5. Neuropathology

All DIAN subjects that come to autopsy will have their pathology reports, including diagnosis, entered by the Neuropathology Core into secure forms on the DCA portal. Within the DCA, the clinicopathologic and biochemical studies of each autopsied subject will be linked to the subject's clinical, genetic, neuropsychological, biomarker, neuroimaging, and demographic data.

D.1.2.6. Quality control

Multiple levels of QC will be implemented by the Informatics Core, in close collaboration with the CCC and Biostatistics Core. Data entry forms in the DCA and CCC websites will include validation code, to prevent entry of out of range values and to enforce entry of required fields. Once entered into the DCA, the data will be held in quarantine until a trained data manager clears them for general use. If data do not meet QC requirements, the acquisition site will be consulted to resolve concerns. In extreme cases, the data will be marked as unusable and held back from general use. All changes to data will be maintained in an audit trail available via the DCA web interface. Imaging data undergo extensive QC conducted by dedicated sites at Mayo (MRI) and Michigan University (PET), as described in the Imaging Core application. The Biostatistics Core will conduct regular statistical analyses of the DCA, inspecting it for outliers, missing data, discrepancies, and per-site irregularities. Again, any concerns that are identified will be raised with acquisition site. Systematic discrepancies in the data will be brought to the immediate attention of the Administrative Core.

D.2. Informatics Core Aim 2. Provide data distribution and sharing services to distribute project data to internal and external investigators in an efficient and timely manner.

D.2.1. Data access policies

DIAN investigators. All DIAN data will be made available to DIAN investigators via the DCA as soon as they are released from quarantine (i.e. passed all QC steps). Request forms for tissue will be included on the DIAN website. Once an investigator is ready to conduct a formal analysis and/or prepare a publication, the investigator will submit a request to the DIAN Publication Committee (chaired by Richard Mayeux), which will review all requests to ensure that there is no unwitting duplication or competition between investigators. The Committee will encourage investigators to consult with the Biostatistics Core for construction of the analytical datasets and for guidance in formal analyses. This approach provides investigators with dynamic on-demand access to the data via the DCA, while maintaining formal support channels for investigators. All analytical datasets will be released to the public via the DCA website once the manuscripts reporting them have been accepted for publication.

External investigators. External investigators will be granted access to DIAN data on a per case basis by the Steering Committee. Online forms on the DIAN website will be available for submitting data access requests. After review and approval by the Steering Committee, these investigators will be given access to a prepared view of the DCA that has been carefully screened to remove all possible measures that could link the data to a particular individual. The Biostatistics Core will be consulted on an ongoing basis to determine appropriate statistical methods (e.g. summary measures, conversion to categorical measures) for making available as much of the database as possible without compromising the privacy of DIAN participants. Each subject and sample in the database will be assigned a public identifier, which will prevent study participants who may have seen internal identifiers on study forms from gaining access to and recognizing their data. Because the same identifiers will be used each time a sample is distributed, investigators will be able to compare results between analyses and publications. All requests by external investigators will be sent to the Biostatistics core leaders and other appropriate core leaders, who will be available to assist investigators in preparing analytical data sets.

Public data sharing. Public releases of the DIAN data will be made available according to the policies established by the DIAN Data Sharing Committee. The data will be distributed via a public access version of the DCA. Again, the Biostatistics Core will be consulted to produce carefully de-identified data sets. In addition to the public DCA, DIAN data will be shared with the public via a number of widely used public access data sharing resources. The Uniform Data Set (UDS), which includes clinical, demographic, and neuropsychological measures will be distributed by the National Alzheimer's Coordinating Center (NACC). Imaging data will be distributed via the BIRN Data Repository. Genetics data will be distributed via the NIA-funded National Institute on Aging Genetics of Alzheimer's disease Data Storage site (NIAGADS)

(http://zork.wustl.edu/nia/). These data sharing resources have been selected because they place the DIAN data within the context of similar domain data and target investigators who are already familiar with these resources' methods. Use of the data will be subject to the terms specified by each of the resources.

D.2.2. The DCA web portal

The DCA will be made available via a user friendly web-based portal. The portal will include a range of productivity features to support investigators, core leaders, and technical staff. Basic site navigation will allow users to browse the various data types, view reports of specific subjects and drill down to specific studies. A built in viewer will allow users to visualize raw and processed MRI and PET images. Downloads of non-image data will be available as spreadsheets and XML documents, while downloads of image data will be managed by a secure file download manager. The portal's search engine will facilitate locating and joining data within and between study types. XNAT's standard tabular and views and structured reports will be supplemented with a number of custom views for DIAN. A study coordinator view will be created to facilitate data and subject tracking. The view will include a table with rows for each subject visit and columns for each study event (e.g. MMSE, PET scan). Colored icons in each column will indicate the status of that event for the subject visit (red: missing data, yellow: quarantined data, green: approved data). Similarly, an image analysis view will be created to display the state of MR and PET data in the quality control and processing workflow. Longitudinal and family views will also be created to assist investigators in exploring the data.

