HHS Public Access

Author manuscript

Neuropsychol Rev. Author manuscript; available in PMC 2017 April 21.

Published in final edited form as:

Neuropsychol Rev. 2015 September ; 25(3): 356–368. doi:10.1007/s11065-015-9293-x.

Neuroinformatics Software Applications Supporting Electronic Data Capture, Management, and Sharing for the Neuroimaging Community

B. Nolan Nichols[1,2] and **Kilian M. Pohl**[1,2]

[1]Center for Health Sciences, SRI International, Menlo Park, CA, USA

[2]Department of Psychiatry and Behavioral Sciences, School of Medicine, Stanford University, Stanford, CA, USA

Abstract

Accelerating insight into the relation between brain and behavior entails conducting small and large-scale research endeavors that lead to reproducible results. Consensus is emerging between funding agencies, publishers, and the research community that data sharing is a fundamental requirement to ensure all such endeavors foster data reuse and fuel reproducible discoveries. Funding agency and publisher mandates to share data are bolstered by a growing number of data sharing efforts that demonstrate how information technologies can enable meaningful data reuse. Neuroinformatics evaluates scientific needs and develops solutions to facilitate the use of data across the cognitive and neurosciences. For example, electronic data capture and management tools designed to facilitate human neurocognitive research can decrease the setup time of studies, improve quality control, and streamline the process of harmonizing, curating, and sharing data across data repositories. In this article we outline the advantages and disadvantages of adopting software applications that support these features by reviewing the tools available and then presenting two contrasting neuroimaging study scenarios in the context of conducting a cross-sectional and a multisite longitudinal study.

Keywords

neuroimaging; neuropsychology; biomedical informatics; neuroinformatics; mri; data sharing

1. Introduction

Making data from biomedical studies freely available to the research community is an increasingly prevalent mandate of funding agencies (Collins & Tabak, 2014) and publishers (Bloom, Ganley, & Winker, 2014). For example, the National Institutes of Health (NIH)1 stated in 2003 that “all investigator-initiated applications with direct costs greater than $500,000 in any single year will be expected to address data sharing in their application” and

Corresponding author: B. Nolan Nichols, 333 Ravenswood Ave, Menlo Park, CA 94025, nolan.nichols@gmail.com.

Conflict of interest: Neither author has conflicts of interest with the information presented herein.

[1]http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html

that “the timely release and sharing to be no later than the acceptance for publication of the main findings from the final data set.” More recently the NIH Director Dr. Collins commented on “…the failure of funding agencies to establish or enforce policies that insist on data access” (Collins & Tabak, 2014) and called to “embrace an era in which transparency and responsible data sharing are common value” (Hudson & Collins, 2015). These sharing directives are supported by recent neuroimaging studies demonstrating that the reusability of data is a key scientific resource (Breeze, Poline, & Kennedy, 2012; Mennes, Biswal, Castellanos, & Milham, 2013; Poline et al., 2012). The cultural shift from data ownership by a closed group toward data sharing within an open community is particularly relevant for recent “Big Data” studies (Fjell et al., 2012; D. S. Marcus et al., 2011; Mennes et al., 2013; Toga, Crawford, Alzheimer’s Disease Neuroimaging Initiative, 2010), neuroimaging data repository efforts (Gorgolewski et al., 2015; Hall, Huerta, McAuliffe, & Farber, 2012; Poldrack et al., 2013), and authors who publish their work at journals such as Proceedings of the National Academy of Sciences (Cozzarelli, 2004), Journal of Neuroscience (Shepherd, 2002), Journal of Cognitive Neuroscience (D’Esposito, 2000), and Public Library of Science (Bloom et al., 2014). In summary, a key element for high impact research in neuroimaging is becoming the integration of data sharing into the study design. This article supports this task by reviewing software tools, including data repositories, aiding in electronic data capture, management, and sharing within the neuroimaging community.

The motivation behind data sharing requirements is likely driven by the promise to maximize the knowledge gleaned from neuroimaging studies through exploration of ‘reusable data’ (Breeze et al., 2012; Kennedy, Haselgrove, Riehl, Preuss, & Buccigrossi, 2015; Poldrack & Gorgolewski, 2014; Poline et al., 2012). Reusable data includes primary data (i.e., raw observations such as brain images or neuropsychological measures) and secondary data (i.e., derived measurements such as image segmentations or composite scores) that are curated in a format easily and freely accessible to the research community. Reusable data can augment the information available in databases, such as BrainMap (Fox & Lancaster, 2002; Laird, Lancaster, & Fox, 2005) and NeuroSynth (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011), allow researchers to explore alternative hypothesis, and reduce concerns about the reproducibility of discoveries by performing independent replication studies (Ioannidis, 2005). To date, however, neuroimaging studies generally reduce access to their data to summary statistics published in journal papers, such as p-values associated with brain atlas coordinates (Lancaster et al., 2000; Tzourio-Mazoyer et al., 2002). Without access to neurocognitive data (i.e., primary and secondary data) the potential to aggregate datasets, boost statistical power of findings (Button et al., 2013), or to inspect the dataset for non-significant findings omitted from the original manuscript (David et al., 2013; Ioannidis, 2011) is mostly limited to meta-analysis, whereby the results of comparable studies are examined collectively to corroborate findings (Caspers, Zilles, Laird, & Eickhoff, 2010; Salimi-Khorshidi, Smith, Keltner, Wager, & Nichols, 2009).

Sharing neurocognitive data is generally a resource intensive activity as it requires curating data so that it is meaningful to the research community (Howe et al., 2008). This situation poses a quandary for principal investigators of neuroimaging studies needing to choose between using their assets to deliver reusable data or to pursue new scientific questions and

hypotheses, particularly when only the latter may lead to new funding opportunities. To reduce barriers associated with creating reusable data, the Neuroinformatics community provides software for electronic data capture, management, and sharing (Poline et al., 2012). The software packages are generally based on best practices developed by that community, which include standardized data formats and metadata representations (Bjaalie & Grillner, 2007).

Neuroinformatics started in the 1990's, when scientists applied the principles of Biomedical Informatics (Kulikowski et al., 2012) to develop reusable tools supporting the data analysis needs of the Human Brain Project (HBP) (JE Brinkley & Rosse, 2002; Huerta & Koslow, 1996; Shepherd et al., 1998) and individual labs investigating relationships between brain and behavior (Young & Scannell, 2000). In the neuroimaging domain, those data analysis tools lowered computational barriers to advances in brain science by enabling investigators without training in software development to harness increasingly complex analysis methodologies (Cox, 1996; Fischl et al., 2004; Smith et al., 2004). Since the 90's, the scale and scope of neuroimaging studies have dramatically increased as have the number and complexity of the tools needed to process those data (Ferguson, Nielson, Cragin, Bandrowski, & Martone, 2014; Gomez-Marin, Paton, Kampff, Costa, & Mainen, 2014; Van Horn & Toga, 2014). In addition to data analysis, early Neuroinformatics efforts focused on experiment management systems that helped to address challenges of complex studies for which the state-of-the-art at the time (i.e., spreadsheets and images stored in directory structures on a file system) became insufficient for effectively fulfilling study objectives (JE Brinkley & Rosse, 2002). Today, Neuroinformatics continues to adapt to the changing requirements of neuroimaging studies by providing software tools that simplify the capture, management, and sharing of neurocognitive data.

This article reviews the state-of-the-art for capturing, managing, and sharing data of neuroimaging studies. Specifically, Section 2 provides an overview of the informatics approaches designed to facilitate electronic data capture and management (Section 2.1), complying with data sharing policies (Section 2.2), and repositories specializing on distributing neurocognitive data (Section 2.3). We then embed these tools in a practical setting by reviewing two study scenarios (Section 3) that contrasts a cross-sectional study (Section 3.1) with a multisite longitudinal study (Section 3.2). In each scenario, we present a study description, requirements, and neuroinformatics approaches available to assist researchers in developing best practices in their own lab and for adhering to more stringent data sharing policies. We complete this review with a discussion of the relative advantages and disadvantages of deploying the systems detailed herein with the goal of helping neuroimaging labs make informed decisions on choosing neuroinformatics tools for electronic data capture, management, and sharing.

2. Electronic Capturing, Managing, and Sharing of Neurocognitive Data

Studies that examine brain and behavior relations are increasing in complexity with the agencies funding projects with a larger number of subjects, cognitive tests, and imaging modalities (Jack et al., 2008; Thompson et al., 2014; Van Essen et al., 2012). As mentioned earlier, neuroimaging labs traditionally rely on spreadsheets and a file system to capture,

manage, and share data (Poline et al., 2012). Increasing the scale of research projects, many research labs are confronted with new technical (e.g., data size, quality control, analysis complexity) and social (e.g., employee turnover, data sharing requirements) challenges (Buckow, Quade, Rienhoff, & Nussbeck, 2014). To address these challenges, many labs develop homegrown data management systems to create immediate solutions that may not address long-term socio-technical issues related to scalability (Franklin, Guidry, & Brinkley, 2011). For example, a lab may develop a system for data entry where information from paper forms is typed into a single "data entry computer." As the lab grows or the complexity of a given study so will the need for multi-user access to the database, paperless electronic data capture, and automated uploading of computerized neuropsychological assessments, for example see (Gur et al., 2010; Kane & Kay, 1992), to central data repositories (Hall et al., 2012). Rather than investing time in the development of new software, a more practical approach is to use existing solutions (Franklin et al., 2011). We now review these technologies by describing electronic data capture and management systems (Section 2.1), data management plans complying with NIH policies regarding the protection, management, and sharing of data (Section 2.2), and data repositories (Section 2.3) that can be used to maintain and distribute the data.

2.1 Electronic Data Capture and Management

Choosing the right electronic data capture and management systems (EDCMS) for a specific lab environment requires carefully evaluating current research workflow, the type of data to be captured and managed by the system (e.g., clinical/neuropsychological forms or medical imaging), and available information technology (IT) resources (e.g., networking and data management personnel). For example, studies with a small neuroimaging component and an extensive neuropsychological test battery administered using paper and pencil, such as in (Meier et al., 2012), may be best served by an EDCMS with excellent double-data entry support to reduce errors from manual data entry. Alternatively, multisite studies focusing on brain imaging, such as in (Fjell et al., 2012; Jack et al., 2008), may select a research Picture Archiving and Communications System (PACS) (Greenes & Brinkley, 2006) with enhanced support for the Digital Imaging and Communications in Medicine (DICOM) standard (Hussein, Engelmann, Schroeter, & Meinzer, 2004) to help automate, for example, the archival and de-identification of images (Haak, Page, Reinartz, Krüger, & Deserno, 2015). To gain access to such systems, labs with adequate informatics expertise and IT resources may choose to install and maintain an EDCMS on their own computer environment so that they can fine-tune and customize the deployed system. Alternatively, research labs might want to access an EDCMS hosted by another institution on the Web (e.g., access is provided by contract or fee) (Book et al., 2013; Scott et al., 2011; Van Horn & Toga, 2009) or provided by a service center within their own institution (Bernstam et al., 2009) to avoid the operating cost of maintaining an EDCMS. The remainder of this section reviews a subset of the most widely used tools and services targeted towards electronically capturing and managing neurocognitive research data (summarized in Table 1), which were selected from the Neuroimaging Informatics Tools and Resources Clearinghouse (Kennedy et al., 2015) and the Neuroscience Information Framework (Gardner et al., 2008) resource registries:

- **Collaborative Informatics Neuroimaging Suite (COINS):** COINS (Scott et al., 2011) is an EDCMS developed by the Mind Research Network (MRN) to support both medical imaging data and clinical assessments. It consists of Web applications that support sharing documents and monitor study progress, search, filter, and retrieve datasets, support the design of forms for dual-data entry and provide tools for managing studies and subject information including notifications to resolve study issues (e.g., data are uploaded for a subject that is not enrolled in a study). In addition, COINS provides Web services (i.e., automated communication between devices and computers), which can be used to transfer assessment data of tablet devices or medical images directly into COINS. COINS is a hosted EDCMS that requires a contract with MRN, thus eliminating the need to install and maintain the database system.
- **Human Imaging Database (HID):** HID (Keator et al., 2009) is a Web application developed by the Biomedical Informatics Research Network (BIRN) (Helmer et al., 2011). It supports medical imaging using DICOM and entry of clinical measures using a clinical assessment layout manager (CALM). CALM allows researchers to design digital forms that resemble their paper-based counterparts to minimize the difference between these two formats. HID also incorporates data exchange standards (Gadde et al., 2012) and can operate across distributed sites (Keator et al., 2009). While no longer under active development, HID is an open sourced project that has been the inspiration to many other systems.
- **Image Data Archive (IDA):** IDA (Van Horn & Toga, 2009) is a data repository and management system developed by the Laboratory of Neuro Imaging (LONI) to manage single-site and multi-site brain imaging studies. It supports automated and semi-automated import of medical imaging and clinical assessment data using Web-based tools. These tools support data verification and upload, querying the image database, and an image viewer for quality assessment. IDA also interfaces with the Pipeline software package (Dinov, 2009), a computational workflow engine for image processing. IDA uses access control and audit trails of all data access to ensure patient privacy and security. To use the centralized IDA repository requires becoming a LONI Resource Collaborator[9].
- **Longitudinal Online Research and Imaging System (LORIS):** LORIS (Das, Zijdenbos, Harlap, Vins, & Evans, 2011) is a Web-based system supporting multi-center studies with a longitudinal study design. Project coordination is done through an integrated Web-based medical image browser that streamlines the process of data entry and quality control of neuroimaging data. The Web-based tool provides a designer for data entry forms. The forms can link scalar measures (e.g., behavioral and neuropsychological measures) with scoring functions to minimize data entry and calculation errors. Imaging measures can be

[9] http://resource.loni.usc.edu/collaboration

uploaded to a database via a DICOM compatible Web application. Once uploaded, the images are processed for quality control. LORIS also provides a data query tool for filtering data records and downloading information from the database. As reported by (Das et al., 2011), LORIS had already been the data management solution for nine large-scale international projects. Installing this open-source system (i.e., it is fully customizable) within a lab environment requires experience in administration of Web applications.

- **Neuroinformatics Database (NiDB):** NiDB (Book et al., 2013) is a Web application that can be used to upload, download and search neuroimages. It supports processing of DICOM files coupled with a graphical user interface (GUI) for quality control. Non-imaging data (e.g., eye tracking, fMRI behavior files) can be uploaded and associated with these DICOM files. NiDB includes predefined forms to enter basic information about study participants (e.g., sex, age at scan, height, and weight). Packaged as open-source software, NiDB includes an automatic installer similar to many commercial software packages, an online demo10, and a pricing model should one want to externally host the system.
- **Research Electronic Data Capture (REDCap):** REDCap (Harris et al., 2009) is a Web application that supports clinical electronic data capture excluding imaging data. Data capture is defined by forms that can be customized for cross-sectional, longitudinal, and multi-arm designs ranging from small research studies to large clinical trials. Data entry forms are configured through a data dictionary, which can contain supporting variable names, human readable labels, field validation, double-data entry, and branching logic (i.e., skipping or including additional fields based on responses). To simplify the setup of a study, REDCap provides a library with over 650 preconfigured data entry forms11. REDCap also provides an Application Programming Interface (API) for automating data management, integration, and querying. The user evaluation of clinical data management systems by Franklin et al. (Franklin et al., 2011) preferred REDCap's generic design over several other systems with similar features. To gain access to the software, research organizations must join the REDCap Consortium. As of today, the REDCap website lists 1,455 active institutional partners in 90 countries with over 169,000 projects and over 230,000 users. Installation of the REDCap system requires knowledge of Web application administration.
- **eXtensible Neuroimaging Archive Toolkit (XNAT):** XNAT (D. Marcus, Olsen, Ramaratnam, & Buckner, 2007) is a Web application designed primarily as a research PACS that couples DICOM sessions with data validation and metadata extraction (i.e., data about data). Imaging and non-imaging data are organized by subject and projects and include searchable metadata fields depending on the type of data (e.g., MRI scans have information about scan acquisition

[10] http://demo.neuroinfodb.org/login.php
[11] https://redcap.vanderbilt.edu/consortium/library

parameters). XNAT does not include a Web-based tool to design data entry forms that is comparable to LORIS or REDCap; rather a programmer must design and deploy a configuration document (i.e., an extension to XNAT's XML Schema12). XNAT is an open source system, which requires installation, configuration, and maintenance by personnel with knowledge of Web application administration. Labs with software developer resources can add features by implementing extension modules (e.g., protocol validation) and make use of an API to automate tasks. Furthermore, XNAT provides an online marketplace13 where additional plugins and data types can be searched for and installed.