The portal will be password protected and utilize SSL-based data encryption. The portal security system will grant access to specific data based on the user's role within the DIAN project. For example, site coordinators and quality control personnel will be able to access data while still in quarantine, while investigators will be able to access the data only after it is released from quarantine. Approved external investigators will be granted access only to carefully cleaned and deidentified data. User logins and site activity will be saved to log files, which will be regularly checked for inappropriate or suspicious activity.

D.2.3. DCA web services

In addition to the browser-based user interface, the DCA will support web service-based access to the data. Web services for searching the database, retrieving studies, and uploading/downloading images will be available. The services enable the data to be accessed via batch scripts (for example, by the Imaging Core QC sites) and by client applications (e.g. image viewers). The web services, which are part of the XNAT platform, use the SOAP standard and are built on the Apache Axis framework. They will utilize the same security infrastructure as the rest of the DCA infrastructure.

D.2.3. User support

The informatics core will employ a number of user support methods. Manuals describing image upload and data entry procedures will be distributed to sites and made available online. The DCA portal will include "Help" links throughout, describing how to access, search, and download available data. For critical upload and data entry concerns, a telephone "hotline" will be manned 24/7 by on-call support personnel in the NRG. Less pressing inquiries will be directed to online resources including email support and a public web forum. An "FAQ" will be composed on the portal to proactively address common questions.

D.3. Informatics Core Aim 3. Provide support services to each of the cores, including user interface tools and automation services.

We have worked with each of the cores to determine their specific requirements, many of which are reflected in the previous text. A representative though not comprehensive listing of specific services to each of the cores is summarized in Table 1. The Informatics Core will interact with the cores on an ongoing basis throughout the project to support emerging needs. Dr. Marcus works closely with each of the core leaders in his ongoing role in the WU ADRC, and these relationships will be extended to support DIAN. The Informatics Core will meet formally with each of the core leaders via participation on the Steering Committee. Because particularly extensive interactions with the biostatistics and imaging cores are expected, Dr. Marcus will meet monthly with their core leaders. Meetings will be arranged on an as-needed basis to support the other cores.

The Informatics Core will work with the Imaging Core to automate many of the Imaging Core's image processing routines. In brief, the XNAT Pipeline Engine (Marcus et al., 2007d) will be used to specify the sequence of steps, input arguments, input and output files, and user input requirements necessary to complete

the tasks. The engine will monitor the execution of the tasks, update the database with derived data and processing history, request user input for manual steps, and notify users by email upon completion. The primary processing routines will involve: a) registration of the anatomic MR scans with the PET PIB and PET FDG images; b) segmentation and parcellation of the anatomic MRI images via FreeSurfer, a freely available and widely used image processing application; c) generation of regions of interest (ROIs) and gray matter volumes from FreeSurfer; d) application of the ROIs to the coregistered PET PIB and PET FDG scans; e) processing of regional activity values to yield regional PIB Binding Potential (BP) and relative metabolism. Many of these routines have already been automated for the daily operations of the ADRC. For more details see the Imaging Core application.

Core	Informatics services
A. Admin. / Steering Committee	Progress reports (weekly, quarterly, annual, ad hoc); discrepancy notifications; usage statistics; data sharing recommendations
B. Clinical	Progress reports; web-based neuropsych. assessments; integration with CCC database
C. Biostatistics	Data set preparation; meta-data access; joint reporting; upload of derived measures and analytics data sets to DCA; monthly meetings; integration of statistical measures into DCA
D. Neuropathology	Data entry & storage of diagnoses and banked tissue; submission of data to NACC; web-based forms for specimen requests
E. Biomarkers	Data entry; integration with inventory database; sample request forms
F. Genetics	Data entry; integration with inventory database; sample requests forms
G. Imaging	Image processing pipelines; data exchange with QC sites; de-identification of image files; monthly meetings

Table 1. Core specific services

D.4. Informatics Core Aim 4. Provide information hosting services to make documentation and related materials available to DIAN investigators, external scientific users, and the general public.