To summarize, we reviewed seven data capture and management systems that aid neuroimaging studies in electronically capturing and managing data. Each tool affords a Web-based interface enabling researchers to upload, manage, and share data from any computer connected to the Internet. They also provide a mailing list, user documentation, online demos, and an issue tracker with links available from their website. Not under active development anymore, HID provided a proof-of-concept for many other systems. COINS and IDA are unique in that they offer a purely centralized solution without the need to install and maintain the system locally. NiDB is the only system that can both be installed by an individual lab or hosted online for a fee. REDCap is the only system we reviewed that focuses primarily on non-imaging data and has been widely deployed for electronically capturing phenotypic data. XNAT is the most broadly deployed open source system for medical imaging data with strong community-based user support. Both REDCap and XNAT are the only systems providing an API for automating data management tasks. Given the strength of each system, their deployment within a research setting requires careful evaluation of the studies they should serve.

2.2 Data Management Plan

As stated by the NIH in 2003 (see Introduction), applications with direct costs greater than $500,000 in any single year must include a data management plan. The data management plan specifies the protection, management, and sharing of data collected by the proposed study in compliance with the funding agency. For example, the National Institute of Mental Health (NIMH) requires14 that clinical research with human subjects need to submit data to one of the NIMH Data Archive systems. To identify the correct repository, a flow chart is provided15. Another important part in the data management plan is specifying the de-identification of patient data so that the privacy and confidentiality of study participants is maintained. Human subjects' data containing Personal Health Information (PHI) must first be sanitized of specific information before it can be made public according to the Health Information Accountability and Affordability Act16. Ignoring specific data types (e.g., genomic data and protected populations) for a moment, NIMH requires that the following information be removed before sharing the data: 17

12 http://www.w3.org/XML/Schema
13 http://marketplace.xnat.org
14 http://grants.nih.gov/grants/guide/notice-files/NOT-MH-15-012.html
15 http://rdocdb.nimh.nih.gov/wp-content/uploads/repository_flowchart1.png
16 http://www.hhs.gov/ocr/privacy/hipaa/understanding/index.html
17 https://ndar.nih.gov/policies_standard_operating_procedures.html#sop5

1. Participant Names
2. Postal address information, other than town or city and state or 3 digit zip code;
3. Telephone numbers;
4. Fax numbers;
5. Email addresses;
6. Social Security numbers;
7. Medical record numbers;
8. Health plan beneficiary numbers;
9. Account numbers;
10. Certificate or license numbers;
11. Vehicle identifiers and license plate numbers;
12. Device identifiers and serial numbers;
13. URLs associated with an individual;
14. IP addresses;
15. Biometric identifiers; and
16. Full-face photographs and any comparable images.

Point 16 is particularly relevant for neuroimaging studies as the reconstructed 3D images of an anatomical scan can be used to render the facial features of a participant. One approach to resolving this issue is to apply a defacing algorithm to the imaging data to obscure any recognizable facial features (Milchenko & Marcus, 2012). To unlink the original with the de-identified data, the name of each participant in a study is replaced with a unique identification number that does not reveal any PHI. NIMH provides a tool for generating such Globally Unique Identifier (GUID)18. Finally, NIMH requires that Institutional Review Boards (IRB) contain specific language regarding the informed consent of sharing participant's data in a data repository, for which NIMH has created a template19. Templates are also provided by the Data Management Planning Tool (DMPTool)20 that guides researchers through the process of creating a plan that will comply with their funding institution's requirements.

2.3 Data Repositories

The growing availability and accessibility of shared neuropsychological and neuroimaging data is due, in part, to centralized data repositories (Gorgolewski et al., 2015; Hall et al., 2012; D. S. Marcus et al., 2011; Poldrack et al., 2013; Toga et al., 2010). Neurocognitive data repositories not only provide a simple mechanism for archiving and distributing data but also can empower researchers to reproduce findings in the primary literature (Pernet &

[18] https://ndar.nih.gov/standards.html#GUID
[19] http://rdocdb.nimh.nih.gov/wp-content/uploads/RDoCdb_InfCon_Language7141.docx
[20] https://dmptool.org/

Poline, 2015) by providing access to neurocognitive data that can be reanalyzed, used for teaching data analysis methodologies, or to advance hypothesis and data-driven discovery (Biswal et al., 2010). The primary function of a data repository is to provide users with a mechanism to retrieve datasets. From a neuroinformatics perspective, the simplest form of data repository is a collection of datasets that can be downloaded in bulk; however, information systems can improve the experience for users searching for specific datasets and contributors wanting to upload data. For example, a query interface that enables a user to browse for and download neurocognitive data from subjects that completed an anatomical MRI between the ages of 18–25 is an improvement over the bulk download scenario. The functionality of these systems is complementary to the data management tools discussed in Section 2.1 in that the focus is to curate large amounts of heterogeneous data (Hall et al., 2012) rather than support day to day research operations; however, EDCMS may also support data repository efforts (e.g., COINS, IDA, and XNAT). This section focuses on population and modality specific neurocognitive data repositories (i.e., databases that distribute medical images and neuropsychological test scores) using subset of representative data sharing projects selected from the Neuroimaging Informatics Tools and Resources Clearinghouse (Kennedy et al., 2015) and Neuroscience Information Framework (Gardner et al., 2008) resource registries. These data repositories can roughly be divided into two types:

- **Population Specific Repositories**: These data repositories are specific to a given population that typically includes a core set of measures sensitive to primary hypotheses. They collect data as part of a consortium or collaboration and then distributes the data to the research community (Mennes et al., 2013; Van Essen et al., 2012). Examples are the National Alzheimer's Coordinating Center (NACC) (Beekly et al., 2007) that accommodates data from the Alzheimer's Disease Research Centers (Morris et al., 2006), Alzheimer's Disease Neuroimaging Initiative (ADNI) (Toga et al., 2010), Autism Brain Imaging Data Exchange (ABIDE) (Di Martino et al., 2014), National Database for Autism Research (NDAR) (Hall et al., 2012), Childhood and Adolescent NeuroDevelopment Initiative (CANDIShare) (Kennedy et al., 2012), and Pediatric Imaging Neurocognition and Genetics (PING) (Fjell et al., 2012). Uploading data to these population specific repositories by participating in a consortium has the advantage of creating curated and integrated information tuned towards answering questions of interest in a given research community.
- **Modality Specific Repositories:** The goal of a modality specific data repository is to simplify sharing specific types of information, such as imaging data. Examples repositories are OpenfMRI (Poldrack et al., 2013) that provides submission guidelines for primary imaging data for fMRI studies, NeuroVault (Gorgolewski et al., 2015) that focuses on sharing derived imaging data from fMRI (i.e., unthresholded statistical parametric maps (Friston et al., 1994)), and Neuroimaging Informatics Resource Technology Clearinghouse Image Repository (NITRC-IR) (Kennedy et al., 2015) that provides a public XNAT database to host shared neuroimaging studies. Uploading data to these repositories allows neuroscientists to comply with data sharing requirements

when their project may not be tied to a specific consortium focused on a given population.

Each of these repositories constitute as a data sharing initiative. The policies for disseminating data from these repositories generally require accepting a data use agreement that limits who and how a dataset can be used. Examples of data use agreements are *Open Access Data,* which is typically de-identified data that can be easily downloaded after filling out a form online (Di Martino et al., 2014), and *Restricted Data,* which may require institutional review board approval, demonstration of qualified research credentials, or review of an application before gaining access to the data (Hall et al., 2012). Table 2 contains a listing of data repositories and distributors including summary information on the categories listed above.

This section presented Neuroinformatics resources that ranged from tools designed to manage neuropsychology data in a single lab to data repositories that are used to distribute data from thousands of research participants throughout the scientific community. In the context of managing a study, these resources can be leveraged to streamline the collection of high quality data and enhance research productivity by minimizing data management activities. The execution of data management plans are needed to insure the patients' privacy and neuroinformatics data repositories can simplify the submission process and shoulder the burden of data storage and longevity.

3. Examples of Neuroimaging Studies with EDCMS

We now present two scenarios for applying the EDCMS described in Section 2.1 to neuroimaging studies. For each scenario, we specify the study design, the resulting requirements for data capture and management, and a neuroinformatics approaches that meets those requirements. Specifically, Scenario A (Section 3.1) describes a cross-sectional brain-behavior study with EDCMS being optional and Scenario B (Section 3.2) presents a case study on a multi-site, longitudinal study with complex requirements that necessitates an EDCMS. These scenarios are meant to highlight the challenges encountered when executing studies of different scales and to gauge when an EDCMS may or may not help to overcome these challenges.

3.1. Scenario A: Cross-Sectional Study

We now present a hypothetical scenario of a typical brain behavioral study to highlight basic neuroinformatics requirements and approaches used to address issues in data capture, management, and sharing. In this scenario, a lab is conducting a cross-sectional study that is funded by the NIMH to examine the relationship between neuroanatomical volumes extracted from MRIs and measures from the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 1989). The data are collected from a population of participants with Autism Spectrum Disorder (n=40) and healthy controls (n=40). These two sample sets are age and sex matched. The study takes place in the medical school that the lab is part of. The school has technical personnel on staff for implementing scanner sequences and acquiring imaging data from the participants. The lab itself consists of a principle investigator, two graduate students, and an undergraduate research assistant. Graduate student A performs the

basic image processing of the anatomical MRI, graduate student B is trained in administering the ADOS, and the research assistant performs the ADOS scoring and data entry. Specifically, student B administers the ADOS modules and records the behavioral observations via paper and pencil form. The research assistant uses these records to manually score and enter the ADOS data into a spreadsheet, whose variable names were defined by the principal investigator. After a participant is scanned, graduate student A obtains a USB thumb drive with the anatomical MRI data. She transfers the data from the USB thumb drive to the file system of a lab workstation and process the imaging data via FreeSurfer (Fischl et al., 2002) to extract regional brain volumes. Once data acquisition and processing of the study is completed, the principal investigator performs the hypothesis driven analyses and drafts a manuscript for publication. Upon publication, this NIMH funded study is required to be uploaded and shared through NDAR.

Study Requirements—To successfully complete this study, the electronic data capture and management requirements are fairly minimal. Given the relatively small sample size, it is quite reasonable to first record observations via paper and pencil and then captures those observations electronically via data entry into a spreadsheet. However, the lack of double data entry can introduce errors and may impact data analysis (Day, Fayers, & Harvey, 1998). For the imaging data, a directory structure will need to be defined to store the subject data and neuroinformatics tools will need to be installed for converting and processing the data. Finally, the data (e.g. spread sheet) will need to be prepared according to the NDAR guidelines so that it can be uploaded to the corresponding repository.

Neuroinformatics Approach—Given the simple design of the study in this scenario, an informatics evaluation of the requirements for this study would not warrant manually installing an EDCMS; however, there are improvements that can be explored without heavy overhead. First, the investigator could explore the resources available at their institution to identify if an EDCMS is already hosted. Today, many medical schools provide EDCMS as a service that is funded by Clinical and Translational Science Awards (Bernstam et al., 2009). If the EDCMS does not support imaging data, a directory structure (such as the Brain Imaging Data Structure (BIDS)28) on a file system will be adequate. If the lab does not want to use an EDCMS for the ADOS data, spreadsheet software will suffice but the investigator may want to consider using the data collection forms provided by the PhenX Toolkit (Stover, Harlan, Hammond, Hendershot, & Hamilton, 2010). The lab should also adopt the standard data dictionaries for variables that are provided by NDAR. Once the results are published, this will ease uploading the data to a repository as mandated by the NIMH.

3.2 Scenario B: Multisite Longitudinal Study

The second case study is based on our own work where data management tools are deployed to conduct research on neurodevelopment in adolescence. This study is motivated by the observation that alcohol and marijuana remain the most commonly used central nervous system-active substances in the teen years (Johnston, OMalley, Miech, Bachman, & Schulenberg, 2015). To study the influence of adolescent alcohol and marijuana abuse on

[28] http://bids.neuroimaging.io

neurodevelopment, the National Consortium on Alcohol and Neurodevelopment in Adolescence (NCANDA) is a multisite, longitudinal, study that recruited 831 participants (ranging from 12–22 years old) across five data collection sites nationwide (Brown et al., In press.).

Each of the five data collection sites carried out the same core assessment and worked in pairs to conduct additional studies (e.g., overnight sleep evaluation and recovery during monitored abstinence). The 831 study participants completed a core data acquisition protocol at baseline and will complete three annual follow-ups, each of which include a neuropsychological (NP) test battery, neuroimaging session (MRI, DTI, and rsfMRI), bio-samples for genetic analysis, a comprehensive assessment of substance use, psychiatric symptoms and diagnoses, functioning in major life domains, and one parent of each youth completes an interview on the youth and family environment. The NP test battery assesses seven major functional domains including: general intelligence; executive functions; emotion regulation; multimodal and multiple component mnemonic processes; visuospatial abilities; basic visual acuity and color perception; and motor skills of eye-hand coordination, speed, and postural stability. In addition, a mid-year phone interview is conducted between each visit to track substance use. Upon completing data collection, the dataset is expected to reach approximately 6TB of primary data and nearly 20TB of derived data from neuroimaging analyses. In the sections below, we present an overview of the study requirements that needed to be addressed and the neuroinformatics approaches we used to implement a framework that enabled us to collect data rapidly, maintain quality control, and streamline data processing (Rohlfing, Cummins, Henthorn, Chu, & Nichols, 2013).

Study Requirements—To realize the longitudinal experimental design of NCANDA, it was necessary to establish a framework capable of meeting the requirements to capture, integrate, and process multimodal data from five data collection sites. To be economical, we wanted to design a framework consisting of freely available data management tools. The guiding principles in the evaluation of those tools were 1) an active and supportive mailing list, 2) intuitive GUI with training materials for research staff, 3) support for customization for longitudinal data acquisition, and 4) the ability to automate tasks programmatically (e.g., quality control checks, test scoring) using an API. After evaluating available medical imaging data management systems and electronic data capture systems (see Section 2.1), we chose a solution coupling XNAT, which is targeted towards imaging studies, with REDCap, a data management system addressing the needs of the study with respect to NP test data. Both systems met the evaluation criteria and tested well with research staff during an initial evaluation. At the time this framework was developed in 2012, no single system solutions existed that fulfilled our evaluation criteria.