The Informatics Core will host a website for distributing information about DIAN to DIAN investigators, external investigators, and the general public. The primary functions will be facilitating communication among investigators at different sites and recruiting families to participate in the research projects. The website will support investigators by: 1) providing information about the individual projects, 2) providing detailed methods and references that can be shared with other investigators; 3) linking to research and biographical information about investigators, and 4) interfacing project management across all of the sites. The website will support recruitment of families into DIAN by: 1) providing clear information about the projects; 2) linking directly to the study coordinators, and 3) interfacing seamlessly with the ADAD section of the AlzForum website (http://www.alzforum.org/eFAD/default.asp).

The functions outlined above have been developed for the Antecedent Alzheimer Biomarkers website and will be adapted for DIAN. The site is built on a content management system that allows new content to be easily added by users with little or no web programming experience. A front page for the DIAN study will be created that outlines the structure of the program and introduces all of the participating sites. Each site will have a separate project page that can be customized to the needs of the investigators. The DIAN website will be updated regularly as new content emerges. All resources contained in the website can be accessed by one of three methods: e-mail, print or PDF to encourage dissemination of the DIAN protocols and supporting references.

Identifying and recruiting families to participate in DIAN will be accomplished by leveraging the combined resources of the DIAN website and Alzforum. Information for ADAD families about DIAN will be cross-linked to Alzforum to maximize the exposure of DIAN to potential families. AlzForum will organize and implement an outreach campaign directed to ADAD families. DIAN and AlzForum will coordinate to disseminate updated information to both websites.

D.5. Computing resources

The Informatics Core will expand on the NRG's existing computing environment. This structure includes 12 compute servers, 2 redundant application/database servers, and redundant load balancing switches. Four additional compute servers and 2 application/database servers will be purchased in each of Year 1 and 3 to support additional DIAN load requirements. The servers will be dual-CPU quad core units running the Linux operating system (Dell, Round Rock, TX). Data storage will be provided by a network-

attached storage (NAS) device with level 5 RAID and expandable to 250 terabytes (BlueArc, San Jose, CA). Data stored on the BlueArc storage system is replicated every 3 hours to a physically disparate disaster recovery site. Snapshots of data that has been changed or deleted on the BlueArc system are taken five nights a week. Thirty snapshots or 6 weeks of changed data are stored insuring that any accidental changes or deletions of data can be recovered. Incremental tape backups of the data will be generated monthly, and full tape snapshots of the data will be taken twice yearly and permanently archived.

The system and network architecture is designed to provide complete HIPAA compliance, near 24/7 uptime with no single point of failure, and grid-enabled compute services. The entire system resides behind the institutional firewall, which restricts incoming traffic by IP address and port. All of the servers will have private network addresses. Incoming network traffic will be directed to redundant public-address load balancing switches (Barracuda, San Jose, CA), and reverse proxy routing will be implemented to direct traffic from these units to the private-address application servers. This configuration ensures that machines holding data have no direct exposure to external network traffic. All network services (e.g. http, ftp) provided by the core will be encrypted using SSL.

The NRG resides within the Washington University Neuroimaging Laboratories, which has a dedicated IT infrastructure and professional support staff. All hardware will be deployed in a cold room with backup power and overseen by professional systems administrators in the Washington University Imaging Research Center Computer Support Group. The data network is connected to the Washington University backbone through a gigabit Ethernet connection. Central servers and critical workstations are also connected to the network at gigabit speeds, enabling high-speed data transfers between devices.

Additional processing capacity will be provided by the BIRN computation and data grid. The BIRN grid provides a hosting environment for running large batches of computationally intensive processing jobs on a secure, distributed infrastructure consisting of several hundred processing nodes. The jobs can be initiated and monitored via a web-based portal.

The ADCS Data Core implements a similar computing architecture. It is comprised of three groups of servers: application, database and storage. Web applications are written in Perl and are served up by load-balanced servers running Linux and Intel systems. The database system is a complement of Postgresql and MySQL. The proposed CCC system will be housed in the ADCS Data Core networking environment. The computing systems are linked to the UCSD WAN via a FastEthernet line. The internal ADCS systems deploy a mixture of FastEthernet and GigabitEthernet network equipment. The LAN employs a firewall appliance between the UCSD feed and the rest of the systems. Secondary firewalling services are enabled locally on the individual servers. Extensive backup systems for ADCS-managed databases are in place. The databases are backed up onto the file server hard disks daily and to tape weekly. Iron Mountain is employed to provide off-site backup to protect the data against theft, or loss, such as fire, flood, or earthquake.

D.6. Timeline

At the start of Year 1, an initial production version of the DCA database and web portal will be deployed with an initial focus on data entry and upload. This version of the system will allow data entry, upload, synchronization with the ADCS database, and distribution of image data to quality control sites. Users at each site will be asked to submit feedback on the usability and feature set of the website and upload tools. At approximately 6 months, participating sites will be required to obtain certification by 1) completing a phone-based website training course and 2) successfully entering the data for all aspects of a study visit by a volunteer subject. During the first 6 months of production, regular patches to the system will be deployed to resolve identified problems, change requests, and new features. Over the course of the project, occasional patches will be deployed as needed. Once the production DCA has stabilized, effort will shift to supporting image processing and analysis and to data sharing. Initial image processing pipelines for PET and MR will be fully implemented by the end of year 1. Pipeline development tasks will be ongoing over the course of the DIAN project, as new methods emerge. Data sharing will become a focus in Years 4, 5, and 6, with close interaction with the Biostatistics Core to develop anonymized and analytic datasets.

The informational website will follow a similar timeline. During the first three months, site design and content requirements will be gathered, culminating in a prototype website. Over the next three months, beta users will be invited to contribute feedback and additional content will be added to the site. The DIAN website will go into production at month 6. Coordination with AlzForum will also occur over the first six months of the project. Regular updates to the DIAN and AlzForum site will occur as needed throughout the life of the project.

E. HUMAN SUBJECTS RESEARCH

1. Risks to Subjects

a. Human subject involvement. No individuals from the families will be excluded from the study based upon gender/ race / ethnicity. Blood samples will be drawn from all individuals over the age of 18 yrs who are participating in the study. Individuals below the age of 18yrs are not recruited as dementia is not a childhood onset disease. In addition to giving blood and answering a family history questionnaire individuals will undergo a neurological evaluation, a neuropsychological test battery, lumbar puncture and neuroimaging.

b. Source of Materials. Blood specimens are collected at the initial assessment for DNA and genotyping. Continuous cell lines are established by NCRAD from all individuals. APOE and other genetic data are obtained for research purposes only; results are not provided to participants, their families or health care providers. All procedures are for research purposes only. Data is stored in locked filing cabinets in a locked office and in a secure database administered by the Informatics Core. There is no billing of insurance companies or Medicare.

c. Potential Risks. Slight discomfort, bruising or bleeding at the site of venipuncture and the possibility of lightheadedness. Occasionally an individual may faint after or during the venipuncture or there may be excessive bleeding at the site of needle insertion. Rarely, if subjects discuss their participation with their doctor, the fact of their participation may become part of their medical record. Because medical records can be available to insurance companies, they might then be unable to obtain health insurance or otherwise denied benefits just because they have participated in dementia-related studies. Because of the possibility that some types of dementia are hereditary, such risks may extend to family members. There are no long-term effects as a result of this procedure. There are no known adverse effects of the neurological exam or the neuropsychological testing. There is a possibility that a participant might find our questions tedious or embarrassing.

2. Adequacy of Protection Against Risk

a. Recruitment and Informed Consent. Subject recruitment occurs through public service announcements (radio, television, and print media), requests to private physicians, and organizations (e.g. the St. Louis Chapter of the Alzheimer's Association, Washington University Volunteers for Health) and word of mouth. All recruitment materials are approved by the Washington University HSC prior to use.

The corner stone of protection in research is the informed consent, but decision-making capacity often is impaired in subjects with AD. We address this issue in several ways (see Core B Human subjects section for details in the clinical core). There is a distinct informed consent document for participation in genetic research. The purposes of the study, the research and assessment procedures, and the risk/ benefit ratio will be explained to the subject or next of kin in the cases of demented individuals. Consent will be documented on the IRB approved consent form. Informed consent is obtained by a research nurse within the DIAN project. The purposes of the study, the research and assessment procedures, and the risk/ benefit ratio are explained to the subject or next of kin in the cases of demented individuals. Consent will be documented on the IRB approved consent form.

b. Future use of Data/Data Sharing. Several important elements are included in the consent form and emphasized in the Intake. These include: 1) the information and samples collected in the assessment are saved indefinitely to be used by researchers at Washington University or investigators outside the University, now and in the future to answer questions about memory, thinking, aging and other health concerns; 2) information is sent to NACC; 3) prospective authorization, i.e. giving consent indicates a willingness to continue even if you lose the ability to make decisions related to participation in research. 4) subjects waive their claim to the blood or products resulting from the blood. Specific attention is drawn to information that blood is frozen, stored indefinitely, and potentially shared with other investigators including NACC. 5) for those participating in the NCRAD study an addendum consent form is signed which specifies the use of tissue and data for that study.

c. Protection Against Risk—All data are safeguarded in accordance with HIPAA and the principles and practices of strict confidentiality. Studies are done for research purposes only. No services are billed to Medicare or Insurance Companies. Data and samples from the assessment are shared with DIAN-approved investigators. Data and samples are maintained by numerical code rather than personal identifiers. All data (including diagnostic information) are protected. Core activities are research related only. The ADRC has a Certificate of Confidentiality (#AG 99-02A) from the Department of Health and Human Services to protect the research data (including genetic data) and medical information from subpoena. No report is generated about participation in the genetic studies, nor is such participation included in the clinical summary paragraph sent to the private physician from the Clinical Core.