Neuroinformatics Approach—Building upon XNAT and REDCap, we designed a framework (Figure 1) that automated electronic data capture, management, harmonization, quality control, analysis, and distribution across the five data collection sites of the NCANDA consortium (Rohlfing et al., 2013). Specifically, the NCANDA sites collected the non-imaging data via the University of Pennsylvania Web-based Computerized Neurocognitive Battery (WebCNP) (Gur et al., 2010), LimeSurvey29, Blaise30, ePrime31,

and REDCap. Test scores not collected directly through entry forms in REDCap were automatically transformed into a REDCap compliant format and uploaded from the laptops used for data capturing at the collection sites to the REDCap server hosted by the NCANDA Data Analysis Component at SRI International via encrypted connections to Subversion32, a secure and persistent data uploading system. Imaging data was first uploaded from the site-specific PACS to the XNAT server hosted at SRI International. All imaging data underwent a quality control that included automatic test scoring, range validation, and a neuroradiologist report for incidental imaging findings. Finally, the outcome of the quality control was uploaded into REDCap and merged with the corresponding non-imaging data for each session. Any updates to information in the REDCap database automatically triggered the generation of reports regarding data integrity. Identified issues were resolved with site consultation for scoring irregularities, incorrectly entered IDs, visit dates, and any data that were not uploaded properly. Once data passed the initial quality control, the data was processed in further analyses and backed it up via Amazon Web Services (AWS).

To distribute the collected data with the NCANDA consortium, the platform has an integrated data release mechanism. To create the release, all entries in REDCap are manually checked one more time for entry errors. Entries passing this quality control are immediately locked in the database (i.e., changes to these records required prior approval by the investigators of the NCANDA Data Analysis component at SRI International). With respect to incorrect or questionable entries, a data manager at SRI International resolves the issues by contacting the collection sites and locks the record once the error is resolved. After all entries requested for the data release are locked, the data is provided to the members of the NCANDA consortium via a set of comma-separated-value (CSV) files exported from REDCap with corresponding data dictionaries for each data element. Plans for sharing the data with the broader research community include technology to facilitate interoperability with neuroinformatics resources, such as the Neuroimaging Data Model (NIDM) standard for data exchange (Keator et al., 2013), the Cognitive Atlas ontology (Poldrack et al., 2011) for data annotation, and the Neuroimaging Informatics Resource Technology Clearinghouse Image Repository (NITRC) (Kennedy et al., 2015) and OpenfMRI (Poldrack et al., 2013) data repositories. The resulting organization of this data set would then align with the approach proposed by the Research Domain Criteria (RDoC) Initiative (Insel et al., 2010), where data can be explored at different levels of analysis (e.g., from circuit-level to family environment) and by broad domains of function (e.g., cognitive systems or working memory).

4. Conclusions

This manuscript provided an introduction to electronic data capture and management tools, data management plan, and data repositories to facilitate compliance of neurocognitive studies with data sharing mandates of funding agencies and publishers and to decrease the setup time and improve quality control of studies, and streamline the process of

[29] https://www.limesurvey.org
[30] http://www.blaise.com
[31] http://www.pstnet.com/eprime.cfm
[32] https://subversion.apache.org

harmonizing, curating, and sharing data across data repositories. Today, many researchers see freely sharing data as a key scientific resource with the goal of maximizing the knowledge gleaned from neuroimaging studies. However, sharing neurocognitive data is generally viewed as a resource intensive activity as it requires curating data so that it is meaningful to the research community. The Neuroinformatics tools, data management plans, and repositories reviewed here aim to reduce this burden. Furthermore, they enable large-scale studies as highlighted by one of the neuroimaging study scenarios. Finally, readers wanting to gain a more complete view of this topic should visit resource registries (Belleau, Nolin, Tourigny, Rigault, & Morissette, 2008; Ferguson et al., 2014; Gardner et al., 2008; Kennedy et al., 2015; Stover et al., 2010), which catalogue shared data repositories (Gardner et al., 2008), data analysis software (Kennedy et al., 2015), ontology resources (Fox et al., 2005; Larson & Martone, 2013; B. N. Nichols et al., 2014; Poldrack et al., 2011), and utilities for simplifying system configuration (Stover et al., 2010); (Gershon et al., 2010).

Acknowledgments

This work was supported by the U.S. National Institute on Alcohol Abuse and Alcoholism (NIAAA) (U01 AA021697, R01 AA005965, R01 AA012388, U01 AA013521, U01 AA017347, U01 AA017923). It was also supported by the Creative and Novel Ideas in HIV Research Program (CNIHR) through a supplement to the University of California at San Francisco (UCSF) Center For AIDS Research funding (P30 AI027763). This funding was made possible by collaborative efforts of the Office of AIDS Research, the National Institutes of Allergies and Infectious Diseases, and the International AIDS Society.

References

Beekly D, Ramos E, Lee W, Deitrich W, Jacka M, Wu J, et al. The National Alzheimer's Coordinating Center (NACC) database: The uniform data set. Alzheimer Disease & Associated Disorders. 2007; 21(3):249. [PubMed: 17804958]

Belleau F, Nolin M-A, Tourigny N, Rigault P, Morissette J. Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. Journal of Biomedical Informatics. 2008; 41(5) http://doi.org/10.1016/j.jbi.2008.03.004.

Bernstam EV, Hersh WR, Johnson SB, Chute CG, Nguyen H, Sim I, et al. Synergies and distinctions between computational disciplines in biomedical research: perspective from the Clinical andTranslational Science Award programs. Presented at the Academic medicine: journal of the Association of American Medical Colleges. 2009; 84:964–970. http://doi.org/10.1097/ACM.0b013e3181a8144d.

Biswal BB, Mennes M, Zuo XN, Gohel S, Kelly C, Smith SM, et al. Toward discovery science of human brain function. Proceedings of the National Academy of Sciences. 2010; 107(10):4734–4739. http://doi.org/10.1073/pnas.0911855107.

Bjaalie JG, Grillner S. Global neuroinformatics: the International Neuroinformatics Coordinating Facility. Journal of Neuroscience. 2007; 27(14):3613–3615. http://doi.org/10.1523/JNEUROSCI.0558-07.2007. [PubMed: 17409224]

Bloom, T., Ganley, E., Winker, M. Data access for the open access literature: PLOS's data policy. PLoS Biology. 2014. http://doi.org/10.1371/journal.pmed.1001607

Book GA, Anderson BM, Stevens MC, Glahn DC, Assaf M, Pearlson GD. Neuroinformatics Database (NiDB) - A Modular, Portable Database for the Storage, Analysis, and Sharing of Neuroimaging Data. Neuroinformatics. 2013; 11(4):495–505. http://doi.org/10.1007/s12021-013-9194-1. [PubMed: 23912507]

Breeze JL, Poline JB, Kennedy DN. Data sharing and publishing in the field of neuroimaging. GigaScience. 2012; 1(1):9. http://doi.org/10.1186/2047-217X-1-9. [PubMed: 23587272]

Brinkley JE, Rosse C. Imaging and the Human Brain Project: a review. Methods of Information in Medicine. 2002; 41(4):245–260. [PubMed: 12425235]

Brown SA, Brumback T, Tomlinson K, Cummins K, Thompson WK, Nagel BJ, et al. The National Consortium on Alcohol and NeuroDevelopment in Adolescence (NCANDA): A multi-site study of adolescent development and substance use. Journal of Studies on Alcohol and Drugs. In press.

Buckow, K., Quade, M., Rienhoff, O., Nussbeck, SY. Changing requirements and resulting needs for IT-infrastructure for longitudinal research in the neurosciences. Neuroscience Research. 2014. http://doi.org/10.1016/j.neures.2014.08.005

Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR. Power failure: why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience. 2013; 14(5):365–376. http://doi.org/10.1038/nrn3475. [PubMed: 23571845]

Caspers S, Zilles K, Laird AR, Eickhoff SB. ALE meta-analysis of action observation and imitation in the human brain. NeuroImage. 2010; 50(3):1148–1167. http://doi.org/10.1016/j.neuroimage.2009.12.112. [PubMed: 20056149]

Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. Nature. 2014; 505(7485):612–613. [PubMed: 24482835]

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research, an International Journal. 1996; 29(3):162–173.

Cozzarelli NR. UPSIDE: Uniform principle for sharing integral data and materials expeditiously. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101(11):3721–3722. http://doi.org/10.1073/pnas.0400437101. [PubMed: 15026576]

D'Esposito M. Letter from the Special Issue Editor. Journal of Cognitive Neuroscience. 2000; 12(supplement 2):1–1. http://doi.org/10.1162/089892900563966.

Das S, Zijdenbos AP, Harlap J, Vins D, Evans AC. LORIS: a web-based data management system for multi-center studies. Frontiers in Neuroinformatics. 2011; 5:37. http://doi.org/10.3389/fninf.2011.00037. [PubMed: 22319489]

David SP, Ware JJ, Chu IM, Loftus PD, Fusar-Poli P, Radua J, et al. Potential reporting bias in fMRI studies of the brain. PLoS ONE. 2013; 8(7):e70104. http://doi.org/10.1371/journal.pone.0070104. [PubMed: 23936149]

Day S, Fayers P, Harvey D. Double data entry: what value, what price? Controlled Clinical Trials. 1998; 19(1):15–24. [PubMed: 9492966]

Di Martino A, Yan CG, Li Q, Denio E, Castellanos FX, Alaerts K, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular Psychiatry. 2014; 19(6):659–667. http://doi.org/10.1038/mp.2013.78. [PubMed: 23774715]

Dinov I. Efficient, distributed and interactive neuroimaging data analysis using the LONI Pipeline. Frontiers in Neuroinformatics. 2009; 3:1–10. http://doi.org/10.3389/neuro.11.022.2009. [PubMed: 19198661]

Ferguson AR, Nielson JL, Cragin MH, Bandrowski AE, Martone ME. Big data from small data: data-sharing in the "long tail" of neuroscience. Nature Neuroscience. 2014; 17(11):1442–1447. http://doi.org/10.1038/nn.3838. [PubMed: 25349910]

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002; 33(3):341–355. [PubMed: 11832223]

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, et al. Automatically parcellating the human cerebral cortex. Cerebral Cortex (New York, NY: 1991). 2004; 14(1):11–22.

Fjell AM, Walhovd KB, Brown TT, Kuperman JM, Chung Y, Hagler DJ, et al. Multimodal imaging of the self-regulating developing brain. Proceedings of the National Academy of Sciences. 2012; 109(48):19620–19625. http://doi.org/10.1073/pnas.1208243109.

Fox PT, Lancaster JL. Opinion: Mapping context and content: the BrainMap model. Nature Reviews Neuroscience. 2002; 3(4):319–321. http://doi.org/10.1038/nrn789. [PubMed: 11967563]

Fox P, Fox PT, Laird AR, Laird A, Fox SP, Fox S, et al. BrainMap taxonomy of experimental design: description and evaluation. Human Brain Mapping. 2005; 25(1):185–198. http://doi.org/10.1002/hbm.20141. [PubMed: 15846810]

Franklin JD, Guidry A, Brinkley JF. A partnership approach for Electronic Data Capture in small-scale clinical trials. Journal of Biomedical Informatics. 2011; 44(Suppl 1):S103–8. http://doi.org/10.1016/j.jbi.2011.05.008. [PubMed: 21651992]

Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RS. Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping. 1994; 2(4):189–210.

Gadde S, Aucoin N, Grethe JS, Keator DB, Marcus DS, Pieper S. FBIRN, MBIRN, BIRN-CC. XCEDE: an extensible schema for biomedical data. Neuroinformatics. 2012; 10(1):19–32. http://doi.org/10.1007/s12021-011-9119-9. [PubMed: 21479735]

Gardner D, Akil H, Ascoli G, Bowden D, Bug W, Donohue D, et al. The neuroscience information framework: a data and knowledge environment for neuroscience. Neuroinformatics. 2008; 6(3): 149–160. [PubMed: 18946742]

Gershon RC, Gershon RC, Cella D, Cella D, Fox NA, Fox NA, et al. Assessment of neurological and behavioural function: the NIH Toolbox. Lancet Neurology. 2010; 9(2):138–139. http://doi.org/10.1016/S1474-4422(09)70335-7. [PubMed: 20129161]

Gomez-Marin A, Paton JJ, Kampff AR, Costa RM, Mainen ZF. Big behavioral data: psychology, ethology and the foundations of neuroscience. Nature Neuroscience. 2014; 17(11):1455–1462. http://doi.org/10.1038/nn.3812. [PubMed: 25349912]

Gorgolewski, KJ., Varoquaux, G., Rivera, G., Schwartz, Y., Sochat, VV., Ghosh, SS., et al. NeuroVault.org: A repository for sharing unthresholded statistical maps, parcellations, and atlases of the human brain. NeuroImage. 2015. http://doi.org/10.1016/j.neuroimage.2015.04.016

Greenes, R., Brinkley, J. Biomedical and Health Informatics: Computer Applications in Healthcare. 2006. Imaging Systems in Radiology; p. 1-34.

Gur RC, Richard J, Hughett P, Calkins ME, Macy L, Bilker WB, et al. A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. Journal of Neuroscience Methods. 2010; 187(2):254–262. http://doi.org/10.1016/j.jneumeth.2009.11.017. [PubMed: 19945485]

Haak, D., Page, C-E., Reinartz, S., Krüger, T., Deserno, TM. DICOM for Clinical Research: PACS-Integrated Electronic Data Capture in Multi-Center Trials; Journal of Digital Imaging. 2015. p. 1-9.http://doi.org/10.1007/s10278-015-9802-8

Hall D, Huerta MF, McAuliffe MJ, Farber GK. Sharing Heterogeneous Data: The National Database for Autism Research. Neuroinformatics. 2012; 10(4):331–339. http://doi.org/10.1007/s12021-012-9151-4. [PubMed: 22622767]

Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics. 2009; 42(2):377–381. http://doi.org/10.1016/j.jbi.2008.08.010. [PubMed: 18929686]

Helmer KG, Ambite JL, Ambite JL, Ames J, Ames J, Ananthakrishnan R, et al. Enabling collaborative research using the Biomedical Informatics Research Network (BIRN). Journal of the American Medical Informatics Association. 2011; 18(4):416–422. http://doi.org/10.1136/amiajnl-2010-000032. [PubMed: 21515543]

Howe D, Costanzo M, Fey P, Gojobori T, Hannick L, Hide W, et al. Big data: The future of biocuration. Nature. 2008; 455(7209):47–50. http://doi.org/10.1038/455047a. [PubMed: 18769432]

Hudson KL, Collins FS. Sharing and reporting the results of clinical trials. Jama. 2015; 313(4):355–356. http://doi.org/10.1001/jama.2014.10716. [PubMed: 25408371]

Huerta MF, Koslow SH. Neuroinformatics: Opportunities across Disciplinary and National Borders. NeuroImage. 1996; 4(3):S4–S6. http://doi.org/10.1006/nimg.1996.0040. [PubMed: 9345515]

Hussein R, Engelmann U, Schroeter A, Meinzer H. DICOM Structured Reporting. Radiographics: a Review Publication of the Radiological Society of North America, Inc. 2004; 24(3):897.

Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. The American Journal of Psychiatry. 2010; 167(7):748–751. http://doi.org/10.1176/appi.ajp.2010.09091379. [PubMed: 20595427]

Ioannidis JPA. Why Most Published Research Findings Are False. PLoS Medicine. 2005; 2(8):e124. http://doi.org/10.1371/journal.pmed.0020124. [PubMed: 16060722]

Ioannidis JPA. Excess significance bias in the literature on brain volume abnormalities. Archives of General Psychiatry. 2011; 68(8):773–780. http://doi.org/10.1001/archgenpsychiatry.2011.28. [PubMed: 21464342]

Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging: JMRI. 2008; 27(4):685–691. http://doi.org/10.1002/jmri.21049. [PubMed: 18302232]

Johnston, LD., OMalley, PM., Miech, RA., Bachman, JG., Schulenberg, JE. Monitoring the Future national survey results on drug use: 1975–2014: Overview, key findings on adolescent drug use. 2015.

Kane RL, Kay GG. Computerized assessment in neuropsychology: A review of tests and test batteries. Neuropsychology Review. 1992; 3(1):1–117. http://doi.org/10.1007/BF01108787. [PubMed: 1300218]

Keator DB, Helmer K, Steffener J, Turner JA, Van Erp TG, Gadde S, et al. Towards structured sharing of raw and derived neuroimaging data across existing resources. NeuroImage. 2013; 82:647–661. http://doi.org/10.1016/j.neuroimage.2013.05.094. [PubMed: 23727024]

Keator DB, Wei D, Gadde S, Bockholt J, Grethe JS, Marcus D, et al. Derived Data Storage and Exchange Workflow for Large-Scale Neuroimaging Analyses on the BIRN Grid. Frontiers in Neuroinformatics. 2009; 3:30. http://doi.org/10.3389/neuro.11.030.2009. [PubMed: 19826494]

Kennedy DN, Haselgrove C, Hodge SM, Rane PS, Rane PS, Makris N, Frazier JA. CANDIShare: a resource for pediatric neuroimaging data. Neuroinformatics. 2012; 10(3):319–322. http://doi.org/10.1007/s12021-011-9133-y. [PubMed: 22006352]

Kennedy, DN., Haselgrove, C., Riehl, J., Preuss, N., Buccigrossi, R. The Three NITRCs: A Guide to Neuroimaging Neuroinformatics Resources; Neuroinformatics. 2015. p. 1-4.http://doi.org/10.1007/s12021-015-9263-8

Kulikowski CA, Shortliffe EH, Currie LM, Elkin PL, Hunter LE, Johnson TR, et al. AMIA Board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline. Journal of the American Medical Informatics Association. 2012; 19(6):931–938. http://doi.org/10.1136/amiajnl-2012-001053. [PubMed: 22683918]

Laird AR, Lancaster JL, Fox PT. BrainMap: the social evolution of a human brain mapping database. Neuroinformatics. 2005; 3(1):65–78. [PubMed: 15897617]

Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, et al. Automated Talairach atlas labels for functional brain mapping. Human Brain Mapping. 2000; 10(3):120–131. http://doi.org/10.1002/1097-0193(200007)10:3<120::AID-HBM30>3.0.CO;2-8. [PubMed: 10912591]

Larson SD, Martone ME. NeuroLex.org: an online framework for neuroscience knowledge. Frontiers in Neuroinformatics. 2013; 7:18. http://doi.org/10.3389/fninf.2013.00018. [PubMed: 24009581]

Lord C, Rutter M, Goode S, Heemsbergen J, Jordan H, Mawhood L, Schopler E. Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. Journal of Autism and Developmental Disorders. 1989; 19(2):185–212. [PubMed: 2745388]

Marcus DS, Harwell J, Olsen T, Hodge M, Glasser MF, Prior F, et al. Informatics and data mining tools and strategies for the human connectome project. Frontiers in Neuroinformatics. 2011; 5:4. http://doi.org/10.3389/fninf.2011.00004. [PubMed: 21743807]

Marcus D, Olsen T, Ramaratnam M, Buckner R. The extensible neuroimaging archive toolkit. Neuroinformatics. 2007; 5(1):11–33. [PubMed: 17426351]

Meier MH, Caspi A, Ambler A, Harrington H, Houts R, Keefe RSE, et al. Persistent cannabis users show neuropsychological decline from childhood to midlife. Proceedings of the National Academy of Sciences. 2012; 109(40):E2657–64. http://doi.org/10.1073/pnas.1206820109.

Mennes M, Biswal BB, Castellanos FX, Milham MP. Making data sharing work: the FCP/INDI experience. NeuroImage. 2013; 82:683–691. http://doi.org/10.1016/j.neuroimage.2012.10.064. [PubMed: 23123682]

Milchenko, M., Marcus, D. Obscuring Surface Anatomy in Volumetric Imaging Data. Neuroinformatics. 2012. http://doi.org/10.1007/s12021-012-9160-3

Morris J, Weintraub S, Chui H, Cummings J, DeCarli C, Ferris S, et al. The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. Alzheimer Disease & Associated Disorders. 2006; 20(4):210. [PubMed: 17132964]

Nichols BN, Mejino JL Jr, Detwiler L, Nilsen TT, Martone ME, Turner JA, et al. Neuroanatomical domain of the foundational model of anatomy ontology. J Biomedical Semantics. 2014; 5:1. http://doi.org/10.1186/2041-1480-5-1.

Pernet C, Poline JB. Improving functional magnetic resonance imaging reproducibility. GigaScience. 2015; 4(1):15. http://doi.org/10.1186/s13742-015-0055-8. [PubMed: 25830019]

Poldrack RA, Gorgolewski KJ. Making big data open: data sharing in neuroimaging. Nature Neuroscience. 2014; 17(11):1510–1517. http://doi.org/10.1038/nn.3818. [PubMed: 25349916]

Poldrack RA, Barch DM, Mitchell JP, Wager TD, Wagner AD, Devlin JT, et al. Toward open sharing of task-based fMRI data: the OpenfMRI project. Frontiers in Neuroinformatics. 2013; 7:12. http://doi.org/10.3389/fninf.2013.00012. [PubMed: 23847528]

Poldrack RA, Kittur A, Kalar D, Miller E, Seppa C, Gil Y, et al. The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. Frontiers in Neuroinformatics. 2011; 5:17. http://doi.org/10.3389/fninf.2011.00017. [PubMed: 21922006]

Poline JB, Breeze JL, Ghosh S, Gorgolewski K, Halchenko YO, Hanke M, et al. Data sharing in neuroimaging research. Frontiers in Neuroinformatics. 2012; 6:9–9. http://doi.org/10.3389/fninf.2012.00009. [PubMed: 22493576]

Rohlfing, T., Cummins, K., Henthorn, T., Chu, W., Nichols, BN. N-CANDA data integration: anatomy of an asynchronous infrastructure for multi-site, multi-instrument longitudinal data capture. Journal of the American Medical Informatics Association. 2013. amiajnl–2013–002367. http://doi.org/10.1136/amiajnl-2013-002367

Salimi-Khorshidi G, Smith SM, Keltner JR, Wager TD, Nichols TE. Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. NeuroImage. 2009; 45(3):810–823. http://doi.org/10.1016/j.neuroimage.2008.12.039. [PubMed: 19166944]

Scott A, Courtney W, Wood D, la Garza De R, Lane S, King M, et al. COINS: An Innovative Informatics and Neuroimaging Tool Suite Built for Large Heterogeneous Datasets. Frontiers in Neuroinformatics. 2011; 5:33–33. http://doi.org/10.3389/fninf.2011.00033. [PubMed: 22275896]

Shepherd GM. Supporting databases for neuroscience research. Journal of Neuroscience. 2002; 22(5): 1497. [PubMed: 11880479]

Shepherd GM, Mirsky JS, Healy MD, Singer MS, Skoufos E, Hines MS, et al. The Human Brain Project: neuroinformatics tools for integrating, searching and modeling multidisciplinary neuroscience data. Trends in Neurosciences. 1998; 21(11):460–468. [PubMed: 9829685]

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage. 2004; 23(Suppl 1):S208–19. http://doi.org/10.1016/j.neuroimage.2004.07.051. [PubMed: 15501092]

Stover PJ, Harlan WR, Hammond JA, Hendershot T, Hamilton CM. PhenX: a toolkit for interdisciplinary genetics research. Current Opinion in Lipidology. 2010; 21(2):136–140. http://doi.org/10.1097/MOL.0b013e3283377395. [PubMed: 20154612]

Thompson PM, Stein JL, Medland SE, Hibar DP, Vasquez AA, Rentería ME, et al. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain Imaging and Behavior. 2014; 8(2):153–182. http://doi.org/10.1007/s11682-013-9269-5. [PubMed: 24399358]

Toga AW, Crawford KL. Alzheimer's Disease Neuroimaging Initiative. The informatics core of the Alzheimer's Disease Neuroimaging Initiative. Alzheimer"S & Dementia: the Journal of the Alzheimer"S Association. 2010; 6(3):247–256. http://doi.org/10.1016/j.jalz.2010.03.001.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage. 2002; 15(1):273–289. [PubMed: 11771995]

Van Essen DC, Ugurbil K, Auerbach E, Barch D, Behrens TEJ, Bucholz R, et al. The Human Connectome Project: a data acquisition perspective. NeuroImage. 2012; 62(4):2222–2231. http://doi.org/10.1016/j.neuroimage.2012.02.018. [PubMed: 22366334]

Van Horn JD, Toga AW. Is it time to re-prioritize neuroimaging databases and digital repositories? NeuroImage. 2009; 47(4):1720–1734. http://doi.org/10.1016/j.neuroimage.2009.03.086. [PubMed: 19371790]

Van Horn JD, Toga AW. Human neuroimaging as a "Big Data" science. Brain Imaging and Behavior. 2014; 8(2):323–331. http://doi.org/10.1007/s11682-013-9255-y. [PubMed: 24113873]

Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. Large-scale automated synthesis of human functional neuroimaging data. Nature Methods. 2011; 8(8):665–670. http://doi.org/10.1038/nmeth.1635. [PubMed: 21706013]

Young MP, Scannell JW. Brain structure-function relationships: advances from neuroinformatics. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. 2000; 355(1393):3–6. http://doi.org/10.1098/rstb.2000.0545. [PubMed: 10703040]

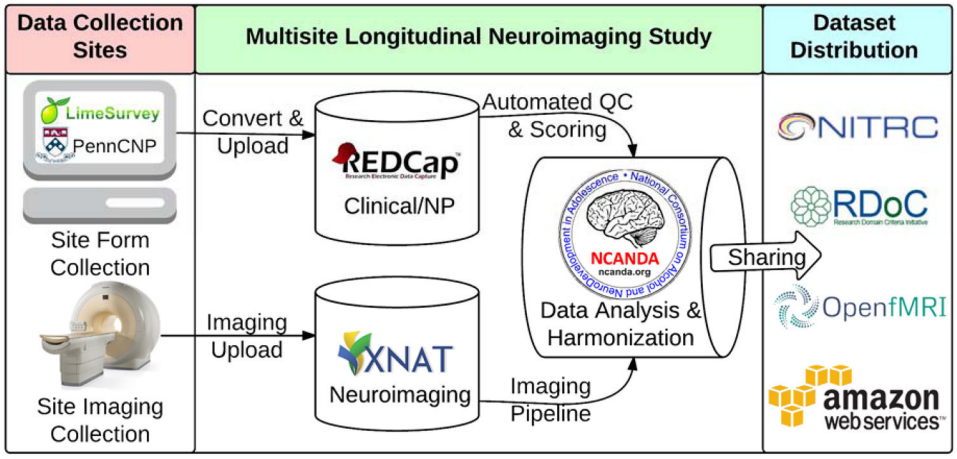

Figure 1.

Scenario B: Multisite Longitudinal Study Framework. Each of the NCANDA Data Collection Sites (left) collects form-based clinical and neuropsychological test data and multi-modal neuroimaging data. The form-based data is programmatically converted into a compliant format and automatically uploaded to a central REDCap server using the API. Imaging data is transmitted to a central XNAT server manually or automatically using the DICOM network protocol. Data from both REDCap and XNAT undergo quality control (QC) procedures before analysis and harmonization (center). After this stage, the processed data can be distributed to the broader community (right).

Table 1

A summary of the Electronic Data Capture and Management Tools described in Section 2.1

All systems provide a mailing list, user documentation, demos, and an issue tracker with links available from their website. For each system, the table also lists the institution primarily developing the system, the latest release year of the software, the type of software license, the availability of an Application Programming Interface (API), if the software is installable, and if there is an option to have the software hosted as a service (i.e., no installation required).

System	Organization	Latest Release	License	API	Installable	Hosted
COINS[2]	Mind Research Network	2015	Custom (closed)	No	No	Yes
HID[3]	Biomedical Informatics Research Network	2010	BSD	No	Yes	No
IDA[4]	University of Southern California	2015	Custom (closed)	No	No	Yes
LORIS[5]	McGill University	2015	GNU GPLv3	No	Yes	No
NiDB[6]	Hartford Hospital	2015	GNU GPLv3	No	Yes	Yes
REDCap[7]	Vanderbilt University	2015	Custom (closed)	Yes	Yes	No
XNAT[8]	Washington University, St. Louis	2014	Custom (open)	Yes	Yes	Yes

[2]Collaborative Neuroinformatics Suite: http://coins.mrn.org
[3]Human Imaging Database: http://www.nitrc.org/projects/hid
[4]Image Data Archive: http://ida.loni.usc.edu
[5]Longitudinal Online Research and Imaging System: http://mcin-cnim.ca/neuroimagingtechnologies/loris
[6]Neuroinformatics Database: http://nidb.sourceforge.net
[7]Research Electronic Data Capture: http://www.project-redcap.org
[8]Extensible Neuroimaging Archive Toolkit: http://www.xnat.org/

Table 2

List of Data Repositories

A selected subset of data repositories and distributors that share open access and restricted neurocognitive datasets including neuropychological tests (NP), anatomical magnetic resonance imaging (MRI), diffusion weighted imaging (DWI), task-based functional MRI (fMRI), resting-state fMRI (rsfMRI) and MRI spectroscopy (MRS). ABIDE, ADNI, NACC, and NDAR all focus on specific disease, while CORR, HCP, NeuroVault, NITRC-IR, and OpenfMRI were designed for reuse of specific types of data. Note: this is not a comprehensive list but a selected sample of frequently used repositories. Not all modalities are available for every participant. Participant count retrieved on May 13, 2015.

Name	Purpose	Participants	Modality	Data Access
ABIDE[21]	Autism	1,112	NP, MRI, fMRI	Open
ADNI[22]	Alzheimer's	2,000+	NP, MRI, fMRI, PET	Restricted
CORR[23]	Reliability	1,630	MRI, rsfMRI	Open
HCP[24]	Connectome	542	NP, MRI, DWI, fMRI, rsfMRI	Open/Restricted
NACC[25]	Alzheimer's	31,872	NP, MRI	Open/Restricted
NDAR	Autism	80,578	NP, MRI, DWI, fMRI, rsfMRI, MRS	Restricted
NeuroVault	Statistical Map	2,029	fMRI	Open
NITRC-IR[26]	Primary Data	6,845	MRI, DWI, fMRI, rsfMRI	Open
OpenfMRI	Primary Data	1,411	MRI, fMRI	Open
PING[27]	Pediatric	1,493	NP, MRI, DWI	Restricted

[21]Autism Brain Imaging Data Exchange: http://fcon_1000.projects.nitrc.org/indi/abide
[22]Alzheimer's Disease neuroimaging Initiative: http://adni.loni.usc.edu
[23]Consortium for Reliability and Reproducibility http://fcon_1000.projects.nitrc.org/indi/CoRR/html/index.html
[24]Human Connectome Project: https://humanconnectome.org
[25]National Alzheimer's Coordinating Center: https://www.alz.washington.edu/WEB/researcher_home.html
[26]Neuroimaging Informatics Resource Technology Clearinghouse Image Repository: http://www.nitrc.org/ir
[27]Pediatric Imaging Neurocognition and Genetics: https://pingstudy.ucsd.edu