Embarrassment is minimized by the efficient, empathetic, professional manner of the DIAN staff. Venipuncture is performed by experienced nurse clinicians to minimize bruising or discomfort.

d. Durable Power of Attorney/Research Proxy (DPOA). (See Core B for details).

3. Potential Benefits of the Proposed Research
a. Potential Benefits to Subjects—Subjects may benefit from the thorough assessment of memory and thinking that accompanies participation in genetic studies. They will be given feedback about the findings of the clinical assessment and recommendations, when appropriate, will be given related to diagnosis, diagnostic testing and dementia treatments. Subjects are informed about any clinically significant condition identified.

b. Potential Benefits to Society—Society will benefit from advances in the genetic understanding of AD proposed by this study. The risks involved in this project are minimal but understanding the causes of dementia could ultimately lead to a better understanding of the pathogenic processes and more effective treatments.

4. Women and Minority Inclusion in Clinical Research
a. Inclusion of Women—Women are included in all studies described. Approximately 58% of Core B's total sample is women. This reflects the percentage of women in our community (60.5%) in the age group studied.

b. Inclusion of Minorities—All minority groups are encouraged to participate in this research. Intensive efforts are on-going to maximize the recruitment of African Americans, which is the largest minority population in the Metropolitan St. Louis area. Core B: Clinical and Core E: Education have formed a staff committee called the Minority Recruitment Committee. This committee develops recruitment materials and suggests strategies to Core B: Clinical to increase enrollment of African American subjects. The ADRC has developed an African American Advisory Board, which consists of representatives of the African American lay and academic community to advise Core B and Core E on how best to diversify the sample. The current % of African Americans in the Core B sample is 13%, identical to that of the community (12.6%).

F. VERTEBRATE ANIMALS
N/A

G. SELECT AGENT RESEARCH
N/A

H. LITERATURE CITED

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A. 2005 Jul 5;102(27):9673-8.

Fox, M.D., Marcus, D.S., Snyder, A.Z., Raichle M.E. (2007a) BrainSCAPE: A spontaneous correlation analysis processing environment for fMRI BOLD data.

Fox M.D., Snyder A.Z., Vincent J.L., Raichle M.E. (2007b) Intrinsic fluctuations within cortical systems account for inter-trial variability in human behavior. Neuron, 56:171-184.

Keator, D.B., Grethe, J.S., Marcus, D., Ozyurt, B., Gadde, S., Murphy, S., Pieper, S., Greve, D., Notestine, R., Bockholt, H.J., Papadopoulos, P., Function BIRN, Morphometry BIRN, BIRN Coordinating Center. (2007) A national human neuroimaging collaboratory enabled by the Biomedical Informatics Research Network (BIRN). IEEE Trans on Info Tech in Biomedicine. doi: 10.1109/TITB.2007.896875.

Marcus, D.S., Olsen, T., Ramaratnam, M., and Buckner, R.L. (2007a) The Extensible Neuroimaging Archive Toolkit (XNAT): An informatics platform for managing, exploring, and sharing neuroimaging data. Neuroinformatics, 5: 11-34.

Marcus, D.S., Wang, T., and Buckner, R.L. (2007b) Open Access Structural Imaging Series (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented and Demented Older Adults. Journal of Cognitive Neuroscience, 19:1-10.

Marcus, D.S., Gadde, S., GRETHE, J.S., Keator, D.B., Pieper, S., Bockholt, H.J., BIRN. (2007c) XML-based clinical experiment data exchange (XCEDE) schema for the storage, retrieval, transport and maintenance of imaging, behavioral and clinical data.

Marcus, D.S., Archie, K.A., Olsen, T., Ramaratnam, M. (2007d) The open source neuroimaging research enterprise. J. Digital Imaging. doi: 10.1007/s10278-007-9066-z.

I. MULTIPLE PI LEADERSHIP PLAN

N/A

J. CONSORTIUM/CONTRACTUAL ARRANGEMENTS

N/A

K. RESOURCE SHARING

N/A

L. CONSULTANTS

Dr. Cynthia Csernansky will consult 250 hours during the first year to gather requirements and content for the informational website